Optimal Selection and Tracking of Generalized Nash Equilibria in Monotone Games

Benenati, Emilio; Ananduta, Wicak; Grammatico, Sergio

DOI
10.1109/TAC.2023.3288372

Publication date
2023

Document Version
Final published version

Published in
IEEE Transactions on Automatic Control

Optimal Selection and Tracking Of Generalized Nash Equilibria in Monotone Games

Emilio Benenati, Wicak Ananduta, and Sergio Grammatico

***Abstract*—A fundamental open problem in monotone game theory is the computation of a specific generalized Nash equilibrium (GNE) among all the available ones, e.g., the optimal equilibrium with respect to a system-level objective. The existing GNE seeking algorithms have in fact convergence guarantees toward an arbitrary, possibly inefficient, equilibrium. In this article, we solve this open problem by leveraging results from fixed-point selection theory and in turn derive distributed algorithms for the computation of an optimal GNE in monotone games. We then extend the technical results to the time-varying setting and propose an algorithm that tracks the sequence of optimal equilibria up to an asymptotic error, whose bound depends on the local computational capabilities of the agents.**

***Index Terms*—Multiagent systems, Nash equilibrium seeking, optimization.**

I. Introduction

Motivation: Numerous engineering systems of recent interest, such as smart electrical grids [1], [2], traffic control systems [3], and wireless communication systems [4], [5], [6] can be modeled as a *generalized game*, i.e., a system of multiple agents aiming at optimizing their individual, interdependent objectives, while satisfying some common constraints. A typical operating point for these systems is the generalized Nash equilibrium (GNE), where no agent can unilaterally improve their objective function [7].

The recent literature has witnessed the development of theory and algorithms for computing a *variational* GNE (v-GNE) [7], [8], [9], which exhibits desirable properties of fairness and stability. *Semidecentralized* GNE seeking algorithms, where a reliable central coordinator gathers and broadcasts aggregate information, have been proposed for strongly monotone [10], [11] and merely monotone games [12], [13], [14]. A breakthrough idea in [15], later generalized for nonstrongly monotone games [16], [17], [18], enables a *distributed* computation of GNEs by exploiting a suitable consensus protocol [19], thus requiring a peer-to-peer information exchange.

Existing results present, however, two fundamental shortcomings that might limit their practical application. First, unless strong assumptions are considered (namely, strong monotonicity of the pseudogradient), a game may have infinitely many v-GNEs *and virtually all the existing algorithms provide no characterization of the equilibrium computed.* For instance, a Nash equilibrium can be arbitrarily inefficient with respect to system-level efficiency metrics (e.g., overall social cost) [20]. Such uncertainty on the obtained equilibrium is often unacceptable. A notable exception is the class of double-layer Tikhonov regularization algorithms, [12], [21], [22]. While the method in [12] works for generalized games, it only guarantees convergence to the minimum-norm solution. On the other hand, the equilibrium selection algorithms in [21] and [22] solve at each (outer) iteration a regularized subproblem where the objectives of the agents are augmented with a convex selection function to be optimized, weighted by a small parameter. However, the latter are only applicable to nongeneralized games. In addition, the double-layer method in [23] and [24] seeks the GNE closest to a desired strategy. It is important to note that double-layer algorithms require the exact solution of a subproblem at each (outer) iteration, and thus they would require a virtually infinite amount of communications per outer iteration in a distributed setting. Recently, a single-layer algorithm based on a regularized projected-pseudogradient dynamics was proposed in [25], which however is only suitable for nongeneralized games and requires nested vanishing stepsizes both on the pseudogradient and on the regularization. Second, decision-making agents often operate in a time-dependent environment and, due to the limited computation capabilities and to the time required to exchange information, it can be impossible to ensure a time-scale separation between the environment and the algorithm dynamics. This results in nonconstant objectives and constraints between the discrete-time algorithmic iterations, as discussed in [26], and the references therein, for the particular case of optimization problems. Only few works, e.g., [27], consider this setting in the case of game equilibrium problems and only with a strong monotonicity assumption on the game pseudogradient mapping.

Optimal equilibrium selection and tracking: We can formulate the first issue, identified in the seminal work [7, Sec. 6], as an *optimal GNE selection* problem, i.e., the problem of computing a GNE of a game (among the potentially infinitely many) that satisfies a selection criterion. This criterion characterizes the desired equilibrium and can be formalized as a system-level

Manuscript received 15 March 2022; revised 14 February 2023; accepted 4 June 2023. Date of publication 21 June 2023; date of current version 5 December 2023. This work was supported in part by the European Research Council under research project COSMOS under Grant 802348. Recommended by Associate Editor U. Topcu. *(Corresponding author: Emilio Benenati.)*

The authors are with the Delft Center of Systems and Control (DCSC), TU Delft, 2624CD Delft, Netherlands (e-mail: e.benenati@tudelft.nl; w.ananduta@tudelft.nl; s.grammatico.tudelft.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2023.3288372.
Digital Object Identifier 10.1109/TAC.2023.3288372

selection function to be optimized over the set of GNEs. For example, the system-level objective of an electricity market can be to minimize the deviation from an efficient operating set-point [26]; for multiple autonomous vehicles, it can be to minimize the overall travel time of the network. Meanwhile, the second issue can be cast as an *optimal GNE tracking* problem, i.e., the problem of tracking the sequence of optimal GNEs of a time-varying game, with finite computation time and limited information on the future instances of the game available. As the GNE set is in general not a singleton, the tracking objective should be again chosen by means of a (time-varying) selection function. These problems, although of high practical interest, have never been addressed in the literature.

Under mild assumptions on the selection function, the optimal GNE selection problem in a monotone game is a special case of a Variational Inequality (VI) [28] defined over the set of v-GNEs. On the other hand, as shown in [13], [14], and [17], operator splitting techniques [29] can be leveraged to characterize v-GNEs as the zeros of a monotone operator and, in turn, as the fixed-point set of a suitable operator. Therefore, here we can cast the problem as that of fixed-point selection [30]. In the literature, e.g., [30], [31], [32], the latter can be solved by the hybrid steepest descent method (HSDM), whose iterations depend on the fixed-point operator (whose definition depends on the primitives of the game) and the monotone operator that defines the VI, namely the gradient of the selection function in our setting.

The contributions: In the first part of the article (Sections III and IV), we propose the first single-layer distributed algorithms for solving the optimal GNE selection problem. Our method employs the forward-backward-forward (FBF) operator [17] combined with the HSDM. We show that the proposed algorithm guarantees convergence to the *optimal* v-GNE set in monotone games. Moreover, for a special class of monotone games, namely cocoercive games with affine coupling constraints, we also show that the preconditioned forward-backward (pFB) [13] can be paired with the HSDM to derive optimal GNE selection algorithms. Technically, our contribution is to show that these operators fulfill special properties that guarantee the convergence of the HSDM toward the solution set of the corresponding fixed-point selection VI. Compared to [12], [23], and [24], our proposed algorithms significantly generalize the class of selection functions; additionally, our method works for generalized games and does not require solving an equilibrium problem at each iteration nor a vanishing stepsize on the pseudogradient dynamics.

In the second part of the article (Section V), we formalize the optimal GNE tracking problem as a time-varying fixed-point selection problem. Thus, as a solution framework, we propose the *restarted HSDM*, which adapts its operators when the problem changes. In line with the results in the time-varying optimization literature [33], we show convergence up to a tracking error which depends on the problem data and can be controlled by a suitable tuning of the algorithm parameters.

II. Mathematical Preliminaries

Notation: The set of real numbers is denoted by $\mathbb{R}$. The vector of all 1 (or 0) with dimension n are denoted by $\mathbf{1}_n$ ($\mathbf{0}_n$). We omit the subscript when the dimension is clear from the context. The operator $\mathrm{col}(\cdot)$ stacks the arguments column-wise. For a group of vectors x_i, $i \in \mathcal{I} = \{1, 2, \dots, N\}$, we use the bold symbol to denote their column concatenation, i.e., $\boldsymbol{x} := \mathrm{col}((x_i)_{i\in\mathcal{I}})$. The cardinality of a set is denoted by $|\cdot|$. The operator $\langle x, y\rangle$ denotes the inner product. We denote by $\|\cdot\|$ the Euclidean norm. Let $P \succ 0$ be symmetric. For $x, y \in \mathbb{R}^n$, $\langle x, y\rangle_P := \langle x, Py\rangle$ and $\|x\|_P := \sqrt{\langle x, x\rangle_P}$ denote the P-weighted Euclidean inner product and the P-weighted Euclidean norm, respectively. The graph of an operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is denoted by $\mathrm{gph}(A)$. $\mathrm{zer}(A)$ defines the set of zeros of operator A, i.e.,

$$\mathrm{zer}(A) := \{x \in \mathrm{dom}(A) \mid 0 \in A(x)\}$$

whereas $\mathrm{fix}(A)$ is the set of fixed points of $A : \mathbb{R}^n \to \mathbb{R}^n$, i.e.,

$$\mathrm{fix}(A) := \{x \in \mathrm{dom}(A) \mid A(x) = x\}.$$

Convex functions: A continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is σ-strongly convex with respect to a Ψ-weighted norm, with $\sigma > 0$ and $\Psi \succ 0$, if, for all $x, x' \in \mathrm{dom}\, f$

$$f(x') \geq f(x) + \langle \nabla f(x), x' - x\rangle_\Psi + \tfrac{\sigma}{2}\|x' - x\|_\Psi^2.$$

Additionally, f is convex if the previous inequality holds for $\sigma = 0$. The projection onto a closed convex set C is denoted by $\mathrm{proj}_C^\Psi(x) = \mathrm{argmin}_{z\in C}\, \|x - z\|_\Psi$, where $\Psi \succ 0$. For a convex function f with subdifferential ∂f and $\Psi \succ 0$, we define the operator [29, Def. 12.23]

$$\mathrm{prox}_{\partial f}^\Psi(x) := \mathrm{argmin}_z\, f(z) + \tfrac{1}{2}\|z - x\|_\Psi^2.$$

For example, for the indicator function of a closed convex set C, ι_C, with $\partial\iota_C = \mathrm{N}_C$ being the normal cone operator, $\mathrm{prox}_{\iota_C}^\Psi(x) = \mathrm{proj}_C^\Psi(x)$ [29, Ex. 1.25, 16.13, 12.25].

Operator theoretic definitions: An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone* [29, Def. 20.1] if

$$\langle y - y', x - x'\rangle \geq 0, \quad \forall (x, y), (x', y') \in \mathrm{gph}(A)$$

and β-*strongly monotone* if $A - \beta\mathrm{Id}$, where Id is the identity operator, is monotone. Let C be a nonempty subset of $\mathbb{R}^n$. A single-valued operator $\mathcal{T} : C \to \mathbb{R}^n$ is *Lipschitz continuous* [29, Def. 1.47] if there exists a constant $L > 0$, such that

$$\|\mathcal{T}(x) - \mathcal{T}(x')\| \leq L\|x - x'\|, \ \forall x, x' \in \mathrm{dom}(\mathcal{T}).$$

In particular, the operator $\mathcal{T}$ is

1) *nonexpansive* if $L = 1$;
2) *attracting nonexpansive* if $\mathcal{T}$ is nonexpansive with $\mathrm{fix}(\mathcal{T}) \neq \varnothing$ and $\|\mathcal{T}(x) - z\| < \|x - z\|, \forall z \in \mathrm{fix}(\mathcal{T})$ and $\forall x \notin \mathrm{fix}(\mathcal{T})$;
3) *quasi-nonexpansive* if $\mathrm{fix}(\mathcal{T}) \neq \varnothing$ and $\|\mathcal{T}(x) - z\| \leq \|x - z\|, \forall z \in \mathrm{fix}(\mathcal{T})$ and $\forall x \in \mathbb{R}^n$.

Moreover, $\mathcal{T}$ is α-averaged nonexpansive, for $\alpha \in (0, 1)$, if there exists a nonexpansive operator $\mathcal{R} : C \to \mathbb{R}^n$ such that $\mathcal{T} = (1 - \alpha)\mathrm{Id} + \alpha\mathcal{R}$. If $\mathcal{T}$ is averaged nonexpansive with $\mathrm{fix}(\mathcal{T}) \neq \varnothing$, then $\mathcal{T}$ is attracting [30, Sec. 2.A]. Additionally $\mathcal{T}$ is β-*cocoercive* if

$$\langle \mathcal{T}(x) - \mathcal{T}(y), x - y\rangle \geq \beta\|\mathcal{T}(x) - \mathcal{T}(y)\|.$$

Now, let C be a nonempty, closed, and convex subset of $\mathbb{R}^n$, $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ be quasi-nonexpansive under the Ψ-induced norm $\|\cdot\|_\Psi$ for some positive definite matrix Ψ, i.e., $\|\mathcal{T}(x) -$

$z\|_\Psi \leq \|x - z\|_\Psi$, for all $z \in \text{fix}(\mathcal{T}) \neq \varnothing$ and $x \in \mathbb{R}^n$. We define the distance of a point $x \in \mathbb{R}^n$ to C by

$$\text{dist}_\Psi(x, C) := \inf_{z\in C}\|x - z\|_\Psi.$$

For $r \geq 0$, we define the set

$$C^\Psi_{\geq r} := \{x \in \mathbb{R}^n \mid \text{dist}_\Psi(x, C) \geq r\}. \quad (1)$$

By considering $A := \{r \in \mathbb{R}_{\geq 0} \mid \text{fix}(\mathcal{T})^\Psi_{\geq r} \cap C \neq \varnothing\}$, we define the the shrinkage function for the operator $\mathcal{T}$ under the norm $\|\cdot\|_\Psi$, which slightly generalizes [Def. 1], as follows:

$$D_\Psi(r) := \iota_A(r) + \inf_{x\in\text{fix}(\mathcal{T})^\Psi_{\geq r}\cap C}\text{dist}_\Psi(x, \text{fix}(\mathcal{T})) - \text{dist}_\Psi(\mathcal{T}(x), \text{fix}(\mathcal{T})). \quad (2)$$

For $\Psi = I$, we omit the subscript of D. The function D_Ψ has the properties stated next in Proposition 1 (see [34, Prop. 2.6] for the case $\Psi = I$).

Proposition 1: Let Ψ be positive definite. For the function D_Ψ defined in (2), it holds that

1) D_Ψ is positive semidefinite and nondecreasing;
2) $D_\Psi(\text{dist}(x, \text{fix}(\mathcal{T}))) \leq \|x - \mathcal{T}(x)\|_\Psi$ for all $x \in C$. □

Definition 1 (Quasi-shrinking [30]): A quasi-nonexpansive operator $\mathcal{T} : \mathbb{R}^n \to \mathbb{R}^n$ is *quasi-shrinking* on a nonempty, closed, and convex set $C \subseteq \mathbb{R}^n$ if $\text{fix}(\mathcal{T}) \cap C \neq \varnothing$ and $D(r) = 0 \Leftrightarrow r = 0$, where $D(r)$ is defined as in (2). □

Remark 1: Suppose that a quasi-nonexpansive operator $\mathcal{T}$ is quasi-shrinking on C, i.e., $D(r) = 0 \Leftrightarrow r = 0$. Then, it also holds that $D_\Psi(r) = 0 \Leftrightarrow r = 0$, for any $\Psi \succ 0$. □

Example 1: The Euclidean projection onto C, proj_C is quasi-shrinking and its shrinkage function (defined in (2)) is

$$D(r) = \inf_{\{u|\text{dist}(u,C)\geq r\}}\text{dist}(u, C) - \underbrace{\text{dist}(\text{proj}_C(u), C)}_{=0} = r.$$

□

Finally, we identify a class of quasi-shrinking operators, as formally stated in Lemma 1, which generalizes the result in [34, Prop. 2.11] and is useful for our analysis.

