Linear convergence in (stochastic) generalized Nash equilibrium problems with coupling inequality constraints

Linear Convergence in (Stochastic) Generalized Nash Equilibrium Problems With Coupling Inequality Constraints

Barbara Franci and Mattia Bianchi, *Member, IEEE*

***Abstract*—We prove the linear convergence of a forward-backward (FB) algorithm for solving generalized Nash equilibrium problems (GNEPs) with full rank coupling inequality constraints. The analysis relies on a contraction argument in an appropriate weighted space, and leverages partial contractivity properties for the forward and backward operators separately. This differs from existing approaches, which can only deal with equality constraints. Contractivity of the full FB algorithm then implies linear convergence of the iterates to the unique fixed point, i.e., to a generalized Nash equilibrium. The contraction perspective also allows us to extend the analysis to the case of stochastic GNEP, where we provide a novel linear convergence result for stochastic games with coupling inequality constraints.**

***Index Terms*—Linear convergence, generalized Nash equilibrium problems, optimization algorithms.**

I. Introduction

GENERALIZED Nash equilibrium problems (GNEPs) [1] and their stochastic counterpart (SGNEPs) [2] model competitive scenarios where multiple agents aim at minimizing their individual objective functions, while taking into account the other participants' behavior and shared feasibility constraints. These problems are motivated by applications in power grids [3], traffic networks [4], ride-hail markets [5], to name a few. The recent literature has seen a surge of algorithms to find a generalized Nash equilibrium (GNE), in a semi-decentralized [6] or distributed [7], [8] fashion. Most methods leverage adaptations of the classic gradient descent mechanism [1], [9], [10]. Analogous developments can be found for the stochastic case, where the agents minimize an expected value function by computing a stochastic approximation of the associated pseudogradient mapping [11], [12]. In both scenarios, to cope with the presence of coupling constraints and split the computation among the agents, the use of Lagrangian methods is the gold standard. In fact, all the references above leverage primal-dual algorithms to solve the Karush–Kuhn–Tucker (KKT) optimality conditions of the GNE problem. Then, one can exploit the extensive literature on variational inequalities and operator theoretic methods for algorithm design and convergence proofs [9], [11], [13], [14]. This letter also follows this prevalent trend.

Specifically, we are interested in establishing linear convergence rates, in generalized games with inequality coupling constraints. Linear convergence was extensively studied for games without coupling constraints, under the assumption of strong monotonicity of the pseudogradient mapping [15], [16], also in variance-reduced stochastic schemes [17], [18]. Some results are also available for (proximal) best response dynamics, both for the deterministic [19] and stochastic case [20]. However, this literature does not consider coupling constraints.

Results for generalized games are much sparser. The critical complication with coupling constraints is that, even when the pseudogradient is strongly monotone, the KKT operator is not strongly monotone in general. One exception is the case of games with full rank equality constraints: in this case, it was shown in [15] that the KKT operator is strongly monotone in a suitable weighted space, which directly yields linear convergence for many established methods. The case of inequality constraints, even if full rank, is more complex, for two reasons: (i) the KKT operator is not strongly monotone; (ii) the extra nonsmooth term makes it impossible to extend the existing arguments based on weighted strong monotonicity. The only linear rate in the case of inequality constraints we are aware of is [21], which however directly assumes that the KKT operator is strongly monotone, which is a restrictive and somewhat abstract condition.

Contributions: We study a primal-dual forward-backward (FB) method for solving GNEP, under the assumptions of strongly monotone pseudogradient and full rank inequality coupling constraints. We show linear convergence to the unique variational GNE (v-GNE), for both the deterministic and stochastic case. To the best of our knowledge, we are the first study linear convergence under these assumptions, and without extra conditions, cf. [21]. In particular, we prove the following:

- The FB dynamics are contractive, with respect to a suitable weighted Euclidean norm. This also implies linear convergence to the unique v-GNE.

Received 17 March 2026; revised 25 May 2026; accepted 12 June 2026. Date of publication 24 June 2026; date of current version 8 July 2026. The work of Barbara Franci was supported by the "Programma Rita Levi Montalcini" of Italian Ministry of University and Research. This article has not been presented at a conference. Recommended by Senior Editor K. Savla. *(Corresponding author: Barbara Franci.)*

Barbara Franci is with the Department of Mathematical Sciences, Politecnico di Torino, 10129 Turin, Italy (e-mail: barbara.franci@polito.it).

