Tracking-Based Distributed Equilibrium Seeking for Aggregative Games

Guido Carnevale , *Member, IEEE*, Filippo Fabiani , Filiberto Fele ,
Kostas Margellos , *Member, IEEE*, and 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 local constraints are present and we design an algorithm combining, for each agent, the projected pseudogradient descent and a tracking mechanism to locally reconstruct the aggregative variable. To handle coupling constraints arising in generalized settings, we propose another distributed algorithm based on a recently emerged augmented primal-dual scheme and 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 and compare them with state-of-the-art distributed equilibrium-seeking algorithms.**

Index Terms—**Distributed algorithms, game theory, network analysis and control, optimization algorithms.**

I. INTRODUCTION

RECENT years have seen an increasing attention to the computation of (generalized) Nash equilibria in games over networks [1], [2], [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], [9], [10], network congestion control [11], and synchronization of coupled oscillators in power grids [12] can be modeled 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 modeled 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], [14], [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], [17], [18], [19], [20].

A. Related Work

In the context of NE problems in aggregative form, first attempts to design equilibrium seeking algorithms involve semidecentralized 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], [22], [23], [24], [25], [26], [27], [28].

To relax the communication requirements, Koshal et al.[29] proposed a gradient-based algorithm for nongeneralized 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 guarantees 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

TABLE I
SETUP WITHOUT COUPLING CONSTRAINTS

	[29]	[33]	[32]	[38]	Our algorithm
Linear rate	✗	✗	✗	✓	✓
Step-size	Diminishing	-	Diminishing	-	Constant
Exactness	✓	✗	✗	✓	✓
Communications per iterate	1	v	1	1	1
Equilibrium assumptions	F strictly monotone	$\exists!$ equilibrium, $x_{i,\mathrm{br}}(\cdot)$ nonexpansive	F strongly monotone	F strongly monotone	F strongly monotone
Local constraint set	Compact and convex	Compact and convex	Unconstrained	Closed and convex	Closed and convex
Gradient unboundedness	✓	✓	✗	✓	✓
Aggregative variable	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} \phi_i(x_i)$
Algorithmic structure	Gradient-based	Best-response-based	Gradient-based	Proximal-based	Gradient-based
Graph	Undirected, time-varying	Directed	Undirected, time-varying	Undirected	Directed

TABLE II
SETUP WITH COUPLING CONSTRAINTS, WHERE SCQ STANDS FOR SLATER'S CONSTRAINT QUALIFICATION

	[35]	[36]	[37]	[38]	Our algorithm
Linear rate	✗	✗	✗	✗	✓
Step-size	Constant	Diminishing	Constant	-	Constant
Exactness	✗	✓	✓	✓	✓
Communications per iterate	v	1	2	1	1
Equilibrium assumptions	F strongly monotone	F cocoercive	F strongly monotone	F strongly monotone	F strongly monotone
Local constraints	Compact and convex	Compact and convex	Compact and convex	Closed and convex	Unconstrained
Coupling constraints	Not specified	SCQ	SCQ	SCQ	$\kappa_1 I \le AA^{\top} \le \kappa_2 I$
Gradient unboundedness	✗	✓	✓	✓	✓
Aggregative variable	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} x_i$	$\frac{1}{N}\sum_{i=1}^{N} \phi_i(x_i)$
Algorithmic structure	Gradient-based	Gradient-based	Gradient-based	Proximal-based	Gradient-based
Graph	Directed	Undirected, time-varying	Undirected	Undirected	Directed

TABLE III
EXECUTION TIME OF A SINGLE ITERATE WITH PRIMAL TRADES
(ALGORITHM 1) AND THE CONSIDERED ALGORITHMS IN [33] AND [38].
MEAN AND 1-STANDARD DEVIATION ARE BASED ON MONTE CARLO
SIMULATIONS OF THE CASE STUDY IN [33]

Time of iterate	Alg. 1	Alg. 2 in [33]	Alg. 4 in [38]
Mean (s)	8×10^{-4}	6.38×10^{-2}	7.67×10^{-2}
Std. Dev. (s)	2×10^{-4}	9.6×10^{-3}	1.18×10^{-2}

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. Finally, distributed equilibrium-seeking algorithms based on proximal best-responses are proposed in [38].

B. Contributions

The main contribution of the article 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 local constraints are present, we combine a projected pseudogradient method with a local, auxiliary variable that compensates for the lack of knowledge of the aggregative variable in each agent. Successively, we deal with the case of coupling constraints, however, no local constraints are present. To achieve this, we take inspiration from a recent augmented primal-dual scheme for centralized,

continuous-time optimization [39] and resort to 1) an averaging step to enforce consensus among the agents' multipliers and 2) 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 (G)NE. To the best of our knowledge, the algorithm proposed for the case where coupling constraints are present is the first distributed scheme in the literature with guaranteed linear convergence to a GNE (see Section A of Appendix for the formal definition). As such, it constitutes the main contribution of our article. Moreover, as discussed in detail in Section II-B, such a linear convergence rate is enabled by our system-theoretic interpretation, which offers a new proof-line perspective. Further, we also guarantee linear convergence when only local constraints are present. A similar result is also achieved by the recent contribution [38] (see [38, Rem. 12]): our algorithm complements the proximal best-response scheme of Bianchi et al. [38] by constituting its gradient-based counterpart. Proximal algorithms require solving some optimization program (which in turn may rely on some iterative method), whereas our projected gradient descent step can allow a simpler update rule if the projection can be performed in an easy manner. As such, gradient-based approaches are often computationally less intensive compared to proximal ones, as verified in the numerical simulations of Section V (see Table III); however, such a conclusion is case-dependent.

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) (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, describes the evolution of the auxiliary variables used to compensate for 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 (G)NE of the whole interconnection. This last step relies on a general theorem (cf., Theorem 2.5) 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 semiglobal properties (see [40, Prop. 8.1], or [41, 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 compare them with state-of-the-art distributed NE-seeking algorithms.

C. Article 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 Section Section B of Appendix; Sections C and D of Appendix 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. 0_m is the all-zero matrix in $\mathbb{R}^{m \times m}$. The vector of N ones is denoted by 1_N, while $\mathbf{1}_{N,d} := 1_N \otimes I_d$ 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 $\nabla_1 f \in \mathbb{R}^{n_1}$ its gradient with respect to its first argument and as $\nabla_2 f \in \mathbb{R}^{n_2}$ its gradient with respect to the second one. The vertical concatenation of column vectors $v_1, \ldots, v_N \in \mathbb{R}^n$ is $\mathrm{COL}(v_1, \ldots, v_N)$. $\mathbb{R}^n_+$ is the positive orthant in $\mathbb{R}^n$. $\mathrm{diag}(v_1, \ldots, v_n)$ denotes the diagonal matrix whose ith diagonal element is given by v_i. $\mathrm{blkdiag}(M_1, \ldots, M_N)$ is the block diagonal matrix whose ith 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. For matrix (resp., vector) $A \in \mathbb{R}^{m \times n}$ ($v \in \mathbb{R}^n$), we denote as $[A]_j$ ($[v]_j$) its jth row (jth component). Given two matrices $A, B \in \mathbb{R}^{m \times m}$, $A \succ B$ (resp. $A \succeq B$) is equivalent to saying that $A - B$ is positive definite (resp. semidefinite). Given $x \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n \times n}$ such that $M = M^\top \succ 0$, $\|x\|_M = \sqrt{x^\top M x}$.

II. MATHEMATICAL PRELIMINARIES

A. Problem Definition and Main Assumptions

We consider a population of $N \in \mathbb{N}$ agents—designated by the set $\mathcal{I} := \{1, \ldots, N\}$—whose interaction is described by the following collection of coupled optimization problems:

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

In words, every agent $i \in \mathcal{I}$ seeks an individual strategy $x_i \in X_i \subseteq \mathbb{R}^{n_i}$ to minimize a local cost defined by the function $J_i : \mathbb{R}^{n_i} \times \mathbb{R}^d \to \mathbb{R}$, which depends on x_i as well as on some aggregate measure of other agents' strategies $\sigma(x) \in \mathbb{R}^d$, where $x := \mathrm{COL}(x_1, \ldots, x_N) \in \mathbb{R}^n$ and $n := \sum_{i=1}^N n_i$. The agents' decisions shall satisfy some global constraints, which can be expressed in the affine form $Ax \leq b$, where $A := [A_1 \cdots A_N] \in \mathbb{R}^{m \times n}$ and $b := \sum_{i \in \mathcal{I}} b_i \in \mathbb{R}^m$. The aggregative variable $\sigma(\cdot)$ formally reads as

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

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:

$$c_i(x_i) = A_i x_i - b_i \qquad (3a)$$

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

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

where $x_{-i} := \mathrm{COL}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \in \mathbb{R}^{n-n_i}$. Then, the collective vector of strategies x belongs to the feasible set $\mathcal{C} := \{x \in X \mid c(x) \leq 0\} \subseteq \mathbb{R}^n$.

We refer to any equilibrium solution to the collection of interdependent 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 2.1 (GNE [3]): A collective vector of strategies $x^\star \in \mathcal{C}$ is a GNE of (1) if we have:

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

for all $i \in \mathcal{I}$, with $\mathcal{C}_i(x_{-i}) := \{x_i \in X_i \mid A_i x_i \leq b_i - c_{-i}(x_{-i})\}$ and $\sigma_{-i}(x_{-i}^\star) := \frac{1}{N} \sum_{j \in \mathcal{I} \setminus \{i\}} \phi_j(x_j^\star)$. □

We remark that the definition of NE follows directly from the above by noting that, in the case without coupling constraints, it holds $\mathcal{C}_i(x_{-i}^\star) \equiv X_i$. An equivalent definition of GNE requires

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

$$x_{i,\mathrm{br}}(x_{-i}) \in \underset{x_i \in \mathcal{C}_i(x_{-i})}{\arg\min} \quad J_i(x_i, \sigma(x))$$

$$= \underset{x_i \in \mathcal{C}_i(x_{-i})}{\arg\min} \quad J_i\left(x_i, \tfrac{1}{N}\phi_i(x_i) + \sigma_{-i}(x_{-i})\right).$$

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

Standing Assumption 2.2 (Cost functions): For all $i \in \mathcal{I}$, 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 $x_{-i} \in \mathbb{R}^{n-n_i}$. $\square$

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

$$F(x) := \mathrm{COL}(\nabla_{x_1} J_1(x_1, \sigma(x)), \ldots, \nabla_{x_N} J_N(x_N, \sigma(x))). \quad (4)$$

With this regard, we also make the following assumption.