Definition 2 (Demiclosed operator [29, Def. 4.26]): Let $C \subseteq \mathbb{R}^n$ be a closed set. An operator $\mathcal{T} : C \to \mathbb{R}^n$ is demiclosed at $u \in \mathbb{R}^n$ if $\mathcal{T}(\boldsymbol{\omega}^\infty) = u$, for any sequence $(\boldsymbol{\omega}_k)_{k\in\mathbb{N}} \in C$ such that $\lim_{k\to\infty} \boldsymbol{\omega}_k = \boldsymbol{\omega}^\infty$ and $\lim_{k\to\infty} \mathcal{T}(\boldsymbol{\omega}_k) = u$. □

Lemma 1: Let $\mathcal{T}$ be an operator with $\text{fix}(\mathcal{T}) \neq \varnothing$. Let $\mathcal{T}_2$ be an operator such that $\text{Id} - \mathcal{T}_2$ is demiclosed at 0 and such that $\text{fix}(\mathcal{T}_2) \subseteq \text{fix}(\mathcal{T})$. Assume that for any $\boldsymbol{\omega}^\star \in \text{fix}(\mathcal{T})$,

$$\|\mathcal{T}(\boldsymbol{\omega}) - \boldsymbol{\omega}^\star\|^2_\Psi \leq \|\boldsymbol{\omega} - \boldsymbol{\omega}^\star\|^2_\Psi - \gamma\|\boldsymbol{\omega} - \mathcal{T}_2(\boldsymbol{\omega})\|^2_\Psi \quad (3)$$

for some $\gamma > 0$ and $\Psi \succ 0$. Then, $\mathcal{T}$ is quasi-shrinking on any compact convex set C such that $C \bigcap \text{fix}(\mathcal{T}) \neq \varnothing$. □

III. Optimal Selection of Generalized Nash Equilibria

A. Generalized Nash Equilibrium Problem

Let us consider N agents, denoted by the set $\mathcal{I} := \{1, 2, \ldots, N\}$, with inter-dependent optimization problems:

$$\forall i \in \mathcal{I} : \begin{cases} \min\limits_{x_i\in\mathcal{X}_i} & J_i(\boldsymbol{x}) := \ell_i(x_i) + f_i(\boldsymbol{x}) \quad (4a) \\ \text{s.t.} & \sum\limits_{j\in\mathcal{I}} g_j(x_j) \leq 0 \quad (4b) \end{cases}$$

where $x_i \in \mathbb{R}^{n_i}$ is the decision variable of agent i whereas $\boldsymbol{x} := \text{col}((x_i)_{i\in\mathcal{I}}) \in \mathbb{R}^n$, with $n = \sum_{i\in\mathcal{I}} n_i$, is a concatenated vector of the decision variables of all agents. Let us use $\boldsymbol{x}_{-i} = \text{col}((x_j)_{j\in\mathcal{I}\setminus\{i\}})$ to denote the concatenated decision variables of all agents except agent i. Let $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ denote the local feasible set of x_i and $J_i : \mathbb{R}^n \to \mathbb{R}$ denote the cost function of agent i that depends on the decision variables of other agents. Moreover, (4b) represents a separable coupling constraint where $g_j : \mathbb{R}^{n_j} \to \mathbb{R}^m$ is associated with agent j. We denote the collective feasible set of the game in (4) by

$$\boldsymbol{\Omega} := \prod_{i\in\mathcal{I}} \mathcal{X}_i \cap \left\{\boldsymbol{x} \mid \textstyle\sum_{j\in\mathcal{I}} g_j(x_j) \leq 0\right\}. \quad (5)$$

Here, we look for equilibrium solutions to (4) where no agent has the incentive to unilaterally deviate, namely, GNE

Definition 3: A set of strategies $\boldsymbol{x}^* := \text{col}((x^*_i)_{i\in\mathcal{I}})$ is a GNE of the game in (4) if $\boldsymbol{x}^* \in \boldsymbol{\Omega}$ and, for each $i \in \mathcal{I}$

$$J_i(\boldsymbol{x}^*) \leq J_i(x_i, \boldsymbol{x}^*_{-i}) \quad (6)$$

for any $x_i \in \mathcal{X}_i \cap \{y \mid g_i(y) \leq -\sum_{j\in\mathcal{I}\setminus\{i\}} g_j(x^*_j)\}$. □

Furthermore, we focus on the class of jointly convex GNE problems and hence, consider the following assumptions on Problem (4) [13, Assumptions 1–2]. We note that [14], [15], [16], [17], and[18] consider the case of affine constraint functions.

Assumption 1: In (4), for each $i \in \mathcal{I}$, the functions $f_i(\cdot, \boldsymbol{x}_{-i})$, for any $\boldsymbol{x}_{-i}$, and $g_i(\cdot)$ are component-wise convex and continuously differentiable; ℓ_i is convex and lower semicontinuous. For each $i \in \mathcal{I}$, the set $\mathcal{X}_i$ is nonempty, compact, and convex. The global feasible set $\boldsymbol{\Omega}$ defined in (5) is nonempty and satisfies Slater's constraint qualification [29, Eq. (27.50)]. □

Assumption 2: The mapping

$$F(\boldsymbol{x}) := \text{col}((\nabla_{x_i} f_i(\boldsymbol{x}))_{i\in\mathcal{I}}) \quad (7)$$

with $(f_i)_{i\in\mathcal{I}}$ as in (4a), is monotone. □

As in [13], [14], [15], [16], [17], and [18], we can formulate the problem of finding a GNE of the game in (4) as that of a monotone inclusion. To this end, we introduce the dual variable $\lambda_i \in \mathbb{R}^m$, for each $i \in \mathcal{I}$, to be associated with the coupling constraint (4b). Furthermore, we focus on a subset of GNEs, namely variational GNE (v-GNE), indicated by equal optimal dual variables, $\lambda^*_i = \lambda^*$, for all $i \in \mathcal{I}$. As discussed in [7] and [9], a v-GNE enjoys several desirable properties, such as fairness and larger social stability than nonvariational ones. Under Assumptions 1–2, the set of v-GNEs of the game in (4) is nonempty [35, Prop. 12.11]. The Karush–Kuhn–Tucker (KKT) optimality conditions of a v-GNE of the game in (4), denoted by $\boldsymbol{x}^*$, are

$$\forall i \in \mathcal{I} : \begin{cases} \mathbf{0} \in \text{N}_{\mathcal{X}_i}(x^*_i) + \partial_{x_i} J_i(\boldsymbol{x}^*) + \langle \nabla g_i(x^*_i), \lambda^* \rangle, & (8a) \\ \mathbf{0} \in \text{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*) - \sum\limits_{j\in\mathcal{I}} g_j(x^*_j). & (8b) \end{cases}$$

To obtain a v-GNE via a fully distributed algorithm, we incorporate a consensus scheme on the dual variables. In the full information case, one typically assumes that there exists a communication network over which the agents exchange information to update their dual variables. Let us represent this communication network as an undirected graph $\mathcal{G}^\lambda = (\mathcal{I}, \mathcal{E}^\lambda)$ and assume that $\mathcal{G}^\lambda$ is connected. Furthermore, we denote the

Laplacian of $\mathcal{G}^\lambda$ by $\mathcal{L}$ and the neighbors of agent i in $\mathcal{G}^\lambda$ by $\mathcal{N}_i^\lambda$, i.e., $\mathcal{N}_i^\lambda := \{j \in \mathcal{I} \mid (i,j) \in \mathcal{E}^\lambda\}$. Additionally, let $\mathcal{N}_i^J$ denote the set of agents whose decision variable x_j influences the cost function J_i. For simplicity, we assume that $\mathcal{N}_i^J \subseteq \mathcal{N}_i^\lambda$.

Now, let us denote $\nu_i \in \mathbb{R}^m$ as the consensus variable of agent i, and $\boldsymbol{\omega} = (\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \in \mathbb{R}^{n_\omega}$, where $\boldsymbol{\lambda} = \mathrm{col}((\lambda_i)_{i\in\mathcal{I}})$, $\boldsymbol{\nu} = \mathrm{col}((\nu_i)_{i\in\mathcal{I}})$, and $n_\omega = n + 2Nm$. Then, we can define the operators $\mathcal{A}, \mathcal{B}, \mathcal{C} : \mathbb{R}^{n_\omega} \to \mathbb{R}^{n_\omega}$, as follows:

$$\mathcal{A}(\boldsymbol{\omega}) := \prod_{i\in\mathcal{I}} (\mathrm{N}_{\mathcal{X}_i} + \partial \ell_i)(x_i) \times \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\boldsymbol{\lambda}) \times \{\mathbf{0}_{Nm}\}, \tag{9}$$

$$\mathcal{B}(\boldsymbol{\omega}) := \mathrm{col}(F(\boldsymbol{x}), (\mathcal{L} \otimes I_m)\boldsymbol{\lambda}, \mathbf{0}_{Nm}), \tag{10}$$

$$\mathcal{C}(\boldsymbol{\omega}) := \mathrm{col}((\langle \nabla g_i(x_i), \lambda_i \rangle)_{i\in\mathcal{I}}, -(g_i(x_i))_{i\in\mathcal{I}} - (\mathcal{L} \otimes I_m)\boldsymbol{\nu}, (\mathcal{L} \otimes I_m)\boldsymbol{\lambda}). \tag{11}$$

We then cast the GNEP in (4) as the inclusion problem:

$$\text{find } \boldsymbol{\omega} \text{ such that } \boldsymbol{\omega} \in \mathrm{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C}). \tag{12}$$

Similarly to [15, Th. 2], we can show that for any $\boldsymbol{\omega}$ such that (12) holds, we obtain the pair $(\boldsymbol{x}, \lambda)$ that satisfies the KKT conditions in (8) if Assumptions 1–2 hold (see Appendix B for details). The zero set of $\mathcal{A} + \mathcal{B} + \mathcal{C}$ is convex following its maximal monotonicity (Lemma 4 in Appendix B) and [29, Prop. 23.39]. This result generalizes the known convexity of the solution set to a convex optimization problem [29, Prop. 11.6], which in our case is recovered by setting $f_i \equiv 0$ for all i. Additionally, as $\boldsymbol{\mathcal{X}}$ is bounded, the set of solutions of (8) and the set $\mathrm{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C})$ are bounded [36, Prop. 3.3].

B. Optimal Equilibrium Selection Problem

The inclusion problem in (12) may have multiple solutions. In this section, we want to find an equilibrium solution that minimizes a selection function, denoted by $\phi : \mathbb{R}^{n_\omega} \to \mathbb{R}$, i.e.,

$$\begin{cases} \underset{\boldsymbol{\omega}}{\mathrm{argmin}} & \phi(\boldsymbol{\omega}) \\ \quad \mathrm{s.\,t.} & \boldsymbol{\omega} \in \mathrm{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C}). \end{cases} \tag{13}$$

For example, we can consider the selection function

$$\phi_{\mathrm{ex}}(\boldsymbol{\omega}) = \|Q\boldsymbol{\omega} - \boldsymbol{\omega}^{\mathrm{ref}}\| \tag{14}$$

for some $Q \succcurlyeq 0$. When $Q = I$ and $\omega^{\mathrm{ref}} = \mathbf{0}$, the objective is to find a minimum norm v-GNE. The vector ω^{ref} can be any desired strategy of the agents, and thus the objective is to find the v-GNE closest to this strategy, as discussed in [23] and [24]. In some engineering applications, such as electrical networks, (14) can represent system level objectives (see Section VI). In the remainder of the article, we consider the following technical assumption on the selection function, which, together with the convexity of $\mathrm{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C})$, guarantees that the optimization problem in (13) is convex.

Assumption 3: The function ϕ in (13) is continuously differentiable, convex, and has L_ϕ-Lipschitz continuous gradient. □

To compute an optimal variational GNE, we first derive operators with the property

$$\boldsymbol{\omega} \in \mathrm{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C}) \Leftrightarrow \boldsymbol{\omega} \in \mathrm{fix}(\mathcal{T}) \tag{15}$$

and such that the Banach-Picard iteration of $\mathcal{T}$ [29, Sec. 5.2] guarantees convergence to a solution of the inclusion in (12). For instance, for cocoercive generalized games, a pFB operator presents the desired characteristics [15], whereas the forward-reflected-backward (FRB) operator [37] or the FBF operator [38] meet these requirements even for general monotone games. Furthermore, we require that the operator $\mathcal{T}$ in (15) can be evaluated in a distributed manner. By (15) and Assumption 3, the optimal equilibrium selection problem in (13) can be cast as a fixed-point selection VI

$$\text{find } \boldsymbol{\omega}^\star \text{ s.t. } \inf_{\boldsymbol{\omega}\in\mathrm{fix}(\mathcal{T})} \langle \boldsymbol{\omega} - \boldsymbol{\omega}^\star, \nabla\phi(\boldsymbol{\omega}^\star) \rangle \geq 0. \tag{16}$$

C. Distributed Optimal Equilibrium Selection Algorithm

With the aim of solving the VI in (16), we consider the HSDM algorithm [30], which is defined by the following iteration:

$$\boldsymbol{\omega}^{(k+1)} = \mathcal{T}(\boldsymbol{\omega}^{(k)}) - \beta^{(k)} \nabla\phi(\mathcal{T}(\boldsymbol{\omega}^{(k)})). \tag{17}$$

The HSDM can solve Problem (16) when $\mathcal{T}$ is quasi-nonexpansive and quasi-shrinking with bounded $\mathrm{fix}(\mathcal{T})$, as formally stated next.

Assumption 4: The step size $\beta^{(k)}$ satisfies the following:
1) $\lim_{k\to\infty} \beta^{(k)} = 0$, $\sum_{k\geq 1} \beta^{(k)} = \infty$;
2) $\sum_{k\geq 1} (\beta^{(k)})^2 < \infty$. □

Remark 2: The sequence $\beta^{(k)} = \beta_0 / k^p$, for any $\beta_0 > 0$ and $p \in (1/2, 1]$, satisfies Assumption 4. □

Assumption 5: $\mathcal{T}$ is quasi-shrinking on a nonempty compact convex set C. □

Lemma 2 (From [30, Th. 5]): Let Assumption 3 hold and $\Omega^\star$ be the set of solutions of the VI in (16), with nonempty and bounded $\mathrm{fix}(\mathcal{T})$. Suppose that $\mathcal{T}$ satisfies Assumptions 5 with compact convex set C such that $(\boldsymbol{\omega}^{(k)})_{k\geq 0} \subset C$. If the step size $\beta^{(k)}$ satisfies Assumption 4.i, then the HSDM in (17) generates a sequence $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \mathrm{dist}(\boldsymbol{\omega}^{(k)}, \Omega^\star) = 0.$$

□

Therefore, our main technical task is to find a suitable operator $\mathcal{T}$ that can be evaluated in a distributed manner and that satisfies both (15) and Assumption 5.

Under mere monotonicity of the pseudogradient mapping (Assumption 2), perhaps the most obvious choice is the FRB splitting, which, however, is not quasi-nonexpansive[1] (and, thus, it is not quasi-shrinking). Another viable option is the FBF splitting method [38], which works for v-GNE seeking in monotone games satisfying Assumptions 1–2, as shown in [13] and [17]. As our first technical result, we show that the FBF algorithm satisfies both the desired property in (15) and Assumptions 5.

[1] FRB iteration does not generate a Fejér monotone sequence [37, Prop. 2.3], implying that it is not quasi-nonxepansive and violates Definition 1.

The FBF operator for (12) reads as follows:

$$\mathcal{T}_{\text{FBF}}(\boldsymbol{\omega}) := ((\text{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C}))(\text{Id} + \Psi^{-1}\mathcal{A})^{-1} \cdot (\text{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C})) + \Psi^{-1}(\mathcal{B}+\mathcal{C}))(\boldsymbol{\omega}) \tag{18}$$

where $\Psi \succ 0$ is a diagonal positive definite matrix. The FBF requires the forward operator, which is $(\mathcal{B}+\mathcal{C})$, to be Lipschitz continuous. A sufficient condition for this requirement is given in Assumption 6 (see Lemma 5 in Appendix B).

Assumption 6: The mapping $F(\boldsymbol{x})$ in (7) is L_F-Lipschitz continuous. Furthermore, for each $i \in \mathcal{I}$, the function g_i in (4b) has a bounded and $L_{\nabla g}$-Lipschitz continuous gradient. □

Under maximal monotonicity and Lipschitz continuity, it holds that $\text{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C}) = \text{fix}(\mathcal{T}_{\text{FBF}})$ (see Lemma 6 in Appendix C). In addition, we define the step-size matrix

$$\Psi := \text{diag}(\rho^{-1}, \tau^{-1}, \sigma^{-1}) \tag{19}$$

where $\rho = \text{diag}((\rho_i I_{n_i})_{i\in\mathcal{I}})$, $\tau = \text{diag}((\tau_i I_m)_{i\in\mathcal{I}})$, and $\sigma = \text{diag}((\sigma_i I_m)_{i\in\mathcal{I}})$ need to be small enough with respect to the Lipschitz constant of $\mathcal{B}+\mathcal{C}$. A sufficient condition on Ψ for the fixed-point iteration with $\mathcal{T}_{\text{FBF}}$ to converge is given in the following Assumption 7 [17, Assumption 2].

Assumption 7: It holds that $|\Psi^{-1}| \leq 1/L_B$, where $L_B > 0$ is the Lipschitz constant of $\mathcal{B}+\mathcal{C}$ and Ψ reads as in (19). □

We are now ready to present the distributed FBF for seeking an optimal variational GNE based on the selection function $\phi(\boldsymbol{\omega})$ via the HSDM, as shown in Algorithm 1. To have a convergence guarantee as stated in Lemma 2, the FBF operator must satisfy Assumption 5. This is shown in the following lemma.

Lemma 3: Let Assumptions 1, 2, 6, and 7 hold. The operator $\mathcal{T}_{\text{FBF}}$ in (18), where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are defined in (9)–(11) and Ψ is defined in (19), is quasi-shrinking on any compact convex set C such that $C \cap \text{fix}(\mathcal{T}_{\text{FBF}}) \neq \varnothing$. □

Thus, Algorithm 1 generates a sequence that converges to the solution set of the problem in (16), as stated next.

Theorem 1: Let Assumptions 1–4 and 6–7 hold. Let $\Omega^\star$ be the set of solutions to Problem (16) with $\mathcal{T} = \mathcal{T}_{\text{FBF}}$ defined in (18), where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are defined in (9)–(11). Let $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$, where $\boldsymbol{\omega}^{(k)} = (\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})$, be the sequence generated by Algorithm 1. Then, $\lim_{k\to\infty} \text{dist}(\boldsymbol{\omega}^{(k)}, \Omega^\star) = 0$, and $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converges to an optimal v-GNE of the game in (4). □

Remark 3: A central coordinator and step 5 of Algorithm 1 are not needed if ϕ is a separable function, i.e., $\phi(\boldsymbol{\omega}) = \sum_{i\in\mathcal{I}} \phi_i(\omega_i)$. In this case, step 6 can be immediately executed by using local information $(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)})$ only, as long as each agent i knows the gradient $\nabla\phi_i$. □

Algorithm 1: Optimal v-GNE Selection via FBF and HSDM.