Mattia Bianchi is with the Automatic Control Laboratory, ETH Zürich, 8092 Zürich, Switzerland (e-mail: mbianch@ethz.ch).

Digital Object Identifier 10.1109/LCSYS.2026.3706919

- Linear convergence also holds for the stochastic case, in the variance reduction regime [18]—where the true pseudogradient is replaced by the average of an increasingly large batch of sampled gradients.

Technically, we show contractivity of the FB algorithm, even though backward-step and forward-step are not separately contractive.The result follows from certain partial contractivity properties for the forward and backward operators, in different norms. Specifically, our approach builds on [22] which deals with saddle-point problems, but extends it to the case of game problems (e.g., resulting in different contractivity and stepsizes bounds); furthermore, we adapt the weighted contractivity argument to the stochastic case.

Notation: The standard inner product is indicated with $\langle \cdot,\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $\langle x, y\rangle = x^\top y$, while $\|\cdot\|$ is the associated Euclidean norm. We indicate a positive definite matrix A, i.e., $x^\top Ax > 0$, with $A \succ 0$. Given a matrix $\Phi \succ 0$, we indicate the inner product induced by Φ as $\langle x, y\rangle_\Phi = \langle \Phi x, y\rangle$ and its associated Φ-induced norm with $\|x\|_\Phi = \sqrt{\langle \Phi x, x\rangle}$. $\mathbf{0}_m$ denotes the vector in $\mathbb{R}^m$ with all entries equal to zero. The projection onto a closed convex set $C \subseteq \mathbb{R}^n$ is $\mathrm{proj}_C(x) = \mathrm{argmin}_{y\in C}\|y-x\|$. The normal cone to C is the set $\mathrm{N}_C(x) = \{u \in \mathbb{R}^n \mid \sup_{y\in C}\langle y-x \mid u\rangle \leqslant 0\}$ if $x \in C$ and $\mathrm{N}_C(x) = \emptyset$ otherwise. $\mathrm{J}_F = (\mathrm{Id} + F)^{-1}$ is the resolvent of the operator $F : \mathbb{R}^n \to \mathbb{R}^n$ where Id indicates the identity operator. A mapping $F : \mathrm{dom}F \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is ℓ-Lipschitz continuous if, for some $\ell > 0$, $\|F(x)-F(y)\| \leq \ell\|x-y\|$ for all $x, y \in \mathrm{dom}(F)$; monotone if $\langle F(x) - F(y), x - y\rangle \geq 0$ for all $x, y \in \mathrm{dom}(F)$; μ-strongly monotone if, for some $\mu > 0$, $\langle F(x) - F(y), x - y\rangle \geq \mu\|x - y\|^2$ for all $x, y \in \mathrm{dom}(F)$. We postulate Standing Assumptions to hold throughout the letter, while we reference Assumptions in the sections where they are required.

II. Problem Description

A. Generalized Nash Equilibrium Problem

We consider a set of agents $\mathcal{N} = \{1,\dots,N\}$ involved in a game, where each agent i chooses an action $x_i \in \mathbb{R}^{n_i}$, with the aim of minimizing the cost function $J_i(x_i, x_{-i})$, $J_i : \mathbb{R}^n \to \mathbb{R}$, where $x_{-i} = \mathrm{col}(\{x_j\}_{j\neq i})$ is the vector collecting the decision variables of the opponents of agent i and $n = \sum_{i\in\mathcal{N}} n_i$. Besides being coupled through their costs, the agents face linear coupling constraints: $\mathcal{X} := \{x \in \mathbb{R}^n \mid Ax - b \leq \mathbf{0}_m\}$, where $A := [A_1,\dots,A_N] \in \mathbb{R}^{m\times n}$ and $b := \sum_{i=1}^N b_i \in \mathbb{R}^m$. Therefore, each agent has a feasible set of the form $\mathcal{X}_i(x_{-i}) := \left\{y_i \in \mathbb{R}^{n_i} \mid A_iy_i \leq b - \sum_{j\neq i}^N A_jx_j\right\}$, which depends on the decision variable of the other agents. In summary, the game we consider in this work is a collection of optimization problems:

$$\forall i \in \mathcal{N} : \quad \begin{cases} \min_{x_i\in\mathbb{R}^{n_i}} & J_i(x_i, x_{-i}) \\ \text{s.t.} & Ax \leq b. \end{cases} \tag{1}$$

The solution concept we consider is that of a generalized Nash equilibrium, i.e., a collective strategy $x^\star = \mathrm{col}(x_1^\star,\dots,x_N^\star) \in \mathcal{X}$ such that for all $i \in \mathcal{N}$,

$$J_i(x_i^\star, x_{-i}^\star) \leq \inf\{J_i(y_i, x_{-i}^\star) \mid y_i \in \mathcal{X}_i(x_{-i}^\star)\}.$$

We postulate standard regularity and convexity assumptions.