Standing Assumption 2.3 (Strong monotonicity and Lipschitz continuity): F is μ-strongly monotone, i.e., there exists $\mu > 0$ such that

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

for all $x, y \in \mathbb{R}^n$. 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

$$\|\nabla_{x_i} J_i(x_i, \phi_i(x_i)/N + y) - \nabla_{x_i'} J_i(x_i', \phi_i(x_i')/N + y')\|$$
$$\leq \beta_1 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|$$

$$\|\nabla_1 J_i(x_i, y) - \nabla_1 J_i(x_i', y')\| \leq \beta_1 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|$$

$$\|\nabla_2 J_i(x_i, y) - \nabla_2 J_i(x_i', y')\| \leq \beta_2 \|\mathrm{COL}(x_i, y) - \mathrm{COL}(x_i', y')\|$$

$$\|\phi_i(x_i) - \phi_i(x_i')\| \leq \beta_3 \|x_i - x_i'\|.$$

$\square$

While assumptions on strong monotonicity and Lipschitz continuity of the game mapping are quite standard in the literature [15], [25], [26], the second part of Standing Assumption 2.3 specializes 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, 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 $\mathcal{N}_i^{\mathrm{out}} := \{j \in \mathcal{I} \mid (i, j) \in \mathcal{E}\}$ 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 next assumption characterizes the considered graphs.

Standing Assumption 2.4 (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

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

$\square$

B. Key Result on Singularly Perturbed Systems

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

Theorem 2.5 (Global exponential stability for singularly perturbed systems): Consider the system

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

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

with $x^t \in \mathcal{D} \subseteq \mathbb{R}^n$, $w^t \in \mathbb{R}^m$, $f : \mathcal{D} \times \mathbb{R}^m \to \mathbb{R}^n$, $g : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$, $\delta > 0$. Let f and g be Lipschitz continuous with respect to both x^t and w^t, with Lipschitz constants L_f and L_g, respectively. Assume that there exists an L_h-Lipschitz continuous function $h : \mathbb{R}^n \to \mathbb{R}^m$ such that, for all $x \in \mathbb{R}^n$,

$$h(x) = g(h(x), x, \delta)$$

and further assume that there exists $x^\star \in \mathbb{R}^n$ such that

$$0 = \delta f(x^\star, h(x^\star)).$$

Then, let

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

be the *reduced system* and

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

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 all $\psi, \psi_1, \psi_2 \in \mathbb{R}^m$, $x \in \mathbb{R}^n$,

$$b_1 \|\psi\|^2 \leq U(\psi) \leq b_2 \|\psi\|^2 \quad (8a)$$

$$U(g(\psi + h(x), x, \delta) - h(x)) - U(\psi) \leq -b_3 \|\psi\|^2 \quad (8b)$$

$$|U(\psi_1) - U(\psi_2)| \leq b_4 \|\psi_1 - \psi_2\| \|\psi_1\|$$
$$+ b_4 \|\psi_1 - \psi_2\| \|\psi_2\|. \quad (8c)$$

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 all $x, x_1, x_2, x_3 \in \mathcal{D}$

$$c_1 \|x - x^\star\|^2 \leq W(x) \leq c_2 \|x - x^\star\|^2 \quad (9a)$$

$$W(x + \delta f(x, h(x))) - W(x) \leq -\delta c_3 \|x - x^\star\|^2 \quad (9b)$$

$$|W(x_1) - W(x_2)| \leq c_4 \|x_1 - x_2\| \|x_1 - x^\star\|$$
$$+ c_4 \|x_1 - x_2\| \|x_2 - x^\star\|. \quad (9c)$$

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 all $(x^0, w^0) \in \mathcal{D} \times \mathbb{R}^m$. $\square$

Theorem 2.5 establishes a stability result for the system in (5), which can be thought of as an interconnection of a fast and a slow subsystem (for sufficiently small $\delta > 0$). This is schematically illustrated in Fig. 1. To analyze this interconnection, we study separately the simplified auxiliary systems (6) and (7). For all $x \in \mathbb{R}^n$, $h(x)$ is a parametric equilibrium of the fast subsystem; we can then fix the slow state $x^t = x$ into the fast dynamics (5b), to obtain the so-called boundary layer system, as pictorially shown in Fig. 2. This is the auxiliary system described by (7), whose state ψ^t encodes the distance of the
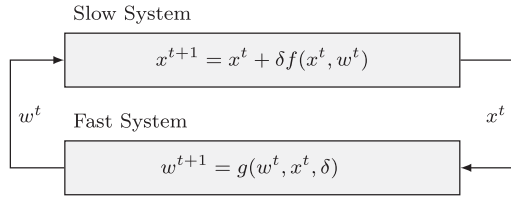
Slow System

$$x^{t+1} = x^t + \delta f(x^t, w^t)$$

w^t Fast System x^t

$$w^{t+1} = g(w^t, x^t, \delta)$$

Fig. 1.   Block diagram of the original interconnected system (5).

Boundary Layer System

$$x^{t+1} = x^t = x$$

x

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

ψ^t
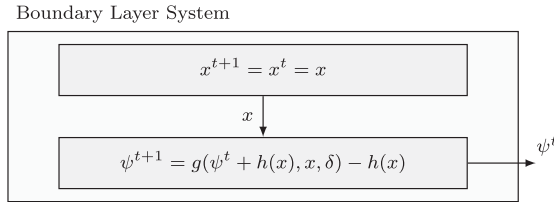
Fig. 2.   Block diagram of the boundary layer system (7).

Reduced System

$$w^t = h(x^t)$$

$h(x^t)$

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

x^t
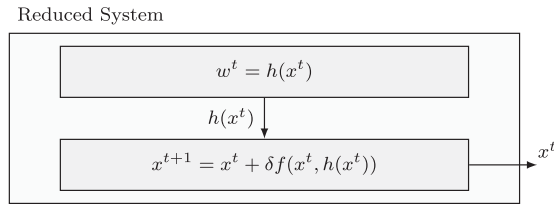
Fig. 3.   Block diagram of the reduced system (6).

state w^t of the fast subsystem from the equilibrium $h(x)$, once x is fixed. Existence of a Lyapunov-like function with properties as in (8) ensures then that for all $x \in \mathbb{R}^n$, the boundary layer system is exponentially stable. Setting now $w^t = h(x^t)$ for all $t \geq 0$ in (5a), i.e., considering the fast state at its parametric equilibrium, we obtain the so-called reduced system, i.e., the auxiliary system (6). This is schematically shown in Fig. 3. Existence of a Lyapunov-like function with properties as in (8) ensures then that $x^\star$ is globally exponentially stable for the reduced system. By properly combining these two Lyapunov functions, Theorem 2.5 ensures that, for sufficiently small values of δ, the point $(x^\star, h(x^\star))$ is globally exponentially stable for the original interconnected system (5). The detailed proof is provided in Section Section B of Appendix. We will show next 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 2.5, and hence prove their convergence in Theorems 3.3 and 4.3, 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 a special case of the aggregative game described by (1), i.e., where the local decision spaces are decoupled. Formally,

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

where $X_i \subseteq \mathbb{R}^{n_i}$, the local feasible set known to agent i only, satisfies the following conditions.

Assumption 3.1: For all $i \in \mathcal{I}$, the feasible set X_i is nonempty, closed, and convex. □

Remark 3.2: The general structure of the aggregative variable $\sigma(x)$ in (2) can accommodate "soft" possibly nonlinear, coupling constraints; these can be incorporated in the game (10) by penalizing their residual in the players' cost functions. □

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 pseudogradient step may be an effective way to steer each agent's strategy to the best response $x_{i,\mathrm{br}}(\sigma_{-i}(x_{-i}^t))$. When applied to problem (10), it reads as

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

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 for 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 $i \in \mathcal{I}$, let the operator $\tilde{F}_i : \mathbb{R}^{n_i} \times \mathbb{R}^d \to \mathbb{R}^{n_i}$ be 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

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

which can be directly implemented without violating the distributed nature of the algorithm. By comparing (11) and (12), we note that the global term $\sigma(x^t)$ has been replaced by the locally available proxy $\phi_i(x_i^t) + z_i^t$. Therefore, if

$$z_i^t \to -\phi_i(x_i^t) + \sigma(x^t) \qquad (13)$$

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 ith 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., [42], where a similar scheme has been used to locally compensate the missing knowledge of the global gradient of a distributed consensus optimization problem):

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

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

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

Algorithm 1: Primal TRADES (Agent i).

Initialization: $x_i^0 \in X_i, z_i^0 = 0$.
for $t = 1, 2, \ldots$ **do**
$$x_i^{t+1} = x_i^t + \delta \left(P_{X_i} \left[x_i^t - \gamma \tilde{F}_i \left(x_i^t, \phi_i(x_i^t) + z_i^t \right) \right] - x_i^t \right)$$
(15a)

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

end for

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

with $X := \prod_{i \in \mathcal{I}} X_i \subseteq \mathbb{R}^n$, $\mathcal{W}_d := \mathcal{W} \otimes I_d \in \mathbb{R}^{Nd \times Nd}$, $z^t :=$ COL$(z_{1,t}, \ldots, z_{N,t})$, $\phi(x^t) := $ COL$(\phi_1(x_1^t), \ldots, \phi_N(x_N^t))$, and $\tilde{F}(x^t, \phi(x^t) + z^t) := $ 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 remark that, since F is μ-strongly monotone (cf., Assumption 2.3) and X nonempty, closed, and convex (cf., Assumption 3.1), there exists a unique Nash equilibrium $x^\star \in \mathbb{R}^n$ for (10). Moreover, for such an equilibrium it holds

$$x^\star = P_X \left[x^\star - \gamma F(x^\star) \right]$$

for any $\gamma > 0$, see [43, Ch. 12]. This result, in turn, guarantees that $x^\star = x^\star + \delta(P_X[x^\star - \gamma F(x^\star)] - x^\star)$ for any $\delta > 0$. We establish next the properties of Primal TRADES in computing the NE $x^\star$ of problem (10).