Initialization: $x_i^{(0)} \in \mathcal{X}_i$, $\lambda_i^{(0)} \in \mathbb{R}^m_{\geq 0}$, and $\nu_i^{(0)} \in \mathbb{R}^m$, $\forall i \in \mathcal{I}$.

Iteration of each agent $i \in \mathcal{I}$**.**

1) Receives $x_j^{(k)}$ from agent $j \in \mathcal{N}_i^J$ and $\lambda_j^{(k)}, \nu_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.
2) Updates:

$$\tilde{x}_i^{(k)} = \text{prox}_{\ell_i+\iota_{\mathcal{X}_i}}^{\rho_i}\left(x_i^{(k)} - \rho_i\left(\nabla_{x_i} f_i(\boldsymbol{x}^{(k)}) + \nabla g_i(x_i^{(k)})^\top \lambda_i^{(k)}\right)\right),$$

$$\tilde{\lambda}_i^{(k)} = \text{proj}_{\geq 0}\left(\lambda_i^{(k)} + \tau_i\left(g_i(x_i^{(k)}) + \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\nu_i^{(k)} - \nu_j^{(k)} - \lambda_i^{(k)} + \lambda_j^{(k)}\right)\right)\right),$$

$$\tilde{\nu}_i^{(k)} = \nu_i^{(k)} - \sigma_i \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\lambda_i^{(k)} - \lambda_j^{(k)}\right).$$

3) Receives $\tilde{x}_j^{(k)}$ from agent $j \in \mathcal{N}_i^J$ and $\tilde{\lambda}_j^{(k)}, \tilde{\nu}_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.
4) Updates:

$$\mathring{x}_i^{(k)} = \tilde{x}_i^{(k)} - \rho_i\left(\nabla_{x_i} f_i(\tilde{\boldsymbol{x}}^{(k)}) - \nabla_{x_i} f_i(\boldsymbol{x}^{(k)}) + \nabla g_i(\tilde{x}_i^{(k)})^\top \tilde{\lambda}_i^{(k)} - \nabla g_i(x_i^{(k)})^\top \lambda_i^{(k)}\right),$$

$$\mathring{\lambda}_i^{(k)} = \tilde{\lambda}_i^{(k)} + \tau_i\left(g_i(\tilde{x}_i^{(k)}) - g_i(x_i^{(k)}) + \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\tilde{\nu}_i^{(k)} - \nu_i^{(k)} - \tilde{\nu}_j^{(k)} + \nu_j^{(k)}\right) - \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\tilde{\lambda}_i^{(k)} - \lambda_i^{(k)} - \tilde{\lambda}_j^{(k)} + \lambda_j^{(k)}\right)\right),$$

$$\mathring{\nu}_i^{(k)} = \tilde{\nu}_i^{(k)} - \sigma_i \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\tilde{\lambda}_i^{(k)} - \lambda_i^{(k)} - \tilde{\lambda}_j^{(k)} + \lambda_j^{(k)}\right).$$

5) Sends $(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)})$ to a coordinator and receives back $\nabla_{\omega_i}\phi(\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)})$, where $\omega_i = (x_i, \lambda_i, \nu_i)$.
6) Updates:

$$(x_i^{(k+1)}, \lambda_i^{(k+1)}, \nu_i^{(k+1)}) = (\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)}) - \beta^{(k)}\nabla_{\omega_i}\phi(\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)}). \tag{20}$$

IV. Optimal Equilibrium Selection in Cocoercive Games

In this section, we discuss a special class of monotone games, namely cocoercive games with affine coupling constraints, characterized by the following Assumptions 8 and 9 constraints. These games arise as a generalization of the widely studied class of strongly monotone games [15], [11]. Differently from the strong monotonicity assumption, however, cocoercivity alone does not guarantee uniqueness of the v-GNE.

Assumption 8 ([14, Assumption 5]): The mapping F in (7) is η-cocoercive. □

Assumption 9 ([14, Eq. (3)]): For each $i \in \mathcal{I}$, the function g_i in (4b) is affine, i.e., $g_i(x_i) := A_i x_i - b_i$, for some matrix $A_i \in \mathbb{R}^{m\times n_i}$ and vector $b_i \in \mathbb{R}^m$. □

For this particular class of games, the pFB splitting [15] can efficiently compute a variational GNE. We note that, although [15] considers games with strongly monotone pseudogradient, the FB splitting only requires cocoercivity of the forward operator[29, Th. 26.14]. Compared with the FBF, the pFB has the advantages of only having one communication round per iteration (as opposed to two) and larger step size bounds. A numerical performance comparison was provided in [17]. Given the particular structure of the coupling constraint as stated in Assumption 9, we can rewrite the operators in (12) as follows:

$$\mathcal{A}(\boldsymbol{\omega}) := \prod_{i\in\mathcal{I}}(\mathrm{N}_{\mathcal{X}_i}+\partial\ell_i)(x_i) \times \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\boldsymbol{\lambda}) \times \{\mathbf{0}_{Nm}\}, \quad (21)$$

$$\mathcal{B}(\boldsymbol{\omega}) := \mathrm{col}(F(\boldsymbol{x}), (\mathcal{L}\otimes I_m)\boldsymbol{\lambda} + \boldsymbol{b}, \mathbf{0}_{Nm}), \quad (22)$$

$$\mathcal{C}(\boldsymbol{\omega}) := \mathrm{col}(\boldsymbol{A}^\top\boldsymbol{\lambda}, -\boldsymbol{A}\boldsymbol{x} - (\mathcal{L}\otimes I_m)\boldsymbol{\nu}, (\mathcal{L}\otimes I_m)\boldsymbol{\lambda}) \quad (23)$$

where $\boldsymbol{A} = \mathrm{diag}((A_i)_{i\in\mathcal{I}})$ and $\boldsymbol{b} = \mathrm{col}((b_i)_{i\in\mathcal{I}})$. Thus, the pFB operator for the monotone inclusion in (12) based on the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in (21)–(23) is given by [15, Eq. (24)]

$$\mathcal{T}_{\text{pFB}}(\boldsymbol{\omega}) := (\mathrm{Id} + \Phi^{-1}(\mathcal{A}+\mathcal{C}))^{-1}(\mathrm{Id} - \Phi^{-1}\mathcal{B})(\boldsymbol{\omega}) \quad (24)$$

where $\Phi \succ 0$ is a preconditioning matrix, defined as

$$\Phi := \Psi + \begin{bmatrix} 0 & -\boldsymbol{A}^\top & 0 \\ -\boldsymbol{A} & 0 & -\mathcal{L}\otimes I_m \\ 0 & -\mathcal{L}\otimes I_m & 0 \end{bmatrix}$$

where Ψ is as in (19). Then, we can have an extension of the pFB for the v-GNE optimal selection of cocoercive games, as stated in Algorithm 2. The step sizes in Ψ need to be small enough with respect to η and to the matrices defining the constraints, as highlighted in Assumption 10, which states the sufficient conditions for the convergence of the pFB (see [15, Eq. (27) and Th. 3]).

Assumption 10: For all $i \in \mathcal{I}$

1) $\rho_i \leq (\max_{j=1,\dots,n_i} \sum_{k=1}^m |[A_i^\top]_{jk}| + \delta)^{-1}$;
2) $\tau_i \leq (\max_{j=1,\dots,n_i} \sum_{k=1}^m |[A_i]_{jk}| + 2|\mathcal{N}_i^\lambda| + \delta)^{-1}$;
3) $\sigma_i \leq (2|\mathcal{N}_i^\lambda| + \delta)^{-1}$, where $\delta > 1/(\min(\eta, (2\max_{i\in\mathcal{I}} |\mathcal{N}_i^\lambda|)^{-1}))$. □

Theorem 2: Let Assumptions 1–4, 6, and 8–10 hold. Let $\Omega^\star$ be the set of solutions to Problem (16) with $\mathcal{T} = \mathcal{T}_{\text{pFB}}$ defined in (24), where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are defined in (21)–(23). Let $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$, where $\boldsymbol{\omega}^{(k)} = (\boldsymbol{x}^{(k)}, \boldsymbol{\lambda}^{(k)}, \boldsymbol{\nu}^{(k)})$, be the sequence generated by Algorithm 2. Then, $\lim_{k\to\infty} \mathrm{dist}(\boldsymbol{\omega}^{(k)}, \Omega^\star) = 0$, and $(\boldsymbol{x}^{(k)})_{k\in\mathbb{N}}$ converges to an optimal v-GNE of the game in (4). □

V. Online Tracking Of Optimal Generalized Nash Equilibria

A. Online Optimal Equilibrium Tracking Problem

In the second part of this article, we consider the online GNE selection problem. Specifically, let us introduce the time-varying game

$$\forall t\in\mathbb{N}, \forall i\in\mathcal{I}: \begin{cases} \min\limits_{x_i\in\mathcal{X}_{i,t}} & J_{i,t}(\boldsymbol{x}) & (26a) \\ \text{s. t.} & \sum\limits_{j\in\mathcal{I}} g_{j,t}(x_j) \leq 0 & (26b) \end{cases}$$

where t denotes the time index. The problem is time-varying in the sense that the objective functions of the agents, as well as the constraints, may vary over time. We assume that each instance of the games in (26) satisfies Assumptions 1 and 2. The time-varying GNE selection problem thus concerns the tracking of the sequence $(\boldsymbol{\omega}_t^\star)_{t\in\mathbb{N}}$

Algorithm 2: Optimal v-GNE Selection via pFB and HDSM for Linearly Coupled Cocoercive Games.

Initialization: $x_i^{(0)} \in \mathcal{X}_i$, $\lambda_i^{(0)} \in \mathbb{R}^m_{\geq 0}$, and $\nu_i^{(0)} \in \mathbb{R}^m$, $\forall i \in \mathcal{I}$.

Iteration of each agent $i \in \mathcal{I}$.

1) Receives $x_j^{(k)}$ from agent $j \in \mathcal{N}_i^J$ and $\lambda_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.
2) Updates:

$$\mathring{x}_i^{(k)} = \mathrm{prox}^{\rho_i}_{\ell_i+\iota_{\mathcal{X}_i}}\left(x_i^{(k)} - \rho_i(\nabla_{x_i} f_i(\boldsymbol{x}^{(k)}) + A_i^\top \lambda_i^{(k)})\right),$$

$$\mathring{\nu}_i^{(k)} = \nu_i^{(k)} - \sigma_i \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(\lambda_i^{(k)} - \lambda_j^{(k)}\right).$$

3) Receives $\mathring{\nu}_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.
4) Updates:

$$\mathring{\lambda}_i^{(k)} = \mathrm{proj}_{\geq 0}\Big(\lambda_i^{(k)} + \tau_i\Big(A_i(2\mathring{x}_i^{(k)} - x_i^{(k)}) - b_i + \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\left(2\mathring{\nu}_i^{(k)} - 2\mathring{\nu}_j^{(k)} - \nu_i^{(k)} + \nu_j^{(k)}\right) - \sum_{j\in\mathcal{N}_i^\lambda}\left(\lambda_i^{(k)} - \lambda_j^{(k)}\right)\Big)\Big).$$

5) Sends $(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)})$ to a coordinator and receives back $\nabla_{\omega_i}\phi(\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)})$, where $\omega_i = (x_i, \lambda_i, \nu_i)$.
6) Updates:

$$(x_i^{(k+1)}, \lambda_i^{(k+1)}, \nu_i^{(k+1)}) = (\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)}) - \beta^{(k)}\nabla_{\omega_i}\phi(\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)}). \quad (25)$$

$$\forall t\in\mathbb{N}: \boldsymbol{\omega}_t^\star := \begin{cases} \operatorname*{argmin}\limits_{\boldsymbol{\omega}} & \phi_t(\boldsymbol{\omega}) & (27a) \\ \text{s. t.} & \boldsymbol{\omega} \in \mathrm{zer}(\mathcal{A}_t + \mathcal{B}_t + \mathcal{C}_t). & (27b) \end{cases}$$

The problems in (26) and (27) are a sequence in time of instances of (4) and (13), respectively. The operators $\mathcal{A}_t$, $\mathcal{B}_t$, and $\mathcal{C}_t$ are defined in (9)–(11), for the game in (26) at time step t. The agents need to compute the action $\boldsymbol{\omega}_{t+1}$, having only access to the game formulation up to time t. This setup describes the case in which the agents act in a variable environment with limited computation capabilities, so that they cannot compute the exact optimal selection before changes in the problem (either in the selection function or in the game) occur. The problem in (27) reduces to an online optimization problem for $|\mathcal{I}| = 1$, see, e.g., [33] and the references therein. Inspired by the online optimization literature, we propose to track the solution sequence $(\boldsymbol{\omega}_t^\star)_{t\in\mathbb{N}}$ by computing at each time step t an approximate solution of the problem at time $t-1$. Such a solution principle is based on the assumption that $\boldsymbol{\omega}_{t-1}^\star$ contains information on $\boldsymbol{\omega}_t^\star$, which is a standard assumption in online optimization, see e.g. [26, Assumption 1], and [39, Assumption 3] and it is introduced next.

Assumption 11: There exist $\delta \geq 0$ such that

$$\sup_{t\in\mathbb{N}} \|\boldsymbol{\omega}_{t+1}^\star - \boldsymbol{\omega}_t^\star\| \leq \delta. \qquad \square$$

For every $t\in \mathbb{N}$, and by choosing $\mathcal{T}_t$ such that

$$\boldsymbol{\omega} \in \text{zer}(\mathcal{A}_t + \mathcal{B}_t + \mathcal{C}_t) \Leftrightarrow \boldsymbol{\omega} \in \text{fix}(\mathcal{T}_t)$$

$\boldsymbol{\omega}_t^\star$ in (27) can be equivalently found as the solution of the time-varying fixed-point selection problem

$$\inf_{\boldsymbol{\omega}\in\text{fix}(\mathcal{T}_t)} \langle \boldsymbol{\omega} - \boldsymbol{\omega}_t^\star, \nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle \geq 0. \tag{28}$$

The sequence $(\boldsymbol{\omega}_t^\star)_{t\in\mathbb{N}}$ is well defined when, for each t, the solution of (27) is unique. Let us then introduce the following assumption, which guarantees uniqueness if $\text{fix}(\mathcal{T}_t)$ is closed and convex for all t [28, Thm. 2.3.3]. This is the case, for example, when $\mathcal{T}_t$ is quasi-nonexpansive [30, Prop. 1a]

Assumption 12: The selection function $\phi_t : \mathbb{R}^{n_\omega} \to \mathbb{R}$ in (28) is continuously differentiable, σ-strongly convex, and has L_ϕ-Lipschitz continuous gradient for all $t \in \mathbb{N}$. $\square$

Remark 4: Under Assumption 12, if $\mathcal{T}_t = \mathcal{T}$, for all $t \in \mathbb{N}$, and the selection function at time t is the sampling of a function that varies continuously over time, that is, $\phi_t(\boldsymbol{\omega}) = \phi(\boldsymbol{\omega}, t)$, then an estimate for δ in Assumption 11 can be found. In fact, if $\phi(\boldsymbol{\omega}, t)$ is continuously differentiable, we find by [40, Thm. 2F.7] that the mapping from t to the solution of (28) is locally Lipschitz continuous with Lipschitz constant $\sigma^{-1}|\nabla_t\phi(\boldsymbol{\omega}_t^\star, t)|$. Thus, if the time variation between two consecutive time steps t_1 and t_2 is small enough, δ can be estimated as $\sigma^{-1}|\nabla_t\phi(\boldsymbol{\omega}_{t_1}^\star, t_1)|(t_2 - t_1)$. The solution mapping is in general discontinuous when $\mathcal{T}_t$ is time-varying; thus, a similar estimate cannot be found in the general case. $\square$

B. Online fixed point tracking via the restarted HSDM

The existing results on the HSDM algorithm study the asymptotic behavior with vanishing step size $(\beta^{(k)})_{k\in\mathbb{N}}$ (see Assumption 4). However, in online scenarios, decision makers may not have the computational capability to exactly compute the fixed point of the algorithm, since that would require an infinite amount of iterations in a limited time span before a new instance of the problem becomes available. Thus, we propose an algorithm that only performs a finite number of HSDM iterations per time step. Consequently, the sequence of step sizes becomes truncated and a sequence of vanishing step sizes, which is required for the convergence of the HSDM, cannot be defined. We therefore simplify the analysis by considering a constant sequence of step sizes.

Let us introduce the restarted HSDM algorithm. Given an initial state $\boldsymbol{\omega}_1$, for each $t \in \mathbb{N}$, we propose the following:

$$\boldsymbol{y}^{(k+1)} := \begin{cases} \boldsymbol{\omega}_t, & \text{for } k = 1, \\ \mathcal{T}_t(\boldsymbol{y}^{(k)}) - \beta\nabla\phi_t(\mathcal{T}_t(\boldsymbol{y}^{(k)})), & \text{for } k = 2, \ldots, K, \end{cases}$$
$$\boldsymbol{\omega}_{t+1} := \boldsymbol{y}^{(K+1)}. \tag{29}$$

In words, at each time step t the auxiliary variable $\boldsymbol{y}^{(k)}$, with $k = 1, \ldots K$, is updated with K iterations of the HSDM. Then, the decision variable at time step $t + 1$ is obtained as $\boldsymbol{\omega}_{t+1} = \boldsymbol{y}^{(K+1)}$. The algorithm is then restarted when the information on the problem for the next time step becomes available. Next, let us postulate the following technical assumptions:

Assumption 13: There exists a compact set $\mathcal{Y}$ such that $\boldsymbol{\omega}_t^\star \in \mathcal{Y}$ for all $t \in \mathbb{N}$. $\square$

Assumption 14: There exists $U \geq 0$ such that $\sup_{\boldsymbol{\omega}\in\bigcup_{\tau\in\mathbb{N}}\text{Im}(\mathcal{T}_\tau), t\in\mathbb{N}} \|\nabla\phi_t(\boldsymbol{\omega})\| \leq U$. $\square$

The set $\mathcal{Y}$ introduced in Assumption 13 is only used in the analysis and its existence is practically reasonable, since we can assume that we do not aim at tracking a divergent sequence. Assumption 14 is in line with the online optimization literature (see [41, Assumption 5], [39, Assumption 5], among others).