Standing Assumption 1 (Convexity): The set $\mathcal{X}$ is nonempty. For each $i \in \mathcal{N}$, J_i is continuous and the function $J_i(\cdot, x_{-i})$ is convex and continuously differentiable for all $x_{-i} \in \mathbb{R}^{n-n_i}$. □

Under this assumption, the GNEs of the game can be characterized by the KKT conditions of the problems in (1) [13]. As standard in the literature, we focus on v-GNEs, where the agents share the same dual variable, and which are more economically justifiable and computationally tractable [23]. In particular, a collective action $x^\star \in \mathbb{R}^n$ is a v-GNE if and only if there is a dual variable $\lambda^\star \in \mathbb{R}^m$ such that

$$\forall i \in \mathcal{N} : \begin{cases} \mathbf{0}_{n_i} \in \nabla_{x_i}J_i(x_i^\star, x_{-i}^\star) + A_i^\top\lambda^\star \\ \mathbf{0}_m \in \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^\star) - (Ax^\star - b). \end{cases} \tag{2}$$

Let us write these KKT conditions in compact form as $\mathbf{0} \in \mathcal{A}(\omega^\star)$, where $\omega = \mathrm{col}(x, \lambda)$,

$$\mathcal{A}(\omega) = \underbrace{\begin{bmatrix} F(x) \\ 0 \end{bmatrix}}_{:=\mathcal{A}_\mathrm{f}} + \underbrace{\begin{bmatrix} A^\top\lambda \\ -(Ax - b) + \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda) \end{bmatrix}}_{:=\mathcal{A}_\mathrm{b}}, \tag{3}$$

and the pseudogradient mapping associated to the game is

$$F(x) := \mathrm{col}(\nabla_{x_1}J_1(x_1, x_{-1}),\dots,\nabla_{x_N}J_N(x_N, x_{-N})). \tag{4}$$

Standing Assumption 2 (Strong monotonicity): F is ℓ-Lipschitz continuous and μ_F-strongly monotone. □

The strong monotonicity assumption is very common for algorithms with linear convergence [1], [11], [22].

Standing Assumption 3 (Full-rank constraints): A is full-row rank, namely $\mu_A = \lambda_{\min}(AA^\top) > 0$. □

The previous condition is well-known in duality theory, and it was for instance exploited for GNEPs with equality constraints [15]. Together with Standing Assumption 2, it implies that there is a unique $\omega^\star = (x^\star, \lambda^\star)$ that satisfies the KKT conditions (2), as it can be inferred from (2) [13], [24].

Remark 1 (On the full-rank assumption): Standing Assumption 3 implies $n \geq m$ and thus that the set $\{x \mid Ax \leq b\}$ is unbounded. While this is restrictive, such full-row-rank coupling inequality constraints are ubiquitous in engineering problems: they model the maximal total energy that a charging station can deliver to multiple electric vehicles [6, §5] ($A = [1\dots1]$); link capacity constraints in communication problems [9, §4]; or market capacity in the classical Cournot game (see Section V). We emphasize that we do not consider local feasible sets in (1): these can be addressed as soft constraints via barrier functions, as shown in [15] and further illustrated in Section V. □

B. The FB Algorithm

In this letter, we consider the following common GNE seeking algorithm for the game in (1):

$$\begin{aligned} x^{k+1} &= x^k - \alpha\left[F(x^k) + A^\top\lambda^k\right] \\ \lambda^{k+1} &= \mathrm{proj}_{\mathbb{R}_{\geq 0}}\left(\lambda^k + \beta\left[2Ax^{k+1} - Ax^k - b\right]\right). \end{aligned} \tag{5}$$

with stepsizes $\alpha,\beta > 0$. It is known that this iteration can be equivalently written as a preconditioned FB scheme [25]:

$$\omega^{k+1} = \mathcal{B}(\mathcal{F}(\omega^k)), \tag{6}$$

featuring a *forward* operator $\mathcal{F}$ and a *backward* operator $\mathcal{B}$,

$$\mathcal{F} = \mathrm{Id} - \Phi^{-1}\mathcal{A}_{\mathrm{f}}, \tag{7a}$$

$$\mathcal{B} = (\mathrm{Id} + \Phi^{-1}\mathcal{A}_{\mathrm{b}})^{-1}, \tag{7b}$$

where $\mathcal{A}_{\mathrm{f}}$ and $\mathcal{A}_{\mathrm{b}}$ are defined in (3), and

$$\Phi = \begin{bmatrix} \frac{1}{\alpha}I_n & -A^\top \\ -A & \frac{1}{\beta}I_m \end{bmatrix}. \tag{8}$$

Lemma 1 [25]: A pair $(x^\star, \lambda^\star)$ is a fixed point for the iteration (5) if and only if it solves the KKT conditions (2); in particular, $x^\star$ is the v-GNE of the game (1).