Theorem 3.3: Consider the dynamics in (16) and Assumption 3.1. 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

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

$\square$

The proof of Theorem 3.3 relies on a *singular perturbation* analysis of system (16), and will be given in the next section.

B. Proof of Theorem 3.3

We build the framework to prove Theorem 3.3 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

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

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

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

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

$$\bar{z}^{t+1} = \frac{\mathbf{1}_{N,d}^\top}{N} z^{t+1} \overset{(a)}{=} \frac{\mathbf{1}_{N,d}^\top}{N} \mathcal{W}_d z^t + \frac{\mathbf{1}_{N,d}^\top}{N} (\mathcal{W}_d - I)\phi(x^t)$$

$$\overset{(b)}{=} \frac{\mathbf{1}_{N,d}^\top}{N} z^t \overset{(c)}{=} \frac{\mathbf{1}_{N,d}^\top}{N} \left(\mathbf{1}_{N,d} \bar{z}^t + R_d z_\perp^t \right) \overset{(d)}{=} \bar{z}^t$$ (19)

where in (a) we exploit the update (16), in (b) we use the facts that, in view of Standing Assumption 2.4, 1) $\mathbf{1}_{N,d}^\top \mathcal{W}_d = \mathbf{1}_{N,d}^\top$

and 2) $\mathbf{1}_{N,d}^\top(\mathcal{W}_d - I) = 0$, 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

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

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

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

$$w^t := z_\perp^t$$

$$f(x^t, w^t) := P_X \left[x^t - \gamma \tilde{F}(x^t, \phi(x^t) + R_d w^t) \right] - x^t$$

$$g(w^t, x^t) := R_d^\top \mathcal{W}_d R_d w^t + R_d^\top (\mathcal{W}_d - I)\phi(x^t).$$ (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: Under the expression for $R_d R_d^\top$ in (18) and since $\mathcal{W}$ is doubly stochastic (cf., Standing Assumption 2.4) notice that for all $x^t = x \in \mathbb{R}^n$,

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

constitutes an equilibrium of (20b). Since $R_d^\top \mathcal{W}_d R_d$ is Schur in view of Standing Assumption 2.4, 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 $z_\perp$ always remains close to the parametrized equilibrium $h(x^t)$.

3) Boundary layer system and satisfaction of (8): the so-called boundary layer system associated with (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 $\tilde{z}^t := z_\perp^t - h(x)$. Using (18), we obtain that

$$\tilde{z}^{t+1} = R_d^\top \mathcal{W}_d R_d \tilde{z}^t.$$ (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 3.4: 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. $\square$

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 all $t \geq 0$. We thus have

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

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

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

Lemma 3.5: 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 $W : \mathbb{R}^n \to \mathbb{R}$ and $\bar{\gamma}, \bar{\delta}_2 > 0$ such that, for any $\gamma \in (0, \bar{\gamma})$ and $\delta \in (0, \bar{\delta}_2)$, W satisfies (9). $\square$

5) Lipschitz continuity of f, g and h: as we will be invoking Theorem 2.5, 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 2.3. To show Lipschitz continuity of g in (21) notice that for all $w, w' \in \mathbb{R}^{(N-1)d}$ and all $x, x' \in \mathbb{R}^n$,

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

where the inequality is due to triangle inequality and the fact that by Standing Assumption 2.3, ϕ is Lipschitz continuous with Lipschitz constant β_3. To show Lipschitz continuity of h, notice that for all $x, x' \in \mathbb{R}^n$,

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

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

By combining Lemmas 3.4 and 3.5 with the Lipschitz conditions expressed above, Theorem 2.5 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. GNE 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 affine 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 4.1 (Feasibility): The set $\mathcal{C}$ is nonempty. ☐

Note that the condition $\mathcal{C} \neq \emptyset$ is weaker than Slater's constraint qualification required by many results in the literature. However, to establish linear convergence of our distributed algorithm, we will enforce an additional assumption on the matrix A (see Assumption 4.2). Consider the following variational inequality, defined by the mapping F in (4) over the domain $\mathcal{C}$:

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

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. 3.9]). The converse, however, does not hold in general. However, since F is strongly monotone (cf., Standing Assumption 2.3) and $\mathcal{C}$ is nonempty (cf., Assumption 4.1), closed and convex (since the constraint are in the form $Ax \leq b$), [43, Prop. 1.4.2 and Th. 2.3.3] guarantee that a unique v-GNE exists, and this satisfies (26).

In the following, we devise an iterative algorithm that asymptotically returns the (unique) v-GNE of (1). Inspired by Qu and Li [39], where an augmented primal-dual scheme was used for continuous-time, centralized optimization, we require the following additional condition on the matrix A, which characterizes the coupling constraints [cf., (1b)].

Assumption 4.2 (Full-row rank): There exist $\kappa_1, \kappa_2 > 0$ such that $\kappa_1 I_m \preccurlyeq AA^\top \preccurlyeq \kappa_2 I_m$. ☐

We note that Assumption 4.2 imposes, as a necessary condition, the fact that $m \leq n$, i.e., that the number of constraints is

at most equal to the total number of components of the global strategy vector.

Following [39], 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 with 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:

$$x_i^{t+1} = x_i^t - \delta \left(\nabla_{x_i} J_i(x_i^t, \sigma(x^t)) + \nabla_{x_i} H(Ax^t - b, \lambda^t) \right) \quad (28a)$$

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

where x_i^t and δ 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 ℓth vector of the canonical basis of $\mathbb{R}^m$, $\ell \in \{1, \ldots, m\}$. The stacked-column form of (28) is

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

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

where $\nabla_x H(Ax^t - b, \lambda^t) := \text{COL}(\nabla_{x_1} H(Ax^t - b, \lambda^t), \ldots, \nabla_{x_N} H(Ax^t - b, \lambda^t))$. By computing the KKT conditions of the VI (26) and using [39, Prop. 1], we obtain that the v-GNE $x^\star$ and the corresponding (unique) optimal multiplier $\lambda^\star \in \mathbb{R}^m$ are such that

$$0 = F(x^\star) + \nabla_x H(Ax^\star - b, \lambda^\star) \quad (31a)$$

$$0 = \nabla_\lambda H(Ax^\star - b, \lambda^\star). \quad (31b)$$

The above result ensures that $\text{COL}(x^\star, \lambda^\star)$ represents an equilibrium point of (30) for any $\delta > 0$.

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 $i \in \mathcal{I}$ 1) 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, 2) a local copy $\lambda_i \in \mathbb{R}^m$ of the multiplier

λ, and 3) an additional averaging step to enforce consensus among the multipliers λ_i [cf., (33b)–(33d)]. As already done in (14), 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

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

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

$$(32)$$

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 (33a) and (33b) in Algorithm 2). The steps of the proposed method are hence summarized in Algorithm 2 from the perspective of agent i, which is then referred to as Primal-Dual TRADES. Note that all the quantities involved in the agent's calculations are purely local, thus making Algorithm 2 fully distributed.

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 $t = 0, 1, \dots$ **do**

$$x_i^{t+1} = x_i^t - \delta \tilde{F}_i(x_i^t, \phi_i(x_i^t) + z_i^t)$$
$$- \delta G_{x,i}(N(A_i x_i^t - b_i) + y_i^t, \lambda_i^t) \qquad (33a)$$

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

$$(33b)$$

$$z_i^{t+1} = \sum_{j \in \mathcal{N}_i} w_{ij} z_j^t + \sum_{j \in \mathcal{N}_i} w_{ij} \phi_j(x_j^t) - \phi_i(x_i^t) \qquad (33c)$$

$$y_i^{t+1} = \sum_{j \in \mathcal{N}_i} w_{ij} y_j^t + \sum_{j \in \mathcal{N}_i} w_{ij} N(A_j x_j^t - b_j)$$

$$- N(A_i x_i^t - b_i) \qquad (33d)$$

end for

Differently from customary primal-dual schemes, (33b) does not need the projection over the positive orthant $\mathbb{R}_+^m$ due to the chosen augmented Lagrangian functions L_i (27). We only need to initialize $\lambda_i^0 \geq 0$ for all $i \in \mathcal{I}$, and choose δ 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 all $j \in \mathcal{N}_i$, 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 $\ell \in \{1, \dots, m\}$. Indeed, specializing (33b) for this case leads to the following update of that ℓth component of λ_i^t:

$$[\lambda_i^{t+1}]_\ell = \left(w_{ii} - \frac{\delta}{\rho}\right)[\lambda_i^t]_\ell. \qquad (34)$$

From (34), we conclude that $[\lambda_i^{t+1}]_\ell$ remains nonnegative if $[\lambda_i^t]_\ell$ is nonnegative, thus alleviating the need for a projection, as long as δ and ρ satisfy $w_{ii} > \delta/\rho$. This feature plays a key role in proving exponential stability properties for the continuous-time,

centralized primal-dual scheme proposed in [39]. As in the case without coupling constraints, the purpose of the initialization step will become clear in the next section. The steps of Algorithm 2 in (33) can be compactly written as

$$x^{t+1} = x^t + \delta f_x(x^t, \lambda^t, z^t, y^t) \qquad (35a)$$

$$\lambda^{t+1} = \mathcal{W}_m \lambda^t + \delta G_\lambda(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t) \qquad (35b)$$

$$z^{t+1} = \mathcal{W}_d z^t + (\mathcal{W}_d - I)\phi(x^t) \qquad (35c)$$

$$y^{t+1} = \mathcal{W}_m y^t + (\mathcal{W}_m - I)N(\bar{A}x^t - \bar{b}) \qquad (35d)$$