As shown in Section III-C, the HSDM method converges to the solution of a selection problem over the fixed point set of a quasi-shrinking operator. In the online scenario, assuming the operator $\mathcal{T}_t$ to be quasi-shrinking for all t is not enough, as the quasi-shrinking property might not hold asymptotically. Thus, we also postulate the technical Assumption 15.

Assumption 15: (Uniformly quasi-shrinking operator) For any closed convex set C such that $C \cap \text{fix}(\mathcal{T}_t) \neq \varnothing$, there exists $D : \mathbb{R}_{\geq 0} \to \mathbb{R}$ positive semidefinite such that $D_t(r) \geq D(r)$ for all $t\in \mathbb{N}$ and for all $r \geq 0$, where $D_t(\cdot)$ is the shrinkage function of $\mathcal{T}_t$ defined as in (2). $\square$

Remark 5: Assumptions 13–15 are satisfied for example when at every time step t, the feasible set of Problem (27) is selected among the GNE sets of finitely many monotone games with a compact decision space. In fact, let $(\mathcal{T}_{\text{FBF}}^h)_{h=1}^H$ be the set of FBF operators such that, for all $t \in \mathbb{N}$, there exists $h \in \{1, \ldots, H\}$ such that $\text{zer}(\mathcal{A}_t + \mathcal{B}_t + \mathcal{C}_t) = \text{fix}(\mathcal{T}_{\text{FBF}}^h)$. The operators $(\mathcal{T}_{\text{FBF}}^h)_{h=1}^H$ are quasi-shrinking. Let us denote with $D^h(\cdot)$ the shrinkage function of $\mathcal{T}_{\text{FBF}}^h$. As the minimum among a finite number of positive semidefinite functions is such, Assumption 15 is then satisfied with $D(r) = \min_{h\in\{1,\ldots,H\}} D^h(r)$. Furthermore, Assumption 13 holds with $\mathcal{Y} = \cup_{h=1}^H \text{Im}(\mathcal{T}_{\text{FBF}}^h)$, which is compact, and Assumption 14 holds as $\nabla\phi_t$ is $L_\phi-$Lipschitz continuous for all t on a compact set. $\square$

We find that the restarted HSDM (29) asymptotically tracks the solution trajectory of the online fixed point selection problem in (28), with an asymptotic error that can be controlled up to the variability of the problem, δ, via an appropriate choice of β and K, as shown in Theorem 3.

Theorem 3: Let Assumptions 11–15 hold. Let the sequence $(\boldsymbol{\omega}_t)_{t\in\mathbb{N}}$ be generated by the restarted HSDM in (29). For any $\gamma > 0$, there exist $\beta \in (0, \frac{2\sigma}{L_\phi^2})$ and $\bar{K}$, such that, for all $K \geq \bar{K}$, the sequence $(\boldsymbol{\omega}_t)_{t\in\mathbb{N}}$ is bounded and

$$\limsup_{t\to\infty} \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^\star\|^2 \leq \frac{(\gamma + \delta^2)}{1/2 - \alpha} \tag{30}$$

where $\alpha = (1 - \tau(\beta))^K < \frac{1}{2}$ and $\tau(\beta) := 1 - \sqrt{1 - \beta(2\sigma - \beta L_\phi^2)} \in (0, 1)$. $\square$

Remark 6: As it emerges from the proof of Theorem 3 (see Remark 10 in the Appendix), to control the approximation error in (30), β must be chosen small so to obtain small values of γ. However, the value $\tau(\beta)$ tends to 0 for small values of β. This leads the denominator in (30) to be small for small stepsizes, unless the number of iterations K is increased. Therefore, a smaller step size leads to a better approximation error only if it is shouldered by an increase in the number of iterations of the algorithm per time step. □

In the next section, we use the restarted HSDM to solve the online GNE tracking problem in (27).

C. Distributed Optimal Equilibrium Tracking Algorithm for Monotone Games

We recall from Section III-C that the set of v-GNEs for a monotone game can be characterized as the set of fixed points of the operator $\mathcal{T}_{\text{FBF}}$ defined in (18). Thus, for the time-varying game in (26) at time t, let $\mathcal{T}_{\text{FBF},t}$ be the FBF operator defined as

$$\mathcal{T}_{\text{FBF},t}(\boldsymbol{\omega}) := ((\text{Id} - \Psi^{-1}(\mathcal{B}_t + \mathcal{C}_t))(\text{Id} + \Psi^{-1}\mathcal{A}_t)^{-1} \cdot (\text{Id} - \Psi^{-1}(\mathcal{B}_t + \mathcal{C}_t)) + \Psi^{-1}(\mathcal{B}_t + \mathcal{C}_t))(\boldsymbol{\omega}) \qquad (31)$$

where $\mathcal{A}_t$, $\mathcal{B}_t$, and $\mathcal{C}_t$ are those in Problem (27) and associated with the game in (26) at time t. The solutions of the time-varying GNE selection problem in (27) are equivalent to the solutions of (28), with $\mathcal{T}_t = \mathcal{T}_{\text{FBF},t}$ for all t. By Lemma 3, $\mathcal{T}_{\text{FBF},t}$, for each t, is a quasi-nonexpansive, quasi-shrinking operator. Therefore, the restarted HSDM algorithm in (29) can be employed for tracking the solution trajectory, with an asymptotic tracking error given by Theorem 3. We can then bound the asymptotic optimal GNE tracking error of Algorithm 3 by using Theorem 3, as formally stated next.

Corollary 1: Let us consider the online GNE tracking problem in (27) for the time-varying game in (26) that satisfies Assumptions 1, 2, 6, for each $t \in \mathbb{N}$. Suppose that Assumptions 11–14 hold and let $\mathcal{T}_t = \mathcal{T}_{\text{FBF},t}$ satisfy Assumption 15. Then, for any $\gamma > 0$ there exist $\beta \in (0, \frac{2\sigma}{L_\phi^2})$ and $\bar{K}$ such that, for any $K \geq \bar{K}$, the asymptotic tracking error of Algorithm 3 is given by (30). □

Remark 7: The solution sequence computed by Algorithm 3, $(\boldsymbol{\omega}_t)_{t\in\mathbb{N}}$, can violate the constraints of the game in (26). Such violation can be estimated using the Lipschitz continuity of $g_{i,t}$ for all $i \in \mathcal{I}$ (which follows from Assumption 1) and Theorem 3. Let us denote with L_g the maximum Lipschitz constant of $g_{i,t}$, for all $i \in \mathcal{I}$ and $t \in \mathbb{N}$, and by $(x_{i,t})_{i\in\mathcal{I}}$ the primal variables associated to $\boldsymbol{\omega}_t$. Then

$$\limsup_{t\to\infty} \sum_{i\in\mathcal{I}} g_{i,t}(x_{i,t}) \leq L_g \sqrt{\frac{\gamma + \delta^2}{1/2 - \alpha}}. \qquad \square$$

Remark 8: The result of this section holds similarly if we substitute the FBF operator with the pFB operator in (24), which is quasi-shrinking (see the proof of Theorem 2), for cocoercive games with affine coupling constraints. □

Algorithm 3: Optimal v-GNE Tracking via FBF and HSDM.

Initialization: $x_{i,0} \in \mathcal{X}_i$, $\lambda_{i,0} \in \mathbb{R}^m_{\geq 0}$, and $\nu_{i,0} \in \mathbb{R}^m$, $\forall i \in \mathcal{I}$.

Iteration at time $t \in \mathbb{N}_0$ **of each agent** $i \in \mathcal{I}$:

1) Receives $J_{i,t}(\cdot)$, $g_{i,t}(\cdot)$, and $\mathcal{X}_{i,t}$.

2) Assigns $\hat{x}_i^{(1)} \leftarrow x_{i,t}$, $\hat{\lambda}_i^{(1)} \leftarrow \lambda_{i,t}$, and $\hat{\nu}_i^{(1)} \leftarrow \nu_{i,t}$.

3) **For** $k = 1, \ldots, K$:

1) Receives $\hat{x}_j^{(k)}$ from agent $j \in \mathcal{N}_i^J$ and $\hat{\lambda}_j^{(k)}, \hat{\nu}_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.

2) Updates:

$$\tilde{x}_i^{(k)} = \text{prox}^{\rho_i}_{\ell_{i,t}+\iota_{\mathcal{X}_{i,t}}}\Big(\hat{x}_i^{(k)} - \rho_i(\nabla_{x_i} f_{i,t}(\hat{\boldsymbol{x}}^{(k)}) + \nabla g_{i,t}(\hat{x}_i^{(k)})^\top \hat{\lambda}_i^{(k)})\Big),$$

$$\tilde{\lambda}_i^{(k)} = \text{proj}_{\geq 0}\Big(\hat{\lambda}_i^{(k)} + \tau_i \Big(g_{i,t}(\hat{x}_i^{(k)}) + \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\Big(\hat{\nu}_i^{(k)} - \hat{\nu}_j^{(k)} - \hat{\lambda}_i^{(k)} + \hat{\lambda}_j^{(k)}\Big)\Big)\Big),$$

$$\tilde{\nu}_i^{(k)} = \hat{\nu}_i^{(k)} - \sigma_i \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\Big(\hat{\lambda}_i^{(k)} - \hat{\lambda}_j^{(k)}\Big).$$

3) Receives $\tilde{x}_j^{(k)}$ from agent $j \in \mathcal{N}_i^J$ and $\tilde{\lambda}_j^{(k)}, \tilde{\nu}_j^{(k)}$ from agent $j \in \mathcal{N}_i^\lambda$.

4) Updates:

$$\mathring{x}_i^{(k)} = \tilde{x}_i^{(k)} - \rho_i\Big(\nabla_{x_i} f_{i,t}(\tilde{\boldsymbol{x}}^{(k)}) - \nabla_{x_i} f_{i,t}(\hat{\boldsymbol{x}}^{(k)}) + \nabla g_{i,t}(\tilde{x}_i^{(k)})^\top \tilde{\lambda}_i^{(k)} - \nabla g_{i,t}(\hat{x}_i^{(k)})^\top \hat{\lambda}_i^{(k)}\Big),$$

$$\mathring{\lambda}_i^{(k)} = \tilde{\lambda}_i^{(k)} + \tau_i\Big(g_{i,t}(\tilde{x}_i^{(k)}) - g_{i,t}(\hat{x}_i^{(k)}) + \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\Big(\tilde{\nu}_i^{(k)} - \hat{\nu}_i^{(k)} - \tilde{\nu}_j^{(k)} + \hat{\nu}_j^{(k)}\Big) - \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\Big(\tilde{\lambda}_i^{(k)} - \hat{\lambda}_i^{(k)} - \tilde{\lambda}_j^{(k)} + \hat{\lambda}_j^{(k)}\Big)\Big),$$

$$\mathring{\nu}_i^{(k)} = \tilde{\nu}_i^{(k)} - \sigma_i \textstyle\sum_{j\in\mathcal{N}_i^\lambda}\Big(\tilde{\lambda}_i^{(k)} - \hat{\lambda}_i^{(k)} - \tilde{\lambda}_j^{(k)} + \hat{\lambda}_j^{(k)}\Big).$$

5) Sends $(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)})$ to a coordinator and receives $\nabla_{\boldsymbol{\omega}_i}\phi_t(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)})$, where $\boldsymbol{\omega}_i = (x_i, \lambda_i, \nu_i)$.

6) Updates:

$$(\hat{x}_i^{(k+1)}, \hat{\lambda}_i^{(k+1)}, \hat{\nu}_i^{(k+1)}) = (\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)}) - \beta\nabla_{\boldsymbol{\omega}_i}\phi_t(\mathring{x}_i^{(k)}, \mathring{\lambda}_i^{(k)}, \mathring{\nu}_i^{(k)}).$$

End For

4) Sets $x_{i,t+1} \leftarrow \hat{x}_i^{(K+1)}$, $\lambda_{i,t+1} \leftarrow \hat{\lambda}_i^{(K+1)}$, $\nu_{i,t+1} \leftarrow \hat{\nu}_i^{(K+1)}$.

VI. Illustrative Example

We consider a peer-to-peer electricity market clearing problem with operational constraints of the electrical network, adapted from [2]. We assume that each bus of a distribution network consists of one agent that has access to either a storage unit or a dispatchable generation unit. Each agent $i \in \mathcal{I}$ has decision authority on the power generated $p^{\text{g}}_{i,h}$, the power bought from the main grid $p^{\text{mg}}_{i,h}$, the power drawn from the storage unit $p^{\text{st}}_{i,h}$, the power traded with the trading partners $p^{\text{tr}}_{(i,j),h}$, $j \in \mathcal{N}_i$, and the phase at the bus $\theta_{i,h}$ over the horizon $h = 1, \ldots, H$. Let us denote $\boldsymbol{x}_{i,h} = \text{col}(p^{\text{g}}_{i,h}, p^{\text{mg}}_{i,h}, p^{\text{st}}_{i,h}, (p^{\text{tr}}_{(i,j),h})_{j\in\mathcal{N}_i}, \theta_{i,h})$, for all $i \in \mathcal{I}$ and $h = 1, \ldots, H$, and denote $\boldsymbol{x}_i := \text{col}((\boldsymbol{x}_{i,h})_{h=1,\ldots,H})$, $\boldsymbol{x} := \text{col}((\boldsymbol{x}_i)_{i\in\mathcal{I}})$. Each agent aims at minimizing its local cost function [2, Eq. (17)]

$$J_i(\boldsymbol{x}) = \sum_{h=1}^{H} f_{i,h}^{\mathrm{g}}(p_{i,h}^{\mathrm{g}}) + f_{i,h}^{\mathrm{tr}}((p_{(i,j),h}^{\mathrm{tr}})_{j\in\mathcal{N}_i}) + f_{i,h}^{\mathrm{mg}}(p_{i,h}^{\mathrm{mg}}, p_{-i,h}^{\mathrm{mg}}) \tag{32}$$

where $f_{i,h}^{\mathrm{tr}}$ encodes the cost or revenue of the trading with other agents and $f_{i,h}^{\mathrm{mg}}$ encodes the cost of purchasing energy from the main grid as in [2, Eq. (11)], while $f_{i,h}^{\mathrm{g}}$ is a linear function which encodes the cost of power generation. The local feasible sets $\mathcal{X}_i$, $i = 1,\ldots,N$ include the satisfaction of the power demand at the bus, as well as the operating constraints of the generators and storage units. The shared constraints are of the form $g(\boldsymbol{x}) \leq \mathbf{0}_{n_c}$, with g affine. They include the operating limits of the grid, the trading reciprocity $\{p_{(i,j),h}^{\mathrm{tr}} = -p_{(j,i),h}^{\mathrm{tr}}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}_i\}$, and the linearized power flow equations with dc approximation $\{p_{i,h}^{\mathrm{g}} + p_{i,h}^{\mathrm{st}} + \iota_i^{\mathrm{mg}} \sum_{j\in\mathcal{N}} p_{j,h}^{\mathrm{mg}} + \sum_{j\in\mathcal{B}_i} B_{ij}(\theta_{i,h} - \theta_{j,h}) = 0\}$, where $\iota_i^{\mathrm{mg}} \in \{0,1\}$ is 1, if and only if i is connected to the main grid, $\mathcal{B}_i$ is the set of buses that are connected to bus i on the electric grid, and B is the susceptance matrix. We note that the game satisfies Assumptions 1 and 2. We consider the IEEE 13-bus distribution feeder for our numerical simulations, performed in MATLAB.

We first simulate the day-ahead market clearing (with 24 hourly time steps) via the standard FBF-based algorithm, which can obtain a v-GNE, and Algorithm 1, which solves the optimal selection problem of this game. Specifically, we consider the GNE selection function

$$\phi(\boldsymbol{x}) = \sum_{h=1}^{H}\{\|\boldsymbol{p}_h^{\mathrm{g}} - \bar{\boldsymbol{p}}^{\mathrm{g}}\|_{Q_{\mathrm{d}}}^2 + \|\boldsymbol{p}_h^{\mathrm{mg}}\|_{Q_{\mathrm{mg}}}^2 + \|\boldsymbol{\theta}_h - \bar{\boldsymbol{\theta}}\|_{Q_\theta}^2 + \|G\boldsymbol{\theta}_h\|_{Q_{\mathrm{pf}}}^2 + \|\boldsymbol{p}_h^{\mathrm{tr}}\|_{Q_{\mathrm{tr}}}^2 + \|\boldsymbol{p}_h^{\mathrm{st}}\|_{Q_{\mathrm{st}}}^2\} + \|\boldsymbol{\lambda}\|_{Q_\lambda}^2 + \|\boldsymbol{\nu}\|_{Q_\nu}^2 \tag{33}$$

where we denoted in bold the column stack of the respective variables for each agent and the matrices $Q_\star$ are diagonal positive definite. We choose $\bar{\boldsymbol{p}}^{\mathrm{g}}$ to be the column vector of the maximum generation production for each agent, in order to maximize the renewable energy production, and $\bar{\boldsymbol{\theta}}$ to be a vector which elements are all equal to the phase of the node connected to the main grid, in order to reduce the grid imbalances. The cost factors related to $\boldsymbol{p}^{\mathrm{mg}}, \boldsymbol{p}^{\mathrm{st}}, \boldsymbol{p}^{\mathrm{tr}}$ aim at reducing the burden on the transmission grid, increasing the lifespan of the storage units, and reducing the load of the trading platform, respectively. The terms in $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ act as regularization of the dual variables. Finally, G is a matrix that maps the phase of the nodes to the power flowing through the lines. In this test, we aim at maximizing the lifespan of the grid lines by setting the nonzero elements of Q_{pf} to be large. We observe that, as expected, the solution obtained by Algorithm 1 achieves a 10.8% lower value of the selection function defined in (33) compared with the one achieved by the standard FBF, since the v-GNE computed by Algorithm 1 minimizes (33). In Fig. 1, we observe that Algorithm 1 generates solutions with less congestion (power flow) than that of the standard FBF, as intended by the term of the selection function in (33) weighted by Q_{pf}.

Second, we test Algorithm 3 on a real-time market scenario, formulated as a time-varying game. Because of the variability along the day of the power demand, the local power balance constraint defined in [2, Eq. (6)] depends on t. The constraints of the game are therefore time-varying. We aim at computing a v-GNE that minimizes the power flowing on the line connecting buses 632 and 671 during peak hours. Thus, we consider (33) as the selection function at each t where the element of Q_{pf} related to this line is time-varying, i.e., it is set high between the peak hours, i.e., 8 A.M. and 4 P.M. We note that this setup falls into the case considered in Remark 5, as only a finite umber of game instances are considered, whilst $(\phi_t)_t$ satisfies Assumption 12. The problem is solved every 15 min using the power balance constraints and selection function formulated at time-step t. After the computation is performed, the system implements the computed v-GNE at time $t+1$. The simulation is run over a 24 h span, thus resulting in 96 consecutive instances of GNE selection problems. Due to the relatively short sampling time, the demand is not expected to vary a lot between two consecutive game instances. We can then consider Assumption 11 to be satisfied. We run the simulation for different values of the parameters K and β and we compare the results with the baseline solution $(\boldsymbol{\omega}_t^{\mathrm{FBF}})$ obtained by running at each time-step the standard FBF algorithm for a limited ($9 \cdot 10^4$, that is, the largest K on which we tested the restarted HSDM) number of iterations. Fig. 2(a) illustrates the relative average residual obtained by restarted HSDM with respect to the baseline solution, where the residual

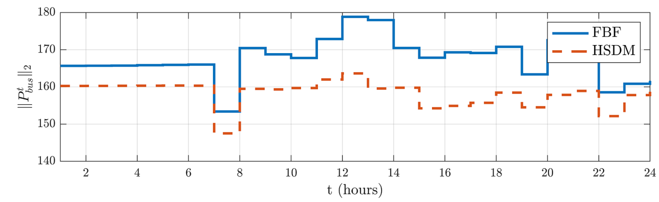

Fig. 1. Total power flow achieved by the proposed algorithm compared to standard FBF in the day-ahead market scenario.

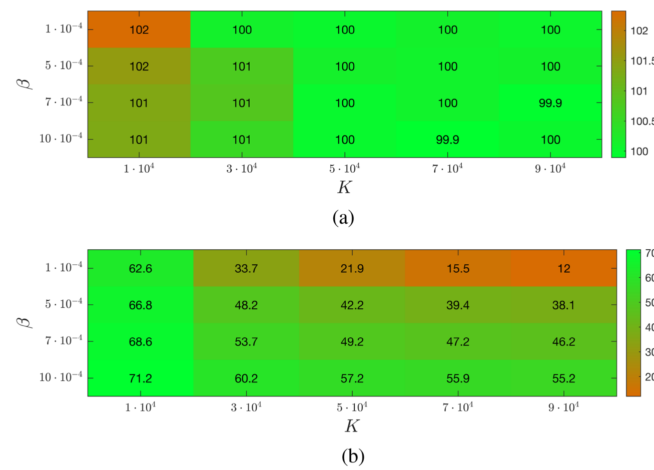

Fig. 2. Algorithm performance for several restarted HSDM parameters. (a) Average residual with respect to the baseline sequence $\frac{\mathfrak{R}_{t+1}(\boldsymbol{\omega}_t)}{\mathfrak{R}_{t+1}(\boldsymbol{\omega}_t^{\mathrm{FBF}})}$ (percentage). (b) Average cost improvement with respect to the baseline sequence $\frac{\phi_{t+1}(\boldsymbol{\omega}_t) - \phi_{t+1}(\boldsymbol{\omega}_t^{\mathrm{FBF}})}{\phi_{t+1}(\boldsymbol{\omega}_t^{\mathrm{FBF}})}$ (percentage).

is computed as

$$\mathfrak{R}_{t+1}(\boldsymbol{\omega}_t) = \|\mathcal{T}_{\mathrm{FBF},t+1}(\boldsymbol{\omega}_t) - \boldsymbol{\omega}_t\|.$$

The residual provides a measure of the constraint satisfaction for the problem in (27) and we observe a comparable performance. However, our algorithm achieves a significant improvement on the selection function values, as shown in Fig. 2(b). Furthermore, increasing K might lead to a reduction in cost advantage, as outlined by Fig. 2(b), because for low values of K the solution approaches the unconstrained minimizer of ϕ_t, while for high values of K it approaches the minimizer within the v-GNE set. We also observe that a diminishing β implies a slower reduction of the cost function, which results in a higher cost as shown in Fig. 2(b). Each iteration of the algorithm is computed in approximately 15 ms, thus in the considered 15 min timestep an agent is able to compute circa $6 \cdot 10^4$ iterations. In the simulations, we consider larger values of K to show the benefit of the iterations on tracking precision.

VII. Conclusion

The optimal generalized Nash equilibrium selection problem in monotone games can be solved distributively by combining the hybrid steepest descent method with an appropriate fixed-point operator. The key requirement to guarantee convergence to the set of optimal generalized Nash equilibria is the quasi-shrinking property, which holds true for certain fixed-point operators. The hybrid steepest descent method can also be modified to track a time-varying optimal generalized Nash equilibria. The resulting approach is suitable for real-time decision making in multi-agent dynamic environments. Future works include: 1) improving the convergence rate of the proposed method via second-order information of the selection function and/or inertial terms; 2) developing distributed Tikhonov-based methods for generalized Nash equilibrium selection problems as benchmarks for the proposed method.

Appendix A
Proof of Lemma 1

We prove by contradiction. We assume that there exists $r>0$ such that $D_\Psi(r)=0$. Thus, by the definition of D_Ψ in (2), there exists a sequence $(\boldsymbol{\omega}_k)_{k\in\mathbb{N}} \in (\mathrm{fix}(\mathcal{T})^{\Psi}_{\geq r})\bigcap C$ such that

$$\lim_{k\to\infty} \mathrm{dist}_\Psi(\boldsymbol{\omega}_k, \mathrm{fix}(\mathcal{T})) - \mathrm{dist}_\Psi(\mathcal{T}(\boldsymbol{\omega}_k), \mathrm{fix}(\mathcal{T})) = 0.$$

By the definition of projection, we have

$$\begin{aligned}\mathrm{dist}_\Psi(\mathcal{T}(\boldsymbol{\omega}_k), \mathrm{fix}(\mathcal{T})) &= \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\mathcal{T}(\boldsymbol{\omega}_k))\|_\Psi \\ &\leq \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi. \end{aligned} \quad (34)$$

By the quasi-nonexpansiveness of $\mathcal{T}$ (implied by (3)) and the latter inequality

$$\begin{aligned} 0 &\leq \underbrace{\|\boldsymbol{\omega}_k - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi}_{=\mathrm{dist}_\Psi(\boldsymbol{\omega}_k, \mathrm{fix}(\mathcal{T}))} - \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi \\ &\leq \mathrm{dist}_\Psi(\boldsymbol{\omega}_k, \mathrm{fix}(\mathcal{T})) - \mathrm{dist}_\Psi(\mathcal{T}(\boldsymbol{\omega}_k), \mathrm{fix}(\mathcal{T})) \xrightarrow{k\to\infty} 0. \end{aligned}$$

It follows that

$$\lim_{k\to\infty} \|\boldsymbol{\omega}_k - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi - \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi = 0.$$

By (3), we then have that

$$\begin{aligned} &\|\boldsymbol{\omega}_k - \mathcal{T}_2(\boldsymbol{\omega}_k)\|^2_\Psi \leq \\ &\tfrac{1}{\gamma}(\|\boldsymbol{\omega}_k - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|^2_\Psi - \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|^2_\Psi) \leq \\ &\tfrac{2d}{\gamma}(\|\boldsymbol{\omega}_k - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi - \|\mathcal{T}(\boldsymbol{\omega}_k) - \mathrm{proj}^{\Psi}_{\mathrm{fix}(\mathcal{T})}(\boldsymbol{\omega}_k)\|_\Psi) \end{aligned}$$

where the latter inequality follows from $a^2 - b^2 = (a-b)(a+b)$ for $a, b \in \mathbb{R}$ and where we substituted $d := \sup_{\boldsymbol{\omega}\in C} \|\boldsymbol{\omega}_k - \boldsymbol{\omega}\|_\Psi$, which is finite since the set C is compact. We conclude

$$\lim_{k\to\infty} \|\boldsymbol{\omega}_k - \mathcal{T}_2(\boldsymbol{\omega}_k)\|^2_\Psi = 0. \quad (35)$$

By the Bolzano–Weierstrass theorem and the boundedness of $\boldsymbol{\omega}_k$, there exists a convergent subsequence $(\boldsymbol{\omega}_{k_l})_{l\in\mathbb{N}}$ with accumulation point $\boldsymbol{\omega}^\infty$. By (35), $\lim_{l\to\infty} \mathcal{T}_2(\boldsymbol{\omega}_{k_l}) = \boldsymbol{\omega}^\infty$.

By the demiclosedness of $\mathrm{Id} - \mathcal{T}_2$ and by $\mathrm{fix}(\mathcal{T}_2) \subset \mathrm{fix}(\mathcal{T})$, $\boldsymbol{\omega}^\infty - \mathcal{T}_2(\boldsymbol{\omega}^\infty) = 0 \Rightarrow \boldsymbol{\omega}^\infty \in \mathrm{fix}(\mathcal{T}_2) \Rightarrow \boldsymbol{\omega}^\infty \in \mathrm{fix}(\mathcal{T})$. However, since $(\mathrm{fix}(\mathcal{T})^{\Psi}_{\geq r})\bigcap C$ is a closed set, then $\boldsymbol{\omega}^\infty \in \mathrm{fix}(\mathcal{T})^{\Psi}_{\geq r}$, which is in contradiction with $\boldsymbol{\omega}^\infty \in \mathrm{fix}(\mathcal{T})$. ∎

Appendix B
Properties of Operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in (9)–(11)

Lemma 4: Let Assumption 1 hold. Then, the operators $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in (9)–(11) are maximally monotone. Thus, $\mathcal{A}+\mathcal{B}+\mathcal{C}$ is also maximally monotone. □

Proof: By Assumption 1, $\mathrm{N}_{\mathcal{X}_i}$ and $\partial\ell_i$ are maximally monotone [29, Thm. 20.25 & Example 20.26]. The operator $\mathcal{A}$ is thus maximally monotone by [29, Prop. 20.23 & Cor. 25.5]. The operator F is maximally monotone by Assumption 2 and by continuity in Assumption 6. Meanwhile $\mathcal{L}$ is a linear positive semidefinite operator and, therefore, it is maximally monotone; thus, the operator $\mathcal{B}$ is maximally monotone. We can write $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$, where $\mathcal{C}_1 = \mathrm{col}((\langle\nabla_{x_i} g_i(x_i), \lambda_i\rangle)_{i\in\mathcal{I}}, -(g_i(x_i))_{i\in\mathcal{I}}, \mathbf{0}_{Nm})$ and $\mathcal{C}_2 = \mathrm{col}(\mathbf{0}_n, -(\mathcal{L}\otimes I_m)\boldsymbol{\nu}, (\mathcal{L}\otimes I_m)\boldsymbol{\lambda})$. The operator $\mathcal{C}_1$ is maximally monotone by continuity and by noting that, for any $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathbb{R}^n \times \mathbb{R}^{Nm}_{\geq 0} \times \mathbb{R}^{Nm}$

$$\begin{aligned} &\langle \mathcal{C}_1(\boldsymbol{\omega}) - \mathcal{C}_1(\boldsymbol{\omega}'), \boldsymbol{\omega} - \boldsymbol{\omega}'\rangle \\ &= \textstyle\sum_{i\in\mathcal{I}} \langle g_i(x_i') - g_i(x_i) - \nabla_{x_i} g_i(x_i)^\top (x_i' - x_i), \lambda_i\rangle \\ &\quad + \textstyle\sum_{i\in\mathcal{I}} \langle g_i(x_i) - g_i(x_i') - \nabla_{x_i} g_i(x_i')^\top (x_i - x_i'), \lambda_i'\rangle \geq 0 \end{aligned}$$

where the inequality follows by the convexity of g_i. As $\mathcal{C}_2$ is a linear skew-symmetric operator, it is maximally monotone [29, Ex. 20.35]. By invoking [29, Cor. 25.5], the result follows.

Lemma 5: Let Assumptions 1 and 6 hold. Then the operators $\mathcal{B}$, $\mathcal{C}$,and $\mathcal{B}+\mathcal{C}$, defined in (10)–(11), are Lipschitz continuous.

Proof: Due to Assumption 6, the operator $\mathcal{B}$ is L_F-Lipschitz continuous. Lipschitz continuity of $\mathcal{C}$ can be evaluated as follows. Similarly to the proof of Lemma 4, let us split $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$. The operator $\mathcal{C}_2$ is Lipschitz continuous by linearity, while Lipschitz continuity of $\mathcal{C}_1$ is shown as follows. Let us denote the bound of $\nabla_{x_i} g_i(x_i)$ by $b_{\nabla g_i}$, i.e., $\|\nabla_{x_i} g_i(x_i)\| \leq b_{\nabla g_i}$ (c.f. Assumption 6) and the bound of λ_i by b_{λ_i}, for all $i \in \mathcal{I}$, which exists due to [36, Prop. 3.3]. For any $\boldsymbol{\omega}, \boldsymbol{\omega}' \in \mathbb{R}^{n+2Nm}$

$$\begin{aligned}&\|\mathcal{C}_1(\boldsymbol{\omega}) - \mathcal{C}_1(\boldsymbol{\omega}')\|^2\\ &\overset{\{1\}}{\leq} \textstyle\sum_{i\in\mathcal{I}} \big(2\|\nabla_{x_i} g_i(x_i)^\top(\lambda_i - \lambda_i')\|^2 + \|g_i(x_i) - g_i(x_i')\|^2\\ &\quad +2\|(\nabla_{x_i} g_i(x_i) - \nabla_{x_i} g_i(x_i'))^\top \lambda_i'\|^2\big)\\ &\overset{\{2\}}{\leq} \textstyle\sum_{i\in\mathcal{I}} \big(2\|\nabla_{x_i} g_i(x_i)^\top\|^2\|\lambda_i - \lambda_i'\|^2 + b_{\nabla g_i}^2\|x_i - x_i'\|^2\\ &\quad +2\|\lambda_i'\|^2\|\nabla_{x_i} g_i(x_i) - \nabla_{x_i} g_i(x_i')\|^2\big)\\ &\overset{\{3\}}{\leq} \textstyle\sum_{i\in\mathcal{I}} \big(2b_{\nabla g_i}^2\|\lambda_i - \lambda_i'\|^2 + (2b_{\lambda_i}^2 L_{\nabla g}^2 + b_{\nabla g_i}^2)\|x_i - x_i'\|^2\big)\\ &\leq \textstyle\sum_{i\in\mathcal{I}} \max(2b_{\nabla g_i}^2, 2b_{\lambda_i}^2 L_{\nabla g}^2 + b_{\nabla g_i}^2)\|\boldsymbol{\omega}_i - \boldsymbol{\omega}_i'\|^2\end{aligned}$$

where $\{1\}$ follows by adding and subtracting the term $\nabla_{x_i} g_i(x_i)^\top \lambda_i'$ and by the bound $\|a+b\|^2 \leq 2\|a\|^2 + 2\|b\|^2$; $\{2\}$ is obtained by the Cauchy–Schwartz inequality and by the fact that g_i is Lipschitz since it has a bounded gradient; $\{3\}$ is obtained by the Lipschitz continuity of $\nabla_{x_i} g_i$. Hence, $\mathcal{C}_1$ is $L_{\mathcal{C}_1}$-Lipschitz continuous, where $L_{\mathcal{C}_1} = \max_{i\in\mathcal{I}}(\max(2b_{\nabla g_i}, \sqrt{2b_{\lambda_i}^2 L_{\nabla g}^2 + b_{\nabla g_i}^2}))$. Since the sum of Lipschitz continuous operators is Lipschitz continuous, the result follows. ∎

Appendix C
Results and Proofs of Section III

The following lemma shows the equivalence between $\mathrm{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C})$ and $\mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$.