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

$$f_x(x, \lambda, z, y) := -\tilde{F}(x, \phi(x) + z) - G_x(N(\bar{A}x - \bar{b}) + y, \lambda)$$

and, similarly to (16), $\lambda : \text{COL}(\lambda_1, \dots, \lambda_N)$, $\mathcal{W}_d := \mathcal{W} \otimes I_d$, $\mathcal{W}_m := \mathcal{W} \otimes I_m$, $G_x(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t) := \text{COL}(G_{x,1}(N(A_1 x_1^t - b_1) + y_1^t, \lambda_1^t), \dots, G_{x,N}(N(A_N x_N^t - b_N) + y_N^t, \lambda_N^t))$, and $G_\lambda(N(\bar{A}x^t - \bar{b}) + y^t, \lambda^t) := \text{COL}(G_{\lambda,1}(N(A_1 x_1^t - b_1) + y_1^t, \lambda_1^t), \dots, G_{\lambda,N}(N(A_N x_N^t - b_N) + y_N^t, \lambda_N^t))$. Next, we establish the convergence properties of Primal-Dual TRADES in computing the v-GNE of (1).

Theorem 4.3: Consider 35 and Assumptions 4.1 and 4.2. 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 $\bar{\delta}, a_1, a_2 > 0$ such that, for any $\delta \in (0, \bar{\delta})$, with $w_{ii} > \frac{\delta}{\rho}$ for all $i \in \{1, \dots, N\}$, it holds

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

$\square$

Note that the additional condition $w_{ii} > \delta/\rho$ needs to be satisfied by δ, given ρ, to ensure the dual variables remain nonnegative, as discussed below (34). As in the case of NE seeking without coupling constraints, the proof of Theorem 4.3 relies on a *singular perturbations* analysis of system (35). We provide this in the next section.

B. Proof of Theorem 4.3

As with the proof of Theorem 3.3, we show that the setting of Theorem 4.3 fits the framework of Theorem 2.5, and organize its proof in five steps.

1) Bringing (35) in the form of (5): we introduce the change of coordinates

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

$$\begin{bmatrix} \bar{\lambda}^t \\ \lambda_\perp^t \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{1}_{N,m}^\top}{N} \\ R_m^\top \end{bmatrix} \lambda^t \qquad (36)$$

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

$$R_d R_d^\top = I - \frac{\mathbf{1}_{N,d}\mathbf{1}_{N,d}^\top}{N}, \quad R_m R_m^\top = I - \frac{\mathbf{1}_{N,m}\mathbf{1}_{N,m}^\top}{N}. \qquad (37)$$

As in the proof of Theorem 4.3, 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 (36), we can therefore rewrite (35) by ignoring the dynamics of $\bar{z}^t$ and $\bar{y}^t$, thus obtaining the system

$$\chi^{t+1} = \chi^t + \delta f(\chi^t, w^t) \qquad (38a)$$

$$w^{t+1} = Sw^t + K(\delta)u(\chi^t) \qquad (38b)$$

in which

$$\chi^t := \begin{bmatrix} x^t \\ \bar{\lambda}^t \end{bmatrix}, \quad w^t := \begin{bmatrix} \lambda_\perp^t \\ z_\perp^t \\ y_\perp^t \end{bmatrix} \tag{39a}$$

$$f(\chi^t, w^t)$$
$$:= \begin{bmatrix} f_x(x^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t, R_d z_\perp^t, R_m y_\perp^t) \\ \frac{\mathbf{1}_{N,m}^\top}{N} G_\lambda(N(\bar{A}x^t - \bar{b}) + R_m y_\perp^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t) \end{bmatrix} \tag{39b}$$

$$S := \begin{bmatrix} R_m^\top \mathcal{W}_m R_m & 0 & 0 \\ 0 & R_d^\top \mathcal{W}_d R_d & 0 \\ 0 & 0 & R_m^\top \mathcal{W}_m R_m \end{bmatrix} \tag{39c}$$

$$K(\delta) := \begin{bmatrix} \delta R_m^\top & 0 & 0 \\ 0 & R_d^\top (\mathcal{W}_d - I) & 0 \\ 0 & 0 & R_m^\top (\mathcal{W}_m - I) \end{bmatrix} \tag{39d}$$

$$u(\chi^t) := \begin{bmatrix} G_\lambda(N(\bar{A}x^t - \bar{b}) + R_m y_\perp^t, \mathbf{1}_{N,m}\bar{\lambda}^t + R_m\lambda_\perp^t) \\ \phi(x^t) \\ N(\bar{A}x^t - \bar{b}) \end{bmatrix}. \tag{39e}$$

We view (38) as a singularly perturbed system, namely the interconnection between the slow dynamics (38a) and the fast one (38b). Indeed, system (38) can be obtained from (5) by considering χ^t as the state of (5a) and setting

$$g(\chi^t, w^t, \delta) := Sw^t + K(\delta)u(\chi^t). \tag{40}$$

2) Equilibrium function h: Under the double stochasticity condition of $\mathcal{W}$, due to Standing Assumption 2.4, and using (37), for all $\chi^t = \chi$,

$$h(\chi) := \begin{bmatrix} 0 \\ -R_d^\top \phi([\ I_n \ \ 0\ ]\chi) \\ -R_m^\top N (\bar{A}[\ I_n \ \ 0\ ]\chi - \bar{b}) \end{bmatrix} \tag{41}$$

constitutes an equilibrium of (38b) (parametrized by χ).

3) Boundary layer system and satisfaction of (8): the so-called boundary layer system associated with (38) 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 (37), we then obtain that

$$\tilde{w}^{t+1} = S\tilde{w}^t + \delta \tilde{u}(\chi, \tilde{w}^t) \tag{42}$$

where

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

The next lemma provides a Lyapunov function for (42).

Lemma 4.4: Consider system (42). 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 $\delta \in (0, \bar{\delta}_1)$, 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 (38a) at steady state, i.e., $w^t = h(\chi^t)$ for all $t \geq 0$. We thus have

$$\chi^{t+1} = \chi^t + \delta f(\chi^t, h(\chi^t)). \tag{43}$$

Let us expand (43). Using (37), we obtain

$$x^{t+1} = x^t - \delta \tilde{F}\left(x^t, \mathbf{1}_{N,d}\sigma(x^t)\right)$$

$$- \delta G_x\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right) \tag{44a}$$

$$\bar{\lambda}^{t+1} = \bar{\lambda}^t + \delta \frac{\mathbf{1}_{N,m}^\top}{N} G_\lambda\left(\mathbf{1}_{N,m}(Ax^t - b), \mathbf{1}_{N,m}\bar{\lambda}^t\right). \tag{44b}$$

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, (43) is identical to the original update (30). Given the unique v-GNE $x^\star$ of (30) (see Assumptions 4.1, 4.2) and the associated multiplier $\lambda^\star \in \mathbb{R}^m$, the next lemma provides a Lyapunov function for (43), hence for (30).

Lemma 5.5: Consider system (43) and Assumptions 4.1, 4.2. Then, there exist a continuous function $W : \mathbb{R}^{n+m} \to \mathbb{R}, \bar{\delta} > 0$ such that, for any $\delta \in (0, \bar{\delta})$, W satisfies (9) with χ in place of x. □

5) Lipschitz continuity of f, g, and h: As we will be invoking Theorem 2.5, we need to ensure that the required Lipschitz properties are satisfied. In particular, we need to show that f, g in (39b) and (40), respectively, and h in (41) 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 2.3), and the Lipschitz continuity of G_x and G_λ (that appear in f and g), which is ensured as shown in (64) within the proof of Lemma 4.4.

By combining Lemmas 4.4 and 4.5 with the Lipschitz continuity properties expressed above, we can apply Theorem 2.5. Then, there exists $\bar{\delta} \in (0, \min(\bar{\delta}_1, \bar{\delta}_2))$ so that, for any $\delta \in (0, \bar{\delta})$, $\mathrm{COL}(x^\star, \lambda^\star, h(x^\star, \lambda^\star))$ is an exponentially stable equilibrium point for (38).

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. In both cases, we performed Monte Carlo simulations consisting of 25 trials. In each trial, we randomly generate the problem parameters, the graph of the network, and the initial conditions of the algorithms' variables.

A. Example Without Coupling Constraints

In this section, we consider an instance of problem (10) and perform numerical simulations in which we compare Primal TRADES with Algorithm 2 proposed in [33] and Algorithm 4 proposed in [38]. We consider the multiagent demand response problem considered in [33]. Consider N loads whose electricity consumption $x_i := \mathrm{COL}(x_{i,1}, \ldots, x_{i,T}) \in \mathbb{R}^T$ with $T \in \mathbb{N}$ has to be chosen to solve

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

where $\hat{u}_i \in \mathbb{R}^T$ denotes some nominal energy profile, $\rho_i > 0$ is a constant weighting parameter, and 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
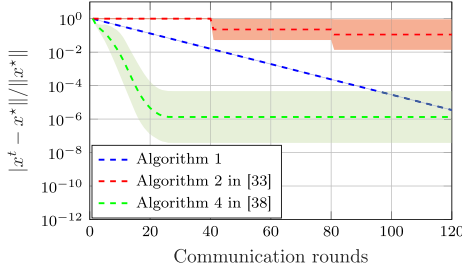
Fig. 4. Mean and 1-standard deviation band (based on Monte Carlo simulations) of the normalized distance of the iterates from the NE achieved by Primal TRADES (Algorithm 1), the algorithm by [33], and the algorithm by [38] on a case study introduced in [33].
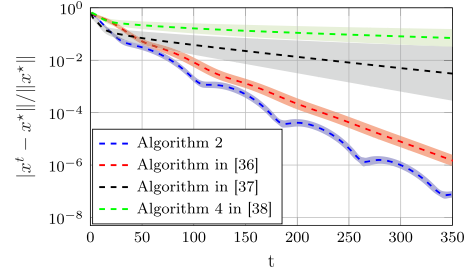

Fig. 5. Mean and 1-standard deviation band (based on Monte Carlo simulations) of the normalized distance of the iterates from the GNE achieved by Primal-Dual TRADES (Algorithm 2), and the algorithms by [36], [37], and [38].