Lemma 6: Let Assumptions 1, 2, 6, and 7 hold. Furthermore, let $\mathcal{T}_{\mathrm{FBF}}$ be defined by (18) while $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ be defined in (9)–(11). Then, $\mathrm{fix}(\mathcal{T}_{\mathrm{FBF}}) = \mathrm{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C})$. □

Proof: The proof is analogous to that of [17, Prop. 1]. ∎

The following lemma is used to prove the quasi-shrinking property of the FBF operator (18).

Lemma 7: Let $\mathcal{A}$ and $\mathcal{B}$ maximally monotone and $\mathcal{B}$ continuous. Let

$$\mathcal{T} = (\mathrm{Id} + \Psi^{-1}\mathcal{A})^{-1}(\mathrm{Id} - \Psi^{-1}\mathcal{B}).$$

Then, $\mathrm{Id} - \mathcal{T}$ is demiclosed at 0. □

Proof: Let us consider a sequence $(v_k)_{k\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} v_k = v, \qquad \lim_{k\to\infty} (\mathrm{Id} - \mathcal{T})(v_k) = 0.$$

We want to prove that $v - \mathcal{T}(v) = 0$ or, equivalently, $v \in \mathrm{fix}(\mathcal{T})$. Let us define $u_k := (\mathrm{Id} - \mathcal{T})(v_k)$. Then,

$$\begin{aligned}&v_k - u_k = (\mathrm{Id} + \Psi^{-1}\mathcal{A})^{-1}(\mathrm{Id} - \Psi^{-1}\mathcal{B})(v_k)\\ &\Leftrightarrow (\mathrm{Id} - \Psi^{-1}\mathcal{B})(v_k) \in (\mathrm{Id} + \Psi^{-1}\mathcal{A})(v_k - u_k)\\ &\Leftrightarrow v_k - \Psi^{-1}\mathcal{B}(v_k) + u_k - v_k \in \Psi^{-1}\mathcal{A}(v_k - u_k)\\ &\Leftrightarrow -\mathcal{B}(v_k) + \Psi u_k \in \mathcal{A}(v_k - u_k).\end{aligned}$$

By the continuity of $\mathcal{B}$ and [29, Fact 1.19], we conclude that $\lim_{k\to\infty} -\mathcal{B}(v_k) + \Psi u_k = -\mathcal{B}(v)$. By [29, Prop. 20.37] and the monotonicity of $\mathcal{A}$, $\mathrm{gph}(\mathcal{A})$ is closed. Therefore, since $\lim_{k\to\infty} v_k - u_k = v$, we conclude that $-\mathcal{B}(v) \in \mathcal{A}(v)$. By [29, Prop. 26.1(iv)], we obtain $v \in \mathrm{fix}(\mathcal{T})$. ∎

A. Proof of Lemma 3

By Lemmas 4 and 5, the operator $\mathcal{A}$ is maximally monotone whereas the operator $\mathcal{B}+\mathcal{C}$ is maximally monotone and Lipschitz continuous with Lipschitz constant denoted by L_B. Then, [17, Cor. 1] shows that $\mathcal{T}_{\mathrm{FBF}}$ is quasi-nonexpansive under Assumption 7. Specifically, it holds that [17, Prop. 2]

$$\|\mathcal{T}_{\mathrm{FBF}}(\boldsymbol{\omega}) - \boldsymbol{\omega}^\star\|_\Psi^2 \leq \|\boldsymbol{\omega} - \boldsymbol{\omega}^\star\|_\Psi^2 - \tfrac{L_B^2}{\mu_{\min}(\Psi)^2}\|\tilde{\boldsymbol{\omega}} - \boldsymbol{\omega}\|_\Psi^2 \tag{36}$$

where $\boldsymbol{\omega}^\star \in \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$, $\mu_{\min}(\Psi)$ is the smallest eigenvalue of Ψ and $\tilde{\boldsymbol{\omega}} = (\mathrm{Id} + \Psi^{-1}\mathcal{A})^{-1}(\mathrm{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C}))(\boldsymbol{\omega})$. Finally, we prove that $\mathcal{T}_{\mathrm{FBF}}$ is quasi-shrinking by invoking Lemma 1. Specifically, we choose $\mathcal{T}_2 = (\mathrm{Id} + \Psi^{-1}\mathcal{A})^{-1}(\mathrm{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C}))$. By [29, Prop. 26.1(iv)] and Lemma 6, $\mathrm{fix}(\mathcal{T}_2) = \mathrm{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C}) = \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$. Moreover, Lemma 7 shows that $\mathrm{Id} - \mathcal{T}_2$ is demiclosed at 0 and (36) is indeed the inequality in (3) for $\mathcal{T}_{\mathrm{FBF}}$. ∎

Remark 9: Although [17, Cor. 1] shows quasi-nonexpansiveness of $\mathcal{T}_{\mathrm{FBF}}$ and [17, Prop. 2] shows the inequality in (36) for Problem (4) with a linear coupling constraint, these results also holds for nonlinear functions $g_i(x_i)$, for all $i \in \mathcal{I}$, as long as Assumption 6 holds, since the operator $\mathcal{C}$ in (11) remains Lipschitz continuous. □

B. Proof of Theorem 1

Let us introduce the following preliminary lemma.

Lemma 8: Let Assumptions 1–4 and 6–7 hold. Then, the sequence $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$ generated by the HSDM method in (17) with $\mathcal{T} = \mathcal{T}_{\mathrm{FBF}}$ in (18), where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are defined in (9)–(11) and Ψ is defined in (19), is bounded, i.e., for any arbitrary $\boldsymbol{\omega}^\star \in \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$, it holds that $\|\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^\star\| \leq R(\boldsymbol{\omega}^\star)$, for some positive finite $R(\boldsymbol{\omega}^\star)$. □

Proof: First, we show that, for an arbitrary $\boldsymbol{\omega}^\star \in \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$,

$$\|\mathcal{T}_{\mathrm{FBF}}(\boldsymbol{\omega}) - \boldsymbol{\omega}^\star\|_\Psi^2 < \|\boldsymbol{\omega} - \boldsymbol{\omega}^\star\|_\Psi^2 \tag{37}$$

for all $\boldsymbol{\omega} \notin \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$. To this end, let us recall the inequality (36) in the proof of Lemma 3, which holds under the considered assumptions

$$\|\mathcal{T}_{\mathrm{FBF}}(\boldsymbol{\omega}) - \boldsymbol{\omega}^\star\|_\Psi^2 \leq \|\boldsymbol{\omega} - \boldsymbol{\omega}^\star\|_\Psi^2 - (L_B/\mu_{\min}(\Psi))^2\|\tilde{\boldsymbol{\omega}} - \boldsymbol{\omega}\|_\Psi^2$$

where $\tilde{\boldsymbol{\omega}} = (\mathrm{Id} + \Psi^{-1}\mathcal{A})^{-1}(\mathrm{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C})))(\boldsymbol{\omega}) =: \mathcal{T}_2(\boldsymbol{\omega})$. By [29, Prop. 26.1(iv)] and Lemma 6, $\mathrm{fix}(\mathcal{T}_2) = \mathrm{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C}) = \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$. Hence, $\tilde{\boldsymbol{\omega}} \neq \boldsymbol{\omega}$ if $\boldsymbol{\omega} \notin \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$. We observe from the preceding inequality that when $\tilde{\boldsymbol{\omega}} \neq \boldsymbol{\omega}$, (37) holds.

We now show that for any arbitrary fixed point $\boldsymbol{\omega}^\star \in \mathrm{fix}(\mathcal{T}_{\mathrm{FBF}})$, there exists $R > 0$ such that

$$\inf_{\|\boldsymbol{\omega}-\boldsymbol{\omega}^\star\|\geq R}(\|\boldsymbol{\omega} - \boldsymbol{\omega}^\star\| - \|\mathcal{T}_{\mathrm{FBF}}(\boldsymbol{\omega}) - \boldsymbol{\omega}^\star\|) > 0. \tag{38}$$

We proceed to prove (38) by contradiction. By the nonexpansiveness of $\mathcal{T}_{\mathrm{FBF}}$, (38) can only be false if for all $R > 0$ there exists a sequence $(\boldsymbol{\omega}_k)_{k\in\mathbb{N}}$ converging to $\bar{\boldsymbol{\omega}}$ such that $\|\boldsymbol{\omega}_k - \boldsymbol{\omega}^\star\| \geq R$ and

$$\lim_{k\to\infty} \|\boldsymbol{\omega}_k - \boldsymbol{\omega}^\star\| - \|\mathcal{T}_{\mathrm{FBF}}(\boldsymbol{\omega}_k) - \boldsymbol{\omega}^\star\| = 0.$$

In particular, the latter holds true for $R > \sup_{\boldsymbol{y}\in \text{fix}(\mathcal{T}_{\text{FBF}})} \|\boldsymbol{y} - \boldsymbol{\omega}^\star\| + \varepsilon$, which implies $\bar{\boldsymbol{\omega}} \notin \text{fix}(\mathcal{T}_{\text{FBF}})$. Then, by (36), $\lim_{k\to\infty} \|\mathcal{T}_2(\boldsymbol{\omega}_k) - \boldsymbol{\omega}_k\| = 0$. As $\text{Id} - \mathcal{T}_2$ is demiclosed at zero by Lemma 7, this implies the contradiction $\bar{\boldsymbol{\omega}} \in \text{fix}(\mathcal{T}_2) \Rightarrow \bar{\boldsymbol{\omega}} \in \text{fix}(\mathcal{T}_{\text{FBF}})$. The inequality in (38) is used in[42, Th. 2] to prove the boundedness of the HSDM sequence with a nonexpansive operator $\mathcal{T}$ that satisfies (37). As (38) holds also for $\mathcal{T}_{\text{FBF}}$, the same proof holds under the remaining assumptions: (i) $\nabla\phi$ is monotone and Lipschitz continuous (Assumption 3), and (ii) the step size $\beta^{(k)}$ is nonsummable but square summable (Assumption 4). ■

We are now ready to proceed with the proof of Theorem 1.

Proof: Let $\tilde{\boldsymbol{\omega}}^{(k)} = (\tilde{\boldsymbol{x}}^{(k)}, \tilde{\boldsymbol{\lambda}}^{(k)}, \tilde{\boldsymbol{\nu}}^{(k)})$ and $\mathring{\boldsymbol{\omega}}^{(k)} = (\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)})$, where $\tilde{\boldsymbol{x}}^{(k)} = \text{col}((\tilde{x}_i)_{i\in\mathcal{I}})$ and the other variables are defined similarly. The updates of $\tilde{\boldsymbol{\omega}}^{(k)}$ in Step 2 of Algorithm 1 can be compactly written as

$$\tilde{\boldsymbol{\omega}}^{(k)} = (\text{Id} + \Psi^{-1}\mathcal{A})^{-1}(\text{Id} - \Psi^{-1}(\mathcal{B}+\mathcal{C}))(\boldsymbol{\omega}^{(k)})$$

whereas the updates of $\mathring{\boldsymbol{\omega}}^{(k)}$ in Step 4 of Algorithm 1 can be compactly written as $\mathring{\boldsymbol{\omega}}^{(k)} = \tilde{\boldsymbol{\omega}}^{(k)} - \Psi^{-1}(\mathcal{B}+\mathcal{C})(\tilde{\boldsymbol{\omega}}^{(k)} - \boldsymbol{\omega}^{(k)})$, implying that $\mathring{\boldsymbol{\omega}}^{(k)} = \mathcal{T}_{\text{FBF}}(\boldsymbol{\omega}^{(k)})$ and the updates in (20) is compactly written as

$$\boldsymbol{\omega}^{(k+1)} = \mathcal{T}_{\text{FBF}}(\boldsymbol{\omega}^{(k)}) - \beta^{(k)}\nabla\phi(\mathcal{T}_{\text{FBF}}(\boldsymbol{\omega}^{(k)})) \tag{39}$$

which is the HSDM applied to $\mathcal{T}_{\text{FBF}}$. We can then invoke Lemma 2 to claim the hypothesis. By Lemma 6, $\text{fix}(\mathcal{T}_{\text{FBF}}) = \text{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C})$; therefore $\text{fix}(\mathcal{T}_{\text{FBF}})$ is nonempty and bounded. Moreover, by Assumption 4, the step size $\beta^{(k)}$ meets the conditions in Lemma 2. Lemma 3 shows that $\mathcal{T}_{\text{FBF}}$ is quasi-nonexpansive and quasi-shrinking on any bounded closed convex set, C such that $C \cap \text{fix}(\mathcal{T}_{\text{FBF}}) \neq \varnothing$. On the other hand, Lemma 8 shows that the FBF-HSDM sequence $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$ obtained by the iterations in (39) is bounded, i.e., for any $\boldsymbol{\omega}^\star \in \text{fix}(\mathcal{T}_{\text{FBF}})$, there exists a positive finite $R(\boldsymbol{\omega}^\star)$ such that $\|\boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^\star\| \le R(\boldsymbol{\omega}^\star)$. Therefore, for an arbitrarily chosen $\boldsymbol{\omega}^\star \in \text{fix}(\mathcal{T}_{\text{FBF}})$, we can construct the following bounded closed set $\mathfrak{B}(\boldsymbol{\omega}^\star) := \{x \in \text{dom}(\mathcal{T}_{\text{FBF}}) \mid \|x - \boldsymbol{\omega}^\star\| \le R(\boldsymbol{\omega}^\star)\}$, on which the sequence $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$ lies. Moreover, we can observe that indeed $\mathfrak{B} \cap \text{fix}(\mathcal{T}_{\text{FBF}}) \neq \varnothing$, since $\boldsymbol{\omega}^\star \in \mathfrak{B}$ is a fixed point of $\mathcal{T}_{\text{FBF}}$. Hence, $\mathcal{T}_{\text{FBF}}$ is quasi-shrinking on $\mathfrak{B}$. ■

Appendix D
Proofs of Section IV

A. Proof of Theorem 2

First, we observe that in Algorithm 2, $\mathring{\boldsymbol{\omega}}^{(k)} = (\mathring{\boldsymbol{x}}^{(k)}, \mathring{\boldsymbol{\lambda}}^{(k)}, \mathring{\boldsymbol{\nu}}^{(k)})$ is updated by using $\mathcal{T}_{\text{pFB}}$ in (24), i.e., $\mathring{\boldsymbol{\omega}}^{(k)} = \mathcal{T}_{\text{pFB}}(\boldsymbol{\omega}^{(k)})$ [15, Section 4, Algorithm 1]. Hence, we can see that $\boldsymbol{\omega}^{(k)}$ is updated via the HSDM method, i.e.,

$$\boldsymbol{\omega}^{(k+1)} = \mathcal{T}_{\text{pFB}}(\boldsymbol{\omega}^{(k)}) - \beta^{(k)}\nabla\phi(\mathcal{T}_{\text{pFB}}(\boldsymbol{\omega}^{(k)})). \tag{40}$$

Similarly to the proof of Theorem 1, due to the boundedness of $\text{fix}(\mathcal{T}_{\text{pFB}}) = \text{zer}(\mathcal{A}+\mathcal{B}+\mathcal{C}) \neq \varnothing$ and the step size rule of $\beta^{(k)}$ in Assumption 4, we can invoke Lemma 2. Specifically, the operator $\mathcal{T}_{\text{pFB}}$ is averaged nonexpansive when Assumptions 1, 2, 6, and 8–10 hold [15, Th. 3]. Therefore, $\mathcal{T}_{\text{pFB}}$ is also quasi-nonexpansive [29, Sec. 4.1]. By [29, Prop. 4.35 (iii)], the condition in (3) holds with $\mathcal{T} = \mathcal{T}_2 = \mathcal{T}_{\text{pFB}}$. By [29, Thm. 4.27], $\text{Id} - \mathcal{T}_{\text{pFB}}$ is demiclosed at 0. Therefore, by Lemma 1, $\mathcal{T}_{\text{pFB}}$ is quasi-shrinking on any closed bounded convex set whose intersection with $\text{fix}(\mathcal{T}_{\text{pFB}})$ is nonempty. Furthermore, since $\mathcal{T}_{\text{pFB}}$ is averaged nonexpansive, $\mathcal{T}_{\text{pFB}}$ is attracting. Therefore, by [42, Th. 2] and due to the choice of the step size $\beta^{(k)}$ in Assumption 4, the sequence generated by (40) is bounded. Following the steps in the proof of Theorem 1, we can find a bounded set $\mathfrak{B}$ such that $\boldsymbol{\omega}^{(k)} \in \mathfrak{B}$ and $\mathcal{T}_{\text{pFB}}$ is quasi-shrinking on $\mathfrak{B}$. ■

Appendix E
Proofs of Section V

A. Preliminary Results

First, we show a series of preliminary results in Lemmas 9–12 that lead to the proof of Theorem 3. The proofs of this section are provided in the standard Euclidean norm for ease of notation. However, the case for any Ψ-induced norm, with $\Psi \succ 0$, follows verbatim.

Lemma 9: Let $\psi : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ be nondecreasing and nonnegative. Let a sequence $(b^{(k)})_{k\in\mathbb{N}}$ be nonincreasing, non negative. Let $(a^{(k)})_{k\in\mathbb{N}} \subset [0,\infty)$ satisfy

$$a^{(k+1)} \le a^{(k)} - \psi(a^{(k)}) + b^{(k+1)}. \tag{41}$$

Let $K \in \mathbb{N}$. If there exists $\xi > 0$ such that $\psi(\xi) \ge \max\{2b^{(1)}, \frac{2}{K-1}a^{(1)}\}$, then

$$a^{(k)} \le \xi + b^{(k)}, \quad \forall\, k \ge K. \tag{42}$$

□

Proof: Let us first show that there exists an $M \in \mathbb{N}$, $M \le K$ such that $a^{(M)} \le \xi$. We proceed by contradiction, assuming that $a^{(k)} > \xi\ \forall k = 1,\ldots,K$. Then, by noting that $\psi(\cdot)$ is nondecreasing and that $\psi(\xi) \ge 2b^{(k)}$ for all $k \in \mathbb{N}$, we have

$$\begin{aligned} a^{(k+1)} &\le a^{(k)} - \psi(a^{(k)}) + b^{(k+1)} \\ &\le a^{(k)} - \psi(\xi) + \tfrac{1}{2}\psi(\xi) = a^{(k)} - \tfrac{1}{2}\psi(\xi). \end{aligned}$$

By iterating the latter relation and recalling that $\psi(\xi) \ge \frac{2}{K-1}a^{(1)}$, we find that

$$a^{(k+1)} \le a^{(1)} - \tfrac{k}{2}\psi(\xi) \le a^{(1)} - \tfrac{k}{K-1}a^{(1)}.$$

For $k = K$, we then obtain the contradiction $a^{(K+1)} < 0$. Thus, there exists $M \le K$ such that $a^{(M)} \le \xi$. We then proceed by induction to prove (42). Let us prove that, if $a^{(k)} \le \xi + b^{(k)}$ then $a^{(k+1)} \le \xi + b^{(k+1)}$ for all $k \ge M$. We distinguish the following two cases.

1) Case $a^{(k)} < \xi$. Then, by (41) and by the nonnegativity of $\psi(\cdot)$, $a^{(k+1)} \le a^{(k)} + b^{(k+1)} < \xi + b^{(k+1)}$.
2) Case $\xi \le a^{(k)} \le \xi + b^{(k)}$. Then, by the nondecreasing property of ψ, $a^{(k)} \ge \xi \Rightarrow \psi(a^{(k)}) \ge \psi(\xi)$. By the assumptions, $\psi(\xi) \ge 2b^{(1)}$ and by the nonincreasing property of $(b^k)_{k\in\mathbb{N}}$, $2b^{(1)} \ge b^{(k)} + b^{(k+1)}$. We thus obtain $\psi(a^{(k)}) \ge b^{(k)} + b^{(k+1)}$. Substituting into (41) leads to

$$a^{(k+1)} \le a^{(k)} - b^{(k)} \le \xi.$$

We conclude by induction that $a^{(k)} \leq \xi + b^{(k)}$ for all $k \geq M$ and, since $M \leq K$, the claim in (42) immediately follows. ■

Lemma 10: Let $\mathcal{T}$ be quasi-nonexpansive and $\mathcal{F}$ be strongly monotone, such that $\|\mathcal{F}(\boldsymbol{\omega})\| \leq U$, for all $\boldsymbol{\omega} \in \operatorname{im}(\mathcal{T})$. Let $(\boldsymbol{\omega}^{(k)})_{k\in\mathbb{N}}$ be generated from (17) with $\beta^{(k)} = \beta > 0$ for all k. Let $K \in \mathbb{N}$ and let $\boldsymbol{\omega}^\star$ be the solution of $\mathrm{VI}(\mathcal{F}, \operatorname{fix}(\mathcal{T}))$. If there exists ξ such that the shrinkage function $D(\cdot)$ of $\mathcal{T}$, defined in (2), satisfies $D(\xi) \geq \max\{2\beta U, 2\frac{\operatorname{dist}(\omega^{(1)},\operatorname{fix}(\mathcal{T}))}{K-1}\}$, then the following inequalities hold:

$$\sup_{k\geq K} \operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T})) \leq \xi + \beta U, \tag{43}$$

$$\sup_{k\geq K} \|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\| \leq 2(\xi + \beta U), \tag{44}$$

$$\sup_{k\geq K} \langle \mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle \leq 3(\xi + \beta U)\|\mathcal{F}(\boldsymbol{\omega}^\star)\|. \tag{45}$$

□

Proof: (i.) For all k, it holds by the definition of distance and by the algorithm definition in (17) that

$$\begin{aligned}&\operatorname{dist}(\boldsymbol{\omega}^{(k+1)}, \operatorname{fix}(\mathcal{T})) \leq \|\boldsymbol{\omega}^{(k+1)} - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\mathcal{T}(\boldsymbol{\omega}^{(k)}))\| =\\ &\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \beta\mathcal{F}(\mathcal{T}(\boldsymbol{\omega}^{(k)})) - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\mathcal{T}(\boldsymbol{\omega}^{(k)}))\| \leq\\ &\underbrace{\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\mathcal{T}(\boldsymbol{\omega}^{(k)}))\|}_{=\operatorname{dist}(\mathcal{T}(\boldsymbol{\omega}^{(k)}),\operatorname{fix}(\mathcal{T}))} + \beta\|\mathcal{F}(\mathcal{T}(\boldsymbol{\omega}^{(k)}))\| \leq\\ &\operatorname{dist}(\mathcal{T}(\boldsymbol{\omega}^{(k)}), \operatorname{fix}(\mathcal{T})) + \beta U.\end{aligned} \tag{46}$$

Let us define $a^{(k)} := \operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T}))$. Then, from (46) we find immediately $a^{(k+1)} - \beta U \leq \operatorname{dist}(\mathcal{T}(\boldsymbol{\omega}^{(k)}), \operatorname{fix}(\mathcal{T}))$. By the definition of shrinkage function in (2) and the latter inequality, we can write

$$\begin{aligned}&D(a^{(k)}) \leq a^{(k)} - \operatorname{dist}(\mathcal{T}(\boldsymbol{\omega}^{(k)}), \operatorname{fix}(\mathcal{T})) \leq\\ &a^{(k)} - a^{(k+1)} + \beta U \Rightarrow a^{(k+1)} \leq a^{(k)} + \beta U - D(a^{(k)})\end{aligned}$$

which defines a sequence of the kind in (41) with $\psi(\cdot) = D(\cdot)$ and $b^{(k)} = \beta U$ for all k. By Lemma 9, then $\operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T})) \leq \xi + \beta U$ for all $k \geq K$.

(ii.) By the triangle inequality, we can write $\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\| \leq \|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)})\| + \|\operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\|$. By quasi-nonexpansiveness of $\mathcal{T}$, we obtain, for all $k \geq K$

$$\begin{aligned}&\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)})\| \leq\\ &\|\boldsymbol{\omega}^{(k)} - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)})\| = \operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T}))\\ &\Rightarrow \|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\| \leq 2\operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T})).\end{aligned}$$

Finally, combining the last inequality and (43) yields (44).

(iii) By the Cauchy–Schwarz inequality, we can write

$$\begin{aligned}&\langle \mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle =\\ &\langle \mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle + \langle \boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle \leq\\ &\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\|\|\mathcal{F}(\boldsymbol{\omega}^\star)\| + \langle \boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle.\end{aligned} \tag{47}$$

Based on (44), for all $k \geq K$, we can bound the first term on the right-hand side of (47) by $\|\mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^{(k)}\|\|\mathcal{F}(\boldsymbol{\omega}^\star)\| \leq 2(\xi + \beta U)\|\mathcal{F}(\boldsymbol{\omega}^\star)\|$ and rewrite the second term as

$$\begin{aligned}\langle \boldsymbol{\omega}^{(k)} - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle &= \langle \boldsymbol{\omega}^{(k)} - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)}), -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle\\ &\quad + \langle \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle.\end{aligned}$$

We observe that the second addend is nonpositive by the definition of VI solution. By applying the Cauchy–Schwarz inequality, the definition of projection, and (44), we obtain

$$\begin{aligned}&\langle \mathcal{T}(\boldsymbol{\omega}^{(k)}) - \boldsymbol{\omega}^\star, -\mathcal{F}(\boldsymbol{\omega}^\star)\rangle\\ &\leq 2(\xi + \beta U)\|\mathcal{F}(\boldsymbol{\omega}^\star)\| + \|\boldsymbol{\omega}^{(k)} - \operatorname{proj}_{\operatorname{fix}(\mathcal{T})}(\boldsymbol{\omega}^{(k)})\|\|\mathcal{F}(\boldsymbol{\omega}^\star)\|\\ &= 2(\xi + \lambda U)\|\mathcal{F}(\boldsymbol{\omega}^\star)\| + \operatorname{dist}(\boldsymbol{\omega}^{(k)}, \operatorname{fix}(\mathcal{T}))\|\mathcal{F}(\boldsymbol{\omega}^\star)\|\\ &\leq 3(\xi + \lambda U)\|\mathcal{F}(\boldsymbol{\omega}^\star)\|.\end{aligned}$$

■

Lemma 11: Let Assumptions 12–14 hold. For any $t \in \mathbb{N}$, let $\boldsymbol{\omega}_{t+1}$ be generated from the step at time t of the restarted HSDM algorithm in (29). Let $D_t(\cdot)$ be the shrinkage function of $\mathcal{T}_t$ as defined in (2). If there exists $\xi > 0$ such that

$$D_t(\xi) \geq \max\left\{2\beta U, 2\frac{\operatorname{dist}(\boldsymbol{\omega}_t,\operatorname{fix}(\mathcal{T}_t))}{K-1}\right\} \tag{48}$$

then

$$\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 \leq (1 - \tau(\beta))^K \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^\star\|^2 + \gamma \tag{49}$$

with

$$\gamma = \tfrac{\beta}{\tau(\beta)} U(6\xi + 11\beta U). \tag{50}$$

□

Proof: Let us define the operator $\mathcal{T}_t^\beta(\boldsymbol{\omega}) := \mathcal{T}_t(\boldsymbol{\omega}) - \beta\nabla\phi_t(\mathcal{T}_t(\boldsymbol{\omega}))$. By $\mathcal{T}_t(\boldsymbol{\omega}_t^\star) = \boldsymbol{\omega}_t^\star$ and by the definition of the algorithm in (29), $\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 = \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t(\boldsymbol{\omega}_t^\star)\|^2$. We sum and subtract $\beta\nabla\phi_t(\boldsymbol{\omega}_t^\star)$ and substitute $\mathcal{T}_t^\beta$ to obtain

$$\begin{aligned}&\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2\\ &= \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t(\boldsymbol{\omega}_t^\star) + \beta\nabla\phi_t(\boldsymbol{\omega}_t^\star) - \beta\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|^2\\ &= \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star) - \beta\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|^2.\end{aligned}$$

Expanding the square {1}, expanding $\mathcal{T}_t^\beta$ {2}, and regrouping {3} leads to

$$\begin{aligned}&\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2\\ &\overset{\{1\}}{=} \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \beta^2\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|^2\\ &\quad + 2\langle \mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star), -\beta\nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle\\ &\overset{\{2\}}{=} \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \beta^2\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|^2 -\\ &\quad 2\beta\langle \mathcal{T}_t(\boldsymbol{y}^{(K)}) - \beta\nabla\phi_t(\mathcal{T}_t(\boldsymbol{y}^{(K)})) -\\ &\quad \mathcal{T}_t(\boldsymbol{\omega}_t^\star) + \beta\nabla\phi_t(\mathcal{T}_t(\boldsymbol{\omega}_t^\star)), \nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle\\ &\overset{\{3\}}{=} \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \beta^2\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|^2\\ &\quad + 2\beta\langle \mathcal{T}_t(\boldsymbol{y}^{(K)}) - \boldsymbol{\omega}_t^\star, -\nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle\\ &\quad + 2\beta^2\langle \nabla\phi_t(\mathcal{T}_t(\boldsymbol{y}^{(K)})) - \nabla\phi_t(\boldsymbol{\omega}_t^\star), \nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle.\end{aligned} \tag{51}$$

We note that, by applying the Cauchy–Schwarz, the triangle inequalities and Assumption 14, we have $\langle \nabla\phi_t(\mathcal{T}_t(\boldsymbol{y}^{(K)})) - \nabla\phi_t(\boldsymbol{\omega}_t^\star),\ \nabla\phi_t(\boldsymbol{\omega}_t^\star)\rangle \leq \|\nabla\phi_t(\mathcal{T}_t(\boldsymbol{y}^{(K)})) - \nabla\phi_t(\boldsymbol{\omega}_t^\star)\| \|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\| \leq (U + \|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|)\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|$. By (48) and Lemma 10, we can substitute in (51) the latter relation and the bound in (45) to obtain

$$\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 \leq \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + 6\beta(\xi + \beta U)\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\| + \beta^2(2U + 3\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|)\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\|.$$

Applying Assumption 14 and rearranging the terms leads to

$$\begin{aligned}\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 &\leq \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \\ &\quad 6\beta(\xi + \beta U)\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\| + \beta^2 5U\|\nabla\phi_t(\boldsymbol{\omega}_t^\star)\| \\ &\leq \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \beta(6\xi + 11\beta U)U \\ &\leq \|\mathcal{T}_t^\beta(\boldsymbol{y}^{(K)}) - \mathcal{T}_t^\beta(\boldsymbol{\omega}_t^\star)\|^2 + \tau(\beta)\gamma. \end{aligned} \tag{52}$$

By quasi-nonexpansiveness of $\mathcal{T}_t$ as well as strong monotonicity and Lipschitz continuity of $\nabla\phi_t$, we can apply [30, Lem. 4a] to obtain $\|\mathcal{T}_t^\beta(\boldsymbol{\omega}) - \mathcal{T}_t^\beta(\bar{\boldsymbol{\omega}})\| \leq (1 - \tau(\beta))\|\boldsymbol{\omega} - \bar{\boldsymbol{\omega}}\|$, for all $\boldsymbol{\omega} \in \mathrm{dom}(\mathcal{T}_t^\beta), \bar{\boldsymbol{\omega}} \in \mathrm{fix}(\mathcal{T}_t)$, which we substitute in (52) to obtain

$$\begin{aligned}\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 &\leq (1 - \tau(\beta))^2 \|\boldsymbol{y}^{(K)} - \boldsymbol{\omega}_t^\star\|^2 + \tau(\beta)\gamma \\ &\leq (1 - \tau(\beta)) \|\boldsymbol{y}^{(K)} - \boldsymbol{\omega}_t^\star\|^2 + \tau(\beta)\gamma.\end{aligned}$$

By iterating, we obtain

$$\begin{aligned}&\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 \leq \\ &(1 - \tau(\beta))^2 \|\boldsymbol{y}^{(K-1)} - \boldsymbol{\omega}_t^\star\|^2 + (1 - \tau(\beta))\tau(\beta)\gamma + \tau(\beta)\gamma \\ &\leq \cdots \leq (1 - \tau(\beta))^K \|\boldsymbol{y}^{(1)} - \boldsymbol{\omega}_t^\star\|^2 + \sum_{j=0}^{K-1} (1 - \tau(\beta))^j \tau(\beta)\gamma \\ &\leq (1 - \tau(\beta))^K \|\boldsymbol{y}^{(1)} - \boldsymbol{\omega}_t^\star\|^2 + \sum_{j=0}^{\infty} (1 - \tau(\beta))^j \tau(\beta)\gamma.\end{aligned}$$

Applying the geometric series convergence and recalling from (29) that $\boldsymbol{y}^{(1)} = \boldsymbol{\omega}_t$ leads to (49). ■

The next lemma outlines a contraction property of the restarted HSDM to the solution sequence of Problem (28) up to an additive error, which can be controlled by an appropriate choice of the step size β and the number of iterations K.

Lemma 12: Let Assumptions 12–15 hold. For any $t \in \mathbb{N}$, let $\boldsymbol{\omega}_{t+1}$ be generated by the restarted HSDM algorithm in (29). For any $\gamma > 0$, there exist $K, \beta > 0$, such that (49) holds. □

Proof: Let us consider $\xi := \frac{\gamma\sigma}{12U}$. Since $\mathcal{T}_t$ is quasi-shrinking, the shrinkage function D_t of $\mathcal{T}_t$ satisfies $D_t(\xi) > 0$. Thus, there exist $\bar{\beta} \in (0, \frac{2\sigma}{L_\phi^2})$ and K such that, for any $\beta \in (0, \bar{\beta}]$, (48) holds.