$\sigma(x) = (1/N) \sum_{i \in \mathcal{I}} x_i$. As for the local feasible set $X_i \subseteq \mathbb{R}^T$, for all $i \in \mathcal{I}$, we pick

$$X_i := \left\{ x_i \in \mathbb{R}^T \mid s_{i,\tau+1}(x_i) \in \mathcal{S}_i \text{ and } x_{i,\tau} \in \mathcal{U}_i \ \forall \tau \in \{1, \ldots, T\} \right.$$

$$\left. \sum_{\tau=1}^{T} x_{i,\tau} = \sum_{\tau=1}^{T} \hat{u}_{i,\tau} \right\}$$

where $\mathcal{U}_i \subseteq \mathbb{R}$, $\mathcal{S}_i \subseteq \mathbb{R}$, and $s_{i,\tau}(x_i)$ is the state of the ith load at time τ; this, 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 ith load. To instantiate the problem, we set $T = 24$ and randomly generate values for $\hat{u}_i, \rho_i, \lambda, p_0, a_i, b_i, s_{i,1}$ 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 Erdos–Rényi graph with parameter 0.3.

This setting satisfies our standing assumptions. We compare our scheme, namely, Primal TRADES with [33, Algorithm 2] and [38, Algorithm 4]. We empirically tune the former with $v_1 = v_2 = 20$ 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). We empirically tuned the method by Bianchi et al. [38] choosing $\alpha = 0.1$, $\beta = 1$, and $\tau_i = 0.1$ for all $i \in \mathcal{I}$. As for the parameters of our scheme, we set $\delta = 0.5$ and $\gamma = 0.001$. Fig. 4 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 $x^\star$ with respect to the method in [33], while it turns out to be slower than the algorithm in [38]. This was anticipated as 1) the method in [33] is not guaranteed to converge to the exact NE (see Table I) and 2) the method in [38] is based on proximal-based updates which are known to exhibit faster behavior compared to gradient-based updates, but are computationally more intensive due to the proximal operator involved. In Table III, we provide a numerical comparison of the considered methods in terms of the mean and standard deviation based on Monte Carlo simulations of the time needed to perform a single iterate. As expected,

Primal TRADES turns out to be much lighter than the other algorithms from a computational point of view. The simulations have been executed on MATLAB, using FMINCON() to solve the optimization steps involved in the methods by [33] and [38].

B. Example With Coupling Constraints

Here, we compare our Primal-Dual TRADES algorithm with the distributed methods proposed in [36], [37], and [38]. 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 focus on a case study inspired by Li et al. [16]—where it was addressed within a cooperative scenario—and adapt it as an instance of (1). In particular, we consider the cost function

$$J_i(x_i, \sigma(x)) = \frac{1}{2} \|x_i - p_i\|^2 + \frac{w}{2} \|x_i - \sigma(x)\|$$

where $w > 0$ and $p_i \in \mathbb{R}^{n_i}$ for all $i \in \mathcal{I}$, while $\sigma(x) = \frac{1}{N} \sum_{i \in \mathcal{I}} x_i$. We consider a communication graph with ring topology. As for the coupling constraints, in each trial of the Monte Carlo simulations, we randomly generate each A_i and b_i by imposing the full row rank property for the former (cf. Assumption 4.2) and extracting the latter from the interval [0,100] with a uniform probability; we set $N = 20$, $n_i = 2$. Moreover, in each trial, we uniformly randomly extract each p_i and w from $[0, 100]^2$ and $[0,1]$, respectively. We empirically tune the algorithm in [36] with $\alpha_i = \beta_i = 0.5$ for all $i \in \mathcal{I}$, and $\gamma^t = 0.1$ for all $t \geq 0$. As for the parameters of the method in [37], we empirically choose $c = 1$, $k = 0.1$, $\tau = 0.2$, $\alpha = 0.2$, and $v = 0.1$. The algorithm in [38] has been empirically tuned setting $\alpha = 0.3$, $\beta = 0.1$, $\tau_i = 0.3$ and $\delta_i = 0.3$ for all $i \in \mathcal{I}$. Finally, as for the parameters of our algorithm, we empirically tune them as $\delta = 0.05$ and $\rho = 0.1$. In Fig. 5, we compare the performance of the algorithms in [36], [37], and [38] with Algorithm 2 in terms of the normalized distance $\|x^t - x^\star\|/\|x^\star\|$ from the GNE $x^\star$. In this case, the proposed scheme outperforms the others in terms of accuracy and convergence speed.

VI. CONCLUSION AND OUTLOOK

We propose 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 does not involve local constraints, however, it allows handling coupling constraints, thus encompassing GNE problems. 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. Current work concentrates on extending our analysis line to allow for local constraint sets, either in a hard manner or by means of dualizing these and satisfying them asymptotically. An additional aspect worth of investigation is the possibility of time-varying communication patterns among the agents. Finally, we perform detailed numerical simulations showing the effectiveness of the proposed methods and that they outperform state-of-the-art distributed methods.

APPENDIX A

A. Q-Linear Rate

Here, we report the definition of Q-linear convergence [44, App. A.2]. For the sake of readability, in the rest of the document, we omit the prefix Q.

Definition A.1: Let $\{x^t\}$ be a sequence in $\mathbb{R}^n$ that converges to $x^\star \in \mathbb{R}^n$. We say that the convergence is Q-linear if there is a constant $r \in (0, 1)$ such that

$$\frac{\|x^{t+1} - x^\star\|}{\|x^t - x^\star\|} \leq r$$

for all t sufficiently large. $\qquad\qquad\square$

B. Proof of Theorem 2.5

Let $\tilde{w}^t := w^t - h(x^t)$ and accordingly rewrite (5) as

$$x^{t+1} = x^t + \delta f(x^t, \tilde{w}^t + h(x^t)) \tag{45a}$$

$$\tilde{w}^{t+1} = g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t) \tag{45b}$$

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 (45a), we obtain

$$\Delta W(x^t) = W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t))) - W(x^t)$$

$$\overset{(a)}{=} W(x^t + \delta f(x^t, h(x^t))) - W(x^t)$$
$$+ W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t)))$$
$$- W(x^t + \delta f(x^t, h(x^t)))$$

$$\overset{(b)}{\leq} -\delta c_3 \|x^t - x^\star\|^2 + W(x^t + \delta f(x^t, \tilde{w}^t + h(x^t)))$$
$$- W(x^t + \delta f(x^t, h(x^t)))$$

$$\overset{(c)}{\leq} -\delta c_3 \|x^t - x^\star\|^2 + 2\delta c_4 L_f \|\tilde{w}^t\| \|x^t - x^\star\|$$
$$+ \delta^2 c_4 L_f \|\tilde{w}^t\| \|f(x^t, \tilde{w}^t + h(x^t))\|$$
$$+ \delta^2 c_4 L_f \|\tilde{w}^t\| \|f(x^t, h(x^t))\| \tag{46}$$

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

$$\|f(x^t, \tilde{w}^t + h(x^t))\| = \|f(x^t, \tilde{w}^t + h(x^t)) - f(x^\star, h(x^\star))\|$$

$$\overset{(a)}{\leq} L_f \|x^t - x^\star\| + L_f \|\tilde{w}^t + h(x^t) - h(x^\star)\|$$

$$\overset{(b)}{\leq} L_f(1 + L_h) \|x^t - x^\star\| + L_f \|\tilde{w}^t\| \tag{47}$$

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

$$\|f(x^t, h(x^t))\| \leq L_f(1 + L_h) \|x^t - x^\star\|. \tag{48}$$

Using inequalities (47) and (48), we then bound (46) as

$$\Delta W(x^t) \leq -\delta c_3 \|x^t - x^\star\|^2 + 2\delta c_4 L_f \|\tilde{w}^t\| \|x^t - x^\star\|$$
$$+ \delta^2 c_4 L_f^2 \|\tilde{w}^t\|^2$$
$$+ 2\delta^2 c_4 L_f^2(1 + L_h) \|\tilde{w}^t\| \|x^t - x^\star\|$$

$$\leq -\delta c_3 \|x^t - x^\star\|^2 + \delta^2 k_3 \|\tilde{w}^t\|^2$$
$$+ (\delta k_1 + \delta^2 k_2) \|\tilde{w}^t\| \|x^t - x^\star\| \tag{49}$$

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 (45b), we obtain

$$\Delta U(\tilde{w})$$
$$= U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)) - U(\tilde{w}^t)$$

$$\overset{(a)}{\leq} U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)) - U(\tilde{w}^t)$$
$$- U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t))$$
$$+ U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t))$$

$$\overset{(b)}{\leq} -b_3 \|\tilde{w}^t\|^2 - U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t))$$
$$+ U(g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t))$$

$$\overset{(c)}{\leq} -b_3 \|\tilde{w}^t\|^2$$
$$+ b_4 \|\Delta h(x^{t+1}, x^t)\|$$
$$\times \|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t) + \Delta h(x^{t+1}, x^t)\|$$
$$+ b_4 \|\Delta h(x^{t+1}, x^t)\| \|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)\|$$

$$\overset{(d)}{\leq} -b_3 \|\tilde{w}^t\|^2 + b_4 \|\Delta h(x^{t+1}, x^t)\|^2$$
$$+ 2b_4 \|\Delta h(x^{t+1}, x^t)\| \|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)\| \tag{50}$$

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 difference of the last two terms, and (d) uses the triangle inequality. By using the definition of $\Delta h(x^{t+1}, x^t)$ and the Lipschitz continuity of h, we write

$$\|\Delta h(x^{t+1}, x^t)\| \leq L_h \|x^{t+1} - x^t\|$$

$$\overset{(a)}{\leq} \delta L_h \|f(x^t, \tilde{w}^t + h(x^t))\|$$

$$\overset{(b)}{\leq} \delta L_h \|f(x^t, \tilde{w}^t + h(x^t)) - f(x^\star, h(x^\star))\|$$

$$\overset{(c)}{\le} \delta L_h L_f(1+L_h)\|x^t-x^\star\|+\delta L_h L_f\|\tilde{w}^t\| \tag{51}$$

where in (a) we use the update (45a), 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

$$\big\|g(\tilde{w}^t + h(x^t), x^t, \delta) - h(x^t)\big\|$$

$$= \big\|g(\tilde{w}^t + h(x^t), x^t, \delta) - g(h(x^t), x^t, \delta)\big\| \le L_g\|\tilde{w}^t\| \tag{52}$$

where the inequality is due to the Lipschitz continuity of g. Using inequalities (51) and (52), we then bound (50) as

$$\Delta U(\tilde{w})$$

$$\le -b_3\big\|\tilde{w}^t\big\|^2 + 2\delta b_4 L_h L_g L_f(1+L_h)\big\|x^t-x^\star\big\|\big\|\tilde{w}^t\big\|$$

$$+ 2\delta b_4 L_h L_g L_f\big\|\tilde{w}^t\big\|^2 + \delta^2 b_4 L_h^2 L_f^2 (1+L_h)^2\big\|x^t-x^\star\big\|^2$$

$$+ 2\delta^2 b_4 L_h^2 L_f^2(1+L_h)\big\|x^t-x^\star\big\|\big\|\tilde{w}^t\big\| + \delta^2 b_4 L_h^2 L_f^2\big\|\tilde{w}^t\big\|^2$$

$$\le (-b_3 + \delta k_6 + \delta^2 k_7)\big\|\tilde{w}^t\big\|^2 + \delta^2 k_8\big\|x^t-x^\star\big\|^2$$

$$+ (\delta k_4 + \delta^2 k_5)\big\|x^t-x^\star\big\|\big\|\tilde{w}^t\big\| \tag{53}$$

where we introduce the constants

$$k_4 := 2b_4 L_h L_g L_f(1+L_h), \qquad k_5 := 2b_4 L_h^2 L_f^2(1+L_h)$$

$$k_6 := 2b_4 L_h L_g L_f, \qquad k_7 := b_4 L_h^2 L_f^2$$

$$k_8 := b_4 L_h^2 L_f^2(1+L_h)^2.$$

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 (45), we can use the results (49) and (53) to write

$$\Delta V(x^t, \tilde{w}^t) \le - \begin{bmatrix} \|x^t-x^\star\| \\ \|\tilde{w}^t\| \end{bmatrix}^\top Q(\delta) \begin{bmatrix} \|x^t-x^\star\| \\ \|\tilde{w}^t\| \end{bmatrix} \tag{54}$$

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

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

with $q_{21}(\delta) := -\frac{1}{2}(\delta(k_1+k_4) + \delta^2(k_2+k_5))$. By relying on the Sylvester criterion [41], we know that $Q \succ 0$ if and only if

$$c_3 b_3 > p(\delta) \tag{55}$$

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

$$p(\delta) := q_{21}(\delta)^2 + \delta^2 c_3 k_6 + \delta^2(\delta c_3(k_3+k_7) + b_3 k_8)$$

$$- \delta^3 k_6 k_8 - \delta^4 k_8(k_3+k_7). \tag{56}$$

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 (55) 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 (54) 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 [45, Th. 13.2], that $(x^\star, 0)$ is an exponentially stable equilibrium point for system (45). 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).

C. Proofs of Technical Lemmas of Section III-B

Proof of Lemma 3.4: 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

$$(R_d^\top \mathcal{W}_d R_d)^\top P R_d^\top \mathcal{W}_d R_d - P = -Q \tag{57}$$

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 (57). Hence, (8b) is satisfied by taking b_3 to be the smallest eigenvalue of Q. To see (8c), notice that

$$\big\|U(\tilde{z}_1^t) - U(\tilde{z}_2^t)\big\| = \big\|(\tilde{z}_1^t)^\top P \tilde{z}_1^t - (\tilde{z}_2^t)^\top P \tilde{z}_2^t\big\|$$

$$\overset{(a)}{\le} \big\|(\tilde{z}_1^t)^\top P \tilde{z}_1^t - (\tilde{z}_1^t)^\top P \tilde{z}_2^t\big\| + \big\|(\tilde{z}_2^t)^\top P \tilde{z}_1^t - (\tilde{z}_2^t)^\top P \tilde{z}_2^t\big\|$$

$$\overset{(b)}{\le} \|P\|\big\|\tilde{z}_1^t - \tilde{z}_2^t\big\|\big\|\tilde{z}_1^t\big\| + \|P\|\big\|\tilde{z}_1^t - \tilde{z}_2^t\big\|\big\|\tilde{z}_2^t\big\| \tag{58}$$

where (a) follows from adding and subtracting $(\tilde{z}_1^t)^\top P \tilde{z}_2^t$ and using the triangle inequality, while (b) from the Cauchy–Schwarz inequality. The bound (8c) follows from (58) by setting b_4 as the largest eigenvalue of P.

We provide here the following technical lemma which is used in the proof of Lemma 3.5.

Lemma A.2 (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 all $x, x' \in \mathbb{R}^n$ it holds

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

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

Proof: We have that

$$\|x - \gamma F(x) - x' + \gamma F(x')\|^2$$

$$= \|x-x'\|^2 + \gamma^2\|F(x) - F(x')\|^2$$

$$- 2\gamma(x-x')^\top(F(x) - F(x'))$$

$$\overset{(a)}{\le} \|x-x'\|^2 - \gamma(2\mu - \gamma L^2)\|x-x'\|^2 \tag{59}$$

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

$$\mu\|x-x'\|^2 \le (F(x) - F(x'))^\top(x-x')$$

$$\le \|F(x) - F(x')\|\|x-x'\| \le L\|x-x'\|^2$$

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

Proof of Lemma 3.5: pick $W : \mathbb{R}^n \to \mathbb{R}$ 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

$$\Delta W(x^t)$$

$$= \frac{1}{2} \left\| (1-\delta)x^t + \delta \left(P_X \left[x^t - \gamma F(x^t) \right] \right) - x^\star \right\|^2 - \frac{1}{2} \left\| x^t - x^\star \right\|^2$$

$$\overset{(a)}{\leq} \frac{(1-\delta)^2}{2} \left\| x^t - x^\star \right\|^2$$

$$- \frac{1}{2} \left\| x^t - x^\star \right\|^2$$

$$+ (\delta - \delta^2) \left\| x^t - x^\star \right\| \left\| P_X \left[x^t - \gamma F(x^t) \right] \right.$$

$$\left. - P_X \left[x^\star - \gamma F(x^\star) \right] \right\|$$

$$+ \frac{\delta^2}{2} \left\| P_X \left[x^t - \gamma F(x^t) \right] - P_X \left[x^\star - \gamma F(x^\star) \right] \right\|^2$$

$$\overset{(b)}{\leq} \frac{(1-\delta)^2}{2} \left\| x^t - x^\star \right\|^2 - \frac{1}{2} \left\| x^t - x^\star \right\|^2$$

$$+ (\delta - \delta^2) \left\| x^t - x^\star \right\| \left\| x^t - \gamma F(x^t) - x^\star + \gamma F(x^\star) \right\|$$

$$+ \frac{\delta^2}{2} \left\| x^t - \gamma F(x^t) - x^\star + \gamma F(x^\star) \right\|^2 \qquad (60)$$

where in (a) we introduce $\delta(x^\star - P_X[x^\star - \gamma F(x^\star)])$ 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 all a, b, we have that $\|P_X[a] - P_X[b]\| \leq \|a - b\|$, since the projection operator is nonexpansive. Since F is μ-strongly monotone and β_1 Lipschitz continuous (cf., Standing Assumption 2.3), set $\bar{\gamma} = 2\mu/(\beta_1)^2$ and choose $\gamma \in (0, \bar{\gamma})$. Applying Lemma A.2 yields

$$\left\| x^t - \gamma F(x^t) - x^\star + \gamma F(x^\star) \right\| \leq (1 - \bar{\mu}) \left\| x^t - x^\star \right\| \quad (61)$$

with $\bar{\mu} = 1 - \sqrt{1 - \gamma(2\mu - \gamma(\beta_1)^2)} \in (0, 1]$. Thus, by using the inequality in (61), we can bound (60) as follows:

$$\Delta W(x^t) \leq \frac{(1-\delta)^2}{2} \left\| x^t - x^\star \right\|^2 - \frac{1}{2} \left\| x^t - x^\star \right\|^2$$

$$+ (\delta - \delta^2)(1 - \bar{\mu}) \left\| x^t - x^\star \right\|^2$$

$$+ \delta^2 (1 - \bar{\mu})^2 / 2 \left\| x^t - x^\star \right\|^2$$

$$\overset{(a)}{=} -\delta\bar{\mu} \left(1 - \delta\bar{\mu}/2 \right) \left\| x^t - x^\star \right\|^2 \qquad (62)$$

where (a) is obtained by rearranging the above terms. Thus, for any $\delta \in (0, \bar{\delta}_2)$ with $\bar{\delta}_2 := 2/\bar{\mu}$, $(1 - \delta\bar{\mu}/2) > 0$ in (62), thus establishing condition (9b) and concluding the proof.