It can be verified that $\lim_{\beta\to 0^+} \frac{\beta}{\tau(\beta)} = \frac{1}{\sigma}$. Then

$$\lim_{\beta\to 0^+} \tfrac{\beta}{\tau(\beta)}(6\xi + 11\beta U)U = \tfrac{6\xi U}{\sigma} = \tfrac{1}{2}\gamma. \tag{53}$$

We thus find $\beta \in (0, \bar{\beta}]$ small enough, such that

$$\tfrac{\beta}{\tau(\beta)}(6\xi + 11\beta U)U \leq \gamma. \tag{54}$$

Hence, the hypothesis holds by invoking Lemma 11. ■

Remark 10: From the proof of Lemma 12, as ξ (and thus $D_t(\xi)$) decreases with γ, it can be seen from (48) that for smaller values of γ a smaller stepsize β and a larger K are necessary. □

B. Proof of Theorem 3

We first construct a suitable stepsize $\bar{\beta}$ and number of iterations $\bar{K}$. We then proceed with proving that the statement holds for the chosen variables. Let us first define the auxiliary variable $\xi = \frac{\gamma\sigma}{12U}$. Following the steps in the proof of Lemma 12, we can choose a small enough $\bar{\beta} \in (0, \min\{\frac{2\sigma}{L_\phi^2}, \frac{D(\xi)}{2U}\})$, where $D(\cdot)$ is defined in Assumption 15, such that

$$\tfrac{\bar{\beta}}{\tau(\bar{\beta})}(6\xi + 11\bar{\beta}U)U \leq \gamma. \tag{55}$$

We now define $\alpha(K) := (1 - \tau(\bar{\beta}))^K$. Since $\tau(\bar{\beta}) \in (0, 1)$, α is decreasing with K. We can then choose K_1, such that $\alpha(K_1) < \frac{1}{2}$. Then, we define the mapping $a : \mathbb{N}_{\geq K_1} \to \mathbb{R}$

$$a(K) = \max\left\{\|\boldsymbol{\omega}_1\| + \sup_{\boldsymbol{\omega}\in\mathcal{Y}}\|\boldsymbol{\omega}\|, \sqrt{\tfrac{2\alpha(K)\delta^2 + \gamma}{1 - 2\alpha(K)}}\right\} \tag{56}$$

We can verify that $a(\cdot)$ is nonincreasing. Consequently, the sequence $(\frac{2(a(K)+\delta)}{K-1})_{K\geq K_1}$ is decreasing. We can then choose any sufficiently large $\bar{K} \geq K_1$, such that

$$D(\xi) \geq \tfrac{2(\bar{a}+\delta)}{\bar{K}-1} \tag{57}$$

where $\bar{a} := a(\bar{K})$. We also define $\bar{\alpha} := \alpha(\bar{K})$. We now prove by induction that

$$\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\| \leq \bar{a} \quad \text{for all } t > 1. \tag{58}$$

To that end, we first show that

$$\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\| \leq \bar{a} \Rightarrow \|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\| \leq \bar{a}. \tag{59}$$

Let us then write

$$\begin{aligned}\mathrm{dist}(\boldsymbol{\omega}_t, \mathrm{fix}(\mathcal{T}_t)) &\overset{\{1\}}{\leq} \|\boldsymbol{\omega}_t - \mathrm{proj}_{\mathrm{fix}(\mathcal{T}_t)}(\boldsymbol{\omega}_{t-1}^\star)\| \\ &\overset{\{2\}}{\leq} \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\| + \|\boldsymbol{\omega}_{t-1}^\star - \mathrm{proj}_{\mathrm{fix}(\mathcal{T}_t)}(\boldsymbol{\omega}_{t-1}^\star)\| \\ &\overset{\{3\}}{\leq} \|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\| + \delta \leq \bar{a} + \delta\end{aligned} \tag{60}$$

where $\{1\}$ follows from the definition of distance, $\{2\}$ from the triangle inequality and $\{3\}$ from Assumption 11 and $\boldsymbol{\omega}_t^\star \in \mathrm{fix}(\mathcal{T}_t)$. Then, by Assumption 15, by the choice $\bar{\beta} \leq \frac{D(\xi)}{2U}$ and (57), it holds that

$$D_t(\xi) \geq \max\left\{2\bar{\beta}U, \tfrac{2(\bar{a}+\delta)}{\bar{K}-1}\right\} \overset{(60)}{\geq} \max\left\{2\bar{\beta}U, \tfrac{2\mathrm{dist}(\boldsymbol{\omega}_t, \mathrm{fix}(\mathcal{T}_t))}{\bar{K}-1}\right\}. \tag{61}$$

By Lemma 11 and (55), we then have

$$\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 \leq \bar{\alpha}\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^\star\|^2 + \gamma. \tag{62}$$

Applying on (62) the triangle inequality, the fact $(a + b)^2 \leq 2a^2 + 2b^2$ and Assumption 11 leads to

$$\begin{aligned}\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 &\leq 2\bar{\alpha}(\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\|^2 + \|\boldsymbol{\omega}_{t-1}^\star - \boldsymbol{\omega}_t^\star\|^2) + \gamma \\ &\leq 2\bar{\alpha}(\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\|^2 + \delta^2) + \gamma \\ &\leq 2\bar{\alpha}(\bar{a}^2 + \delta^2) + \gamma.\end{aligned} \tag{63}$$

Finally, by (56), it holds $\bar{a}^2 \geq \frac{2\bar{\alpha}\delta^2 + \gamma}{1 - 2\bar{\alpha}}$, which implies

$$2\bar{\alpha}(\bar{a}^2 + \delta^2) + \gamma \leq \bar{a}^2. \tag{64}$$

Thus, we obtain $\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 \leq \bar{a}^2$. We now continue the induction argument by proving

$$\|\boldsymbol{\omega}_2 - \boldsymbol{\omega}_1^\star\|^2 \leq \bar{a}^2. \quad (65)$$

From the triangle inequality and from (56), $\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_1^\star\| \leq \|\boldsymbol{\omega}_1\| + \|\boldsymbol{\omega}_1^\star\| \leq \bar{a}$. From the definition of distance, we obtain

$$\mathrm{dist}(\boldsymbol{\omega}_1, \mathrm{fix}(\mathcal{T}_1)) \leq \|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_1^\star\| \leq \bar{a} \leq \bar{a} + \delta. \quad (66)$$

Then, by (57) and the choice $\bar{\beta} \leq \frac{D(\xi)}{2U}$, we have

$$D_t(\xi) \geq D(\xi) \geq \max\left\{2\beta U, \tfrac{2(\bar{a}+\delta)}{K-1}\right\}$$
$$\geq \max\left\{2\beta U, \tfrac{2\mathrm{dist}(\boldsymbol{\omega}_1, \mathrm{fix}(\mathcal{T}_1))}{K-1}\right\}.$$

By Lemma 11 and (55), we find

$$\|\boldsymbol{\omega}_2 - \boldsymbol{\omega}_1^\star\|^2 \leq \bar{\alpha}\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_1^\star\|^2 + \gamma.$$

We then upper bound the right-hand side of the last inequality:

$$\|\boldsymbol{\omega}_2 - \boldsymbol{\omega}_1^\star\|^2 \overset{(66)}{\leq} \bar{\alpha}\bar{a}^2 + \gamma \leq \bar{\alpha}(2\bar{a}^2 + 2\delta^2) + \gamma \overset{(64)}{\leq} \bar{a}^2.$$

Therefore, combining (59) and (65) leads to $\sup_{t>1}\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1}^\star\| \leq \bar{a}$. Recalling that, from Assumption 13, $\boldsymbol{\omega}_t^\star \in \mathcal{Y}$ for all t, this immediately implies $\mathrm{dist}(\boldsymbol{\omega}_t, \mathcal{Y}) \leq \bar{a}$ for all $t > 1$, which proves that the sequence is bounded.

We now proceed with proving (30). We note that the relation in (62) holds for all t. We then observe that, by the triangle inequality, by $(a+b)^2 \leq 2a + 2b$, and by Assumption 11

$$\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_{t+1}^\star\|^2 \leq 2\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 + 2\|\boldsymbol{\omega}_{t+1}^\star - \boldsymbol{\omega}_t^\star\|^2$$
$$\leq 2\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2 + 2\delta^2.$$

By using (62) to upper bound $\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_t^\star\|^2$ and iterating, we find

$$\|\boldsymbol{\omega}_{t+1} - \boldsymbol{\omega}_{t+1}^\star\|^2 \leq 2\bar{\alpha}\|\boldsymbol{\omega}_t - \boldsymbol{\omega}_t^\star\|^2 + 2(\gamma + \delta^2)$$
$$\leq (2\bar{\alpha})^2\|\boldsymbol{\omega}_{t-1} - \boldsymbol{\omega}_{t-1}^\star\|^2 + 2(\gamma + \delta^2) + 2\bar{\alpha}(2\gamma + 2\delta^2)$$
$$\leq \cdots \leq (2\bar{\alpha})^t\|\boldsymbol{\omega}_1 - \boldsymbol{\omega}_1^\star\|^2 + \textstyle\sum_{j=0}^{t-1}(2\bar{\alpha})^j(2\gamma + 2\delta^2).$$

By taking the limit for $t \to \infty$ and by applying the convergence of the geometric sequence, we obtain (30). ■

C. Proof of Corollary 1

Steps i–vi of Algorithm 3 are analogous to Steps 1–6 of Algorithm 1. Analogously to the proof of Theorem 1, we see that the variable $\boldsymbol{y}^{(k)} := (\hat{x}_i^{(k)}, \hat{\lambda}_i^{(k)}, \hat{\nu}_i^{(k)})$ is updated at each time step by K iterations of the HSDM

$$\boldsymbol{y}^{(k+1)} = \mathcal{T}_{\mathrm{FBF},t}(\boldsymbol{y}^{(k)}) - \beta\nabla\phi_t(\mathcal{T}_{\mathrm{FBF},t}(\boldsymbol{y}^{(k)})),\ k = 1, \ldots, K.$$

Then, the variable $\boldsymbol{\omega}_{t+1}$ is updated as $\boldsymbol{\omega}_{t+1} = \boldsymbol{y}^{(K+1)}$. Thus, Algorithm 3 is a particular instance of the restarted HSDM algorithm (29). By Theorem 3, the tracking error is given by (30). ■

References

[1] I. Shilov, H. Le Cadre, and A. Busic, “Privacy impact on generalized Nash equilibrium in peer-to-peer electricity market,” *Operations Res. Lett.*, vol. 49, no. 5, pp. 759–766, 2021.

[2] G. Belgioioso, W. Ananduta, S. Grammatico, and C. Ocampo-Martinez, “Operationally-safe peer-to-peer energy trading in distribution grids: A game-theoretic market-clearing mechanism,” *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2897–2907, Jul. 2022.

[3] B. G. Bakhshayesh and H. Kebriaei, “Decentralized equilibrium seeking of joint routing and destination planning of electric vehicles: A constrained aggregative game approach,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 13265–13274, Aug. 2021, doi: 10.1109/TITS.2021.3123207.

[4] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, “Distributed power allocation with rate constraints in Gaussian parallel interference channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3471–3489, Aug. 2008.

[5] G. Scutari, D. P. Palomar, F. Facchinei, and J.-S. Pang, “Monotone Games for Cognitive Radio Systems,” in *Distributed Decision Making and Control*. London:Springer2012, pp. 83–112.

[6] J. Wang, M. Peng, S. Jin, and C. Zhao, “A generalized Nash equilibrium approach for robust cognitive radio networks via generalized variational inequalities,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3701–3714, Jul. 2014.

[7] F. Facchinei and C. Kanzow, “Generalized Nash equilibrium problems,” *Ann. Operations Res.*, vol. 175, no. 1, pp. 177–211, 2010.

[8] F. Facchinei, A. Fischer, and V. Piccialli, “On generalized Nash games and variational inequalities,” *Operations Res. Lett.*, vol. 35, no. 2, pp. 159–164, 2007.

[9] A. A. Kulkarni and U. V. Shanbhag, “On the variational equilibrium as a refinement of the generalized Nash equilibrium,” *Automatica*, vol. 48, no. 1, pp. 45–55, 2012.

[10] 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. 2018.

[11] G. Belgioioso and S. Grammatico, “Projected-gradient algorithms for generalized equilibrium seeking in aggregative games are preconditioned forward-backward methods,” in *Proc. Eur. Control Conf.*, 2018, pp. 2188–2193.

[12] H. Yin, U. V. Shanbhag, and P. G. Mehta, “Nash equilibrium problems with scaled congestion costs and shared constraints,” *IEEE Trans. Autom. Control*, vol. 56, no. 7, pp. 1702–1708, Jul. 2011.

[13] G. Belgioioso and S. Grammatico, “Semi-decentralized Nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions,” *IEEE Contr. Syst. Lett.*, vol. 1, no. 2, pp. 400–405, Oct. 2017.

[14] 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.

[15] P. Yi and L. Pavel, “An operator splitting approach for distributed generalized Nash equilibria computation,” *Automatica*, vol. 102, pp. 111–121, 2019.

[16] P. Yi and L. Pavel, “Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms,” *IEEE Trans. Control Netw. Syst.*, vol. 6, no. 1, pp. 299–311, Mar. 2019.

[17] B. Franci, M. Staudigl, and S. Grammatico, “Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking,” in *Proc. Eur. Control Conf.*, 2020, pp. 1274–1279.

[18] 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.

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

[20] J. R. Marden and T. Roughgarden, “Generalized efficiency bounds in distributed resource allocation,” *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 571–584, Mar. 2014.

[21] G. Scutari, F. Facchinei, J.-S. Pang, and L. Lampariello, “Equilibrium Selection in Power Control Games on the Interference Channel,” in *Proceedings IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 675–683.

[22] G. Scutari, F. Facchinei, J.-S. Pang, and D. P. Palomar, “Real and complex monotone communication games,” *IEEE Trans. Inf. Theory*, vol. 60, pp. 4197–4231, Jul. 2014.

[23] A. Dreves, “How to select a solution in generalized Nash equilibrium problems,” *J. Optim. Theory Appl.*, vol. 178, no. 3, pp. 973–997, 2018.

[24] A. Dreves, “An algorithm for equilibrium selection in generalized Nash equilibrium problems,” *Comput. Optim. Appl.*, vol. 73, no. 3, pp. 821–837, 2019.

[25] H. D. Kaushik and F. Yousefian, "A method with convergence rates for optimization problems with variational inequality constraints," *SIAM J. Optim.*, vol. 31, no. 3, pp. 2171–2198, 2021.
[26] A. Simonetto, E. Dall'Anese, S. Paternain, G. Leus, and G. B. Giannakis, "Time-varying convex optimization: Time-structured algorithms and applications," *Proc. IEEE*, vol. 108, no. 11, pp. 2032–2048, Nov. 2020.
[27] K. Lu, G. Li, and L. Wang, "Online distributed algorithms for seeking generalized Nash equilibria in dynamic environments," *IEEE Trans. Autom. Control*, vol. 66, no. 5, pp. 2289–2296, May 2021.
[28] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. New York, NY, USA: Springer, 2007.
[29] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York, NY, USA: Springer, 2011.
[30] I. Yamada and N. Ogura, "Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings," *Numer. Funct. Anal. Optim.*, vol. 25, no. 7/8, pp. 619–655, 2005.
[31] H. Xu and T. Kim, "Convergence of hybrid steepest-descent methods for variational inequalities," *J. Optim. Theory Appl.*, vol. 119, no. 1, pp. 185–201, 2003.
[32] A. Cegielski and R. Zalas, "Properties of a class of approximately shrinking operators and their applications," *Fixed Point Theory*, vol. 15, no. 2, pp. 399–426, 2014.
[33] A. Simonetto, A. Koppel, A. Mokhtari, G. Leus, and A. Ribeiro, "Decentralized Prediction-Correction Methods for Networked Time-Varying Convex Optimization," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5724–57381, Nov. 2017, doi: 10.1109/TAC.2017.2694611.
[34] A. Cegielski, A. Gibali, S. Reich, and R. Zalas, "An algorithm for solving the variational inequality problem over the fixed point set of a quasi-nonexpansive operator in euclidean space," *Numer. Funct. Anal. Optim.*, vol. 34, no. 10, pp. 1067–1096, 2013.
[35] D. P. Palomar and Y. C. Eldar, *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[36] A. Auslender and M. Teboulle, "Lagrangian duality and related multiplier methods for variational inequality problems," *SIAM J. Optim.*, vol. 10, no. 4, pp. 1097–1115, 2000.
[37] Y. Malitsky and M. K. Tam, "A forward-backward splitting method for monotone inclusions without cocoercivity," *SIAM J. Optim.*, vol. 30, no. 2, pp. 1451–1472, 2020.
[38] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 431–446, 2000.
[39] E. Dall'Anese and A. Simonetto, "Optimal power flow pursuit," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 942–952, Mar. 2018.
[40] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*. New York, NY, USA: Springer, 2014.
[41] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *Proc. 20th Int. Conf. Mach. Learn.*, 2003, pp. 928–936.
[42] N. Ogura and I. Yamada, "Nonstrictly convex minimization over the bounded fixed point set of a nonexpansive mapping," *Numer. Funct. Anal. Optim.*, vol. 24, no. 1/2, pp. 129–135, 2003.

Emilio Benenati received the bachelor's degree in electrical engineering from the University of Catania, Catania, Italy, and the master's degree in robotics, systems and control from ETH Zürich, Zürich, Switzerland, in 2016 and 2019, respectively. He is currently working toward the Ph.D. degree in automatic control with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands.

From 2019 to 2020, he was a Research Fellow with the Department of Artificial and Mechanical Intelligence, Italian Institute of Technology, Genova, Italy.

Wicak Ananduta received the bachelor's degree in electrical engineering from the University of Indonesia, Jakarta, Indonesia, the master's degree in systems and control from Delft University of Technology (TU Delft), Delft, The Netherlands, and the Ph.D. degree in automatic control, robotics and vision from Universitat Politècnica de Catalunya, Barcelona, Spain, in 2011, 2016, and 2019, respectively.

He is a Postdoctoral Researcher with the Delft Center for Systems and Control, TU Delft.

Sergio Grammatico received the bachelor's degree in computer engineering and the master's and Ph.D. degrees in automatic control from the University of Pisa, Pisa, Italy, in 2008, 2009, and 2013, respectively.

He is an Associate Professor with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands. From 2012 to 2017, he held research positions with the University of California Santa Barbara, CA, USA, with ETH Zürich, Zürich, Switzerland, and with TU Eindhoven, Eindhoven, The Netherlands.

Dr. Grammatico was the recipient of the 2021 Roberto Tempo Best Paper Award. He is currently an Associate Editor for IEEE TRANSACTIONS ON AUTOMATIC CONTROL and IFAC *Automatica*.