D. Proofs of Technical Lemmas of Section IV-B

Proof of Lemma 4.4: Since $R_m^\top \mathbf{1}_{N,m} = 0$, we can write

$$R_m^\top G_\lambda \left(\mathbf{1}_{N,m}(Ax - b) + R_m \tilde{y}_\perp^t, \mathbf{1}_{N,m}\bar{\lambda} + R_m \tilde{\lambda}_\perp^t \right)$$

$$= R_m^\top \bigg(G_\lambda \left(\mathbf{1}_{N,m}(Ax - b) + R_m \tilde{y}_\perp^t, \mathbf{1}_{N,m}\bar{\lambda} + R_m \tilde{\lambda}_\perp^t \right)$$

$$- \mathbf{1}_{N,m} \nabla_\lambda H(Ax - b, \bar{\lambda}) \bigg)$$

$$= R_m^\top \bigg(G_\lambda \left(\mathbf{1}_{N,m}(Ax - b) + R_m \tilde{y}_\perp^t, \mathbf{1}_{N,m}\bar{\lambda} + R_m \tilde{\lambda}_\perp^t \right)$$

$$- G_\lambda (\mathbf{1}_{N,m}(Ax - b), \mathbf{1}_{N,m}\bar{\lambda}) \bigg) \qquad (63)$$

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 [39, Lemma 3], notice that, for all $r_1, r_2 \in \mathbb{R}$, there exists $\epsilon(r_1, r_2) \in [0, 1]$ so that[1]

$$\max\{r_1, 0\} - \max\{r_2, 0\} = \epsilon(r_1, r_2)(r_1 - r_2). \qquad (64)$$

Let us introduce

$$q_i^t := \sum_{\ell=1}^{m} [R_m \tilde{y}_\perp^t]_{\ell + (i-1)m} e_\ell, \quad p_i^t := \sum_{\ell=1}^{m} [R_m \tilde{\lambda}_\perp^t]_{\ell + (i-1)m} e_\ell \qquad (65)$$

and use them to define

$$r_{1,i}^t := \rho(Ax - b + q_i^t) + \bar{\lambda} + p_i^t$$

$$r_{2,i} := \rho(Ax - b) + \bar{\lambda}. \qquad (66)$$

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 (63). Let $n_w := (N - 1)(2m + d)$. As such, for all $\chi \in \mathbb{R}^{n+m}$ and $\tilde{w}^t \in \mathbb{R}^{n_w}$, we use the definition of G_λ in (32), $r_{1,i}^t$ and $r_{2,i}$ in (66), and apply (64) for each component of $\tilde{u}(\chi, \tilde{w}^t)$ obtaining

$$\left\| \tilde{u}(\chi, \tilde{w}^t) \right\|$$

$$\leq \left\| R_m^\top \frac{1}{\rho} \mathrm{COL} \left(\sum_{\ell=1}^{m} \epsilon([r_{1,i}^t]_\ell, [r_{2,i}]_\ell) \left([r_{1,i}^t - \bar{\lambda} - p_i^t]_\ell \right. \right. \right.$$

$$\left. \left. \left. - [r_{2,i} - \bar{\lambda}]_\ell \right) e_\ell \right)_{i=1}^{N} \right\|$$

$$\overset{(a)}{\leq} \left\| R_m^\top \frac{1}{\rho} \mathrm{COL} \left(\sum_{\ell=1}^{m} \left([r_{1,i}^t - \bar{\lambda} - p_i^t]_\ell - [r_{2,i} - \bar{\lambda}]_\ell \right) e_\ell \right)_{i=1}^{N} \right\|$$

$$\overset{(b)}{=} \left\| R_m^\top \frac{1}{\rho} \mathrm{COL} \left(\sum_{\ell=1}^{m} \rho [q_i^t]_\ell e_\ell \right)_{i=1}^{N} \right\|$$

$$\overset{(c)}{=} \left\| R_m^\top R_m \tilde{y}_\perp^t \right\| \overset{(d)}{\leq} \left\| \tilde{w}^t \right\|, \qquad (67)$$

where in (a) we use the fact that $\epsilon([r_{1,i}^t]_\ell, [r_{2,i}]_\ell) \in [0, 1]$ for all $\ell \in \{1, \ldots, m\}$ and $i \in \mathcal{I}$, (b) uses the definitions in (66) to simplify the terms, (c) follows from (65), and (d) uses $R_m^\top R_m = I$ and $\|\tilde{y}_\perp^t\| \leq \|\tilde{w}^t\|$ that holds since $\tilde{y}_\perp^t$ is a component of $\tilde{w}^t$. Now, let $U : \mathbb{R}^{n_w} \to \mathbb{R}$ be

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

where $\mathcal{U} \in \mathbb{R}^{n_w \times n_w}$ with $\mathcal{U} = \mathcal{U}^\top \succ 0$, such that

$$S^\top \mathcal{U} S - \mathcal{U} = -I. \qquad (68)$$

---

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

We remark that such a matrix $\mathcal{U}$ always exists because, in light of Standing Assumption 2.4, 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 (42), obtaining

$$\Delta U(\tilde{w}^t) = (S\tilde{w}^t + \delta\tilde{u}(\chi, \tilde{w}^t))^\top \mathcal{U}(S\tilde{w}^t + \delta\tilde{u}(\chi, \tilde{w}^t))$$
$$\quad - (\tilde{w}^t)^\top \mathcal{U}\tilde{w}^t$$
$$= -\left\|\tilde{w}^t\right\|^2 + 2\delta(\tilde{w}^t)^\top S^\top \mathcal{U}\tilde{u}(\chi, \tilde{w}^t)$$
$$\quad + \delta^2 \tilde{u}(\chi, \tilde{w}^t)^\top \mathcal{U}\tilde{u}(\chi, \tilde{w}^t)$$
$$\leq -(1 - \delta\mu_1 - \delta^2\mu_2)\left\|\tilde{w}^t\right\|^2 \quad (69)$$

where the second equality is due to (68), and the inequality follows from (67) and the Cauchy–Schwarz inequality, with the constants $\mu_1 := 2\|S\|\|\mathcal{U}\|$ and $\mu_2 := \|\mathcal{U}\|$. Thus, there always exists $\bar{\delta}_1 > 0$ small enough so that $(1 - \delta\mu_1 - \delta^2\mu_2) > 0$ for any $\delta \in (0, \bar{\delta}_1)$, concluding the proof.

Proof of Lemma 4.5: the proof is inspired by [39, 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

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

$$\mathcal{H}(\chi^t) := \begin{bmatrix} \nabla_x H\left(A\begin{bmatrix} I & 0 \end{bmatrix}\chi^t - b, \begin{bmatrix} 0 & I \end{bmatrix}\chi^t\right) \\ -\nabla_\lambda H\left(A\begin{bmatrix} I & 0 \end{bmatrix}\chi^t - b, \begin{bmatrix} 0 & I \end{bmatrix}\chi^t\right) \end{bmatrix}. \quad (70a)$$

Applying (64) to each of the components of $\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star)$, for all $\chi^t \in \mathbb{R}^{n+m}$ we obtain

$$\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star)$$
$$= \begin{bmatrix} \rho A^\top E(\chi^t, \chi^\star)A & A^\top E(\chi^t, \chi^\star) \\ -E(\chi^t, \chi^\star)A & -\frac{1}{\rho}(E(\chi^t, \chi^\star) - I) \end{bmatrix}(\chi^t - \chi^\star) \quad (71)$$

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

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

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

$$F(x^t) - F(x^\star) = \int_0^1 \nabla F((1 - \nu)x^\star + \nu x^t)(x^t - x^\star)d\nu$$
$$\overset{(a)}{=} \left[\int_0^1 \nabla F((1 - \nu)x^\star + \nu x^t)d\nu\right](x^t - x^\star)$$
$$\overset{(b)}{=} B(x^t, x^\star)(x^t - x^\star) \quad (72)$$

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 2.3), we can uniformly bound the integrand term of (72) as

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

which leads to

$$\mu I \preccurlyeq \int_0^1 \mu I d\nu \preccurlyeq B(x^t, x^\star) \preccurlyeq \int_0^1 \beta_1 I d\nu \preccurlyeq \beta_1 I. \quad (73)$$

Combining (39b), (70), (71), and (72), we can write

$$f(\chi^t, h(\chi^t)) - f(\chi^\star, h(\chi^\star))$$
$$= -\mathcal{F}(\chi^t) + \mathcal{F}(\chi^\star) - (\mathcal{H}(\chi^t) - \mathcal{H}(\chi^\star))$$
$$= D(\chi^t, \chi^\star)(\chi^t - \chi^\star) \quad (74)$$

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

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

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

$$\left\|D(\chi^t, \chi^\star)(\chi^t - \chi^\star)\right\|_M^2 \leq \mu_1 \left\|\chi^t - \chi^\star\right\|_M^2 \quad (75)$$

where $\mu_1 := (\max\{\beta_1 + \rho\|A\|^2, \frac{1}{\rho}\})^2$ and the inequality follows by inspection of $D(\chi^t, \chi^\star)(\chi^t - \chi^\star)$ and using $\|E(\chi^t, \chi^\star)\| \leq 1$. Now, let $W : \mathbb{R}^{n+m} \to \mathbb{R}$ be defined as

$$W(\chi) = (\chi - \chi^\star)^\top M(\chi - \chi^\star) \quad (76)$$

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

$$M := \begin{bmatrix} cI & A^\top \\ A & cI \end{bmatrix}. \quad (77)$$

Note that $M \succ 0$ for any $c > \sqrt{\kappa_2}$ and, thus, W satisfies (9a) and (9c). To show (9b), we evaluate $\Delta W(\chi^t) := W(\chi^{t+1}) - W(\chi^t)$ along the trajectories of (43), obtaining

$$\Delta W(\chi^t)$$
$$= \left\|\chi^t + \delta f(\chi^t, h(\chi^t)) - \chi^\star\right\|_M^2 - \left\|\chi^t - \chi^\star\right\|_M^2$$
$$\overset{(a)}{=} \left\|\chi^t + \delta f(\chi^t, h(\chi^t)) - \chi^\star - \delta f(\chi^\star, h(\chi^\star))\right\|_M^2$$
$$\quad - \left\|\chi^t - \chi^\star\right\|_M^2$$
$$\overset{(b)}{=} \left\|\chi^t - \chi^\star + \delta D(\chi^t, \chi^\star)(\chi^t - \chi^\star)\right\|_M^2 - \left\|\chi^t - \chi^\star\right\|_M^2$$
$$\overset{(c)}{=} \delta(\chi^t - \chi^\star)^\top(D(\chi^t, \chi^\star)^\top M + MD(\chi^t, \chi^\star))(\chi^t - \chi^\star)$$
$$\quad + \delta^2 \left\|D(\chi^t, \chi^\star)(\chi^t - \chi^\star)\right\|_M^2 \quad (78)$$

where (a) uses the fact that $f(\chi^\star, h(\chi^\star)) = 0$ [cf., (31)], (b) rewrites the quantities using (74), and (c) expands $\|\cdot\|_M^2$. As in [39, Lemma 4], since it holds (73), there exists $\bar{c} > 0$ such that, for any $c > \bar{c}$, it holds

$$D(\chi^t, \chi^\star)^\top M + MD(\chi^t, \chi^\star) \leq -\tau M \quad (79)$$

where $\tau := \frac{\kappa_1}{2c}$. Therefore, by using (79) and (75), we bound the right-hand side of (78) as

$$\Delta W(\chi^t) \leq -\delta(\tau - \delta\mu_1)\left\|\chi^t - \chi^\star\right\|_M^2.$$

Setting $\bar{\delta} := \frac{\tau}{\mu_1}$, (78) 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," *Synth. Lectures Commun. Netw.*, 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 Trans. Inf. Theory*, vol. 60, no. 7, pp. 4197–4231, Jul. 2014.

[3] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Ann. Oper. Res.*, 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 Trans. Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 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 *Proc. Eur. Control Conf.*, 2020, pp. 1324–1329.

[6] K. Okuguchi and F. Szidarovszky, *The Theory of Oligopoly With Multi-Product Firms.*, Berlin, Germany: Springer, 2012.

[7] F. Fabiani, D. Fenucci, and A. Caiti, "A distributed passivity approach to AUV teams control in cooperating potential games," *Ocean Eng.*, 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 Trans. Autom. Control*, vol. 66, no. 9, pp. 4238–4245, Sep. 2021.

[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 *Proc. IEEE 58th Conf. Decis. Control*, 2019, pp. 4904–4909.

[11] J. Barrera and A. Garcia, "Dynamic incentives for congestion control," *IEEE Trans. Autom. Control*, vol. 60, no. 2, pp. 299–310, Feb. 2015.

[12] H. Yin, P. G. Mehta, S. P. Meyn, and U. V. Shanbhag, "Synchronization of coupled oscillators is a game," *IEEE Trans. Autom. Control*, vol. 57, no. 4, pp. 920–935, Apr. 2012.

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

[14] F. Parise and A. Ozdaglar, "Analysis and interventions in large network games," *Annu. Rev. Control Robot. Auton. Syst.*, 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 Syst. Mag.*, vol. 42, no. 4, pp. 87–102, Aug. 2022.

[16] X. Li, L. Xie, and Y. Hong, "Distributed aggregative optimization over multi-agent networks," *IEEE Trans. Autom. Control*, vol. 67, no. 6, pp. 3165–3171, Jun. 2022.

[17] X. Li, X. Yi, and L. Xie, "Distributed online convex optimization with an aggregative variable," *IEEE Trans. Control Netw. Syst.*, vol. 9, no. 1, pp. 438–449, Mar. 2022.

[18] G. Carnevale, A. Camisa, and G. Notarstefano, "Distributed online aggregative optimization for dynamic multi-robot coordination," *IEEE Trans. Autom. Control*, vol. 68, no. 6, pp. 3736–3743, Jun. 2023.

[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," *Int. J. Robust Nonlinear Control*, 2023.

[21] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to Nash equilibria in constrained deterministic mean field control," *IEEE Trans. Autom. Control*, vol. 61, no. 11, pp. 3315–3329, Nov. 2016.

[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 Syst. Lett.*, vol. 1, no. 2, pp. 400–405, Oct. 2017.

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

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

[25] D. Paccagnan, B. Gentile, F. Parise, M. Kamgarpour, and J. Lygeros, "Nash and wardrop equilibria in aggregative games with coupling constraints," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1373–1388, Apr. 2019.

[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 Trans. Autom. Control*, vol. 66, no. 12, pp. 6011–6016, Dec. 2021.

[28] G. Belgioioso and S. Grammatico, "Semi-decentralized generalized Nash equilibrium seeking in monotone aggregative games," *IEEE Trans. Autom. Control*, vol. 68, no. 1, pp. 140–155, Jan. 2023.

[29] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Res.*, 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 Syst. Mag.*, vol. 39, no. 3, pp. 40–72, Jun. 2019.

[32] M. Ye, G. Hu, L. Xie, and S. Xu, "Differentially private distributed Nash equilibrium seeking for aggregative games," *IEEE Trans. Autom. Control*, vol. 67, no. 5, pp. 2451–2458, May 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, 2020, Art. no. 108959.

[34] C. Cenedese, G. Belgioioso, Y. Kawano, S. Grammatico, and M. Cao, "Asynchronous and time-varying proximal type dynamics in multiagent network games," *IEEE Trans. Autom. Control*, vol. 66, no. 6, pp. 2861–2867, Jun. 2021.

[35] F. Parise, B. Gentile, and J. Lygeros, "A distributed algorithm for almost-Nash equilibria of average aggregative games with coupling constraints," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 2, pp. 770–782, Jun. 2020.

[36] G. Belgioioso, A. Nedić, and S. Grammatico, "Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2061–2075, May 2021.

[37] D. Gadjov and L. Pavel, "Single-timescale distributed GNE seeking for aggregative games over networks via forward–backward operator splitting," *IEEE Trans. Autom. Control*, vol. 66, no. 7, pp. 3259–3266, Jul. 2021.

[38] M. Bianchi, G. Belgioioso, and S. Grammatico, "Fast generalized Nash equilibrium seeking under partial-decision information," *Automatica*, vol. 136, 2022, Art. no. 110080.

[39] G. Qu and N. Li, "On the exponential stability of primal-dual gradient dynamics," *IEEE Control Syst. Lett.*, vol. 3, no. 1, pp. 43–48, Jan. 2019.

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

[41] H. K. Khalil, "Nonlinear systems," Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[42] G. Carnevale and G. Notarstefano, "Nonconvex distributed optimization via Lasalle and singular perturbations," *IEEE Control Syst. Lett.*, vol. 7, pp. 301–306, 2022.

[43] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.*, Berlin, Germany: Springer, 2003.

[44] J. Nocedal and S. J. Wright, *Numerical Optimization.*, Berlin, Germany: Springer, 1999.

[45] V. Chellaboina and W. M. Haddad, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach.*, Princeton, NJ, USA: Princeton Univ. Press, 2008.

Guido Carnevale (Student Member, IEEE) received the M.Sc. degree "*summa cum laude*" in automation engineering from the University of Bologna, Bologna, Italy, in 2019, and the Ph.D. degree in biomedical, electrical, and systems engineering from the same university, in 2023.

He is a Junior Assistant Professor with the Department of Electrical, Electronic, and Information Engineering G. Marconi at Alma Mater Studiorum Università di Bologna. He was a Visiting Scholar with the University of Oxford in 2022. His research interests include distributed optimization and games over networks and optimal control.

Filippo Fabiani received the B.Sc. degree in bioengineering, the M.Sc. degree in automatic control engineering, and the Ph.D. degree in automatic control, all from the University of Pisa, Pisa, Italy, in 2012, 2015, and 2019, respectively.

He is an Assistant Professor with the IMT School for Advanced Studies Lucca, IT. In 2018–2019, he was Postdoctoral Research Fellow with the Delft Center for Systems and Control, TU Delft, The Netherlands, while in 2019–2022 he was a postdoctoral Research Assistant in the Control Group, Department of Engineering Science, University of Oxford, U.K. 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.

Kostas Margellos (Member, IEEE) received the Diploma degree in electrical engineering from the University of Patras, Patras, Greece, in 2008, and the Ph.D. degree in control engineering from ETH Zurich, Zurich, Switzerland, in 2012.

He spent 2013, 2014, and 2015 as a Postdoctoral Researcher with 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 and Machine Learning at Reuben College and a Lecturer at Worcester College. He is currently an 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.

Filiberto Fele received the M.Sc. degree in mechatronics engineering from Politecnico di Torino, Turin, Italy, in 2008, and the Ph.D. degree in control engineering from the University of Seville, Seville, Spain, in 2017.

He is Ramón y Cajal Research Fellow with the Department of Systems Engineering and Automation, University of Seville, Spain. He held postdoctoral appointments with the Electrical and Electronic Engineering Department, Imperial College London, London, U.K., in 2017, and the Department of Engineering Science, University of Oxford, 2018–2022, where he was Research Associate with the Energy and Power Group and the Control Group. From 2019 to 2022, he was a Lecturer with Christ Church and St Edmund Hall Colleges, Oxford. He is part of the EUCA Conference Editorial Board. His interests include data-driven robust optimization and control of multiagent systems, with a focus on the smart grid and the transition to net zero.

Giuseppe Notarstefano (Member, IEEE) received the Laurea degree *(summa cum laude)* in electronics engineering from the Università di Pisa, Pisa, Italy, in 2003 and the Ph.D. degree in automation and operation research from the Università di Padova, 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 an Associate Professor (from 2016 to 2018) and previously an Assistant Professor, Ricercatore (from 2007), with the Università del Salento, Italy. He has been a Visiting Scholar with 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 is 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.