The main goal in this letter is to show that (6) is contractive, as done in Section III and informally summarized next, for readability.

Theorem 1: For small enough stepsizes $\alpha, \beta > 0$, the iteration in (6) is contractive: there are a matrix $\Psi \succ 0$ and a constant $\rho \in (0, 1)$, such that, for any $\omega, \omega' \in \mathbb{R}^{n+m}$

$$\|\mathcal{B}(\mathcal{F}(\omega)) - \mathcal{B}(\mathcal{F}(\omega'))\|_\Psi \leq \rho\|\omega - \omega'\|_\Psi. \tag{9}$$

By choosing $\omega' = \omega^\star = \mathcal{B}(\mathcal{F}(\omega^\star))$, this implies linear convergence of the algorithm (5) to the unique fixed point $\omega^\star = (x^\star, \lambda^\star)$, and in particular of x^k to the unique v-GNE.

Remark 2: The x-update in (5) can be split agent-wise, enabling, for example, a semi-decentralized implementation [6]. This is enabled by the primal-dual formulation (2). In fact, primal-dual algorithms such as (5) form the foundations for GNE seeking methods [6], [8]. In contrast, primal-only methods, such as the projected pseudogradient method $x^{k+1} = \mathrm{proj}_{\mathcal{X}}(x^k - \alpha F(x^k))$ (for which proving linear convergence is trivial), require a projection on the set $\mathcal{X}$, which is a non-separable operation. For this reason we focus on primal-dual algorithms here. □

III. Contractivity Analysis

This section is dedicated to the contractivity analysis of the iterative scheme in (5). We start with a remark on the preconditioning matrix Φ.

Lemma 2 (Properties of Φ): If $\alpha\beta\|A\|^2 < 1$, then $\Phi \succ 0$. Moreover,

$$\lambda_{\max}(\Phi) = \|\Phi\| \leq \zeta_{\alpha,\beta} := \max\left\{\frac{1}{\alpha}, \frac{1}{\beta}\right\} + \|A\|. \tag{10}$$

Proof: By the Schur complement, $\Phi \succ 0$ iff $\frac{1}{\alpha}I_n - A^\top(\beta I_m)A = \left(\frac{1}{\alpha}I_n - \beta A^\top A\right) \succ 0$, which is ensured by $\alpha\beta\|A\|^2 < 1$. Since Φ is symmetric, $\|\Phi\| = \lambda_{\max}(\Phi)$. Write $\Phi = D + S$ with $D = \mathrm{diag}\left(\frac{1}{\alpha}I_n, \frac{1}{\beta}I_m\right)$ and $S = [0, -A^\top; -A, 0]$. Then $\|\Phi\| \leq \|D\| + \|S\| = \max\left\{\frac{1}{\alpha}, \frac{1}{\beta}\right\} + \|A\|$, because $\|S\| = \sqrt{\lambda_{\max}(AA^\top)} = \|A\|$. ■

A. Partial Contractivity of the Backward Step

The contractivity properties of the operator $\mathcal{B}$ in (7b) depend on the matrix A, i.e., on the coupling constraints.

Lemma 3 (Partial contractivity of the backward step): Assume $\Phi \succ 0$ and define $\zeta_{\alpha,\beta}$ as in (10). Let $\mathcal{B}$ be as in (7b) and define the matrix

$$\Psi_b := \mathrm{diag}(0_{n\times n}, \gamma_\lambda I_m), \quad \gamma_\lambda := \frac{\mu_A}{\zeta_{\alpha,\beta}} > 0. \tag{11}$$

Then, for all $\omega, \omega' \in \mathbb{R}^{n+m}$, it holds that

$$\|\mathcal{B}(\omega) - \mathcal{B}(\omega')\|^2_{\Phi+\Psi_b} \leq \|\omega - \omega'\|^2_\Phi. \tag{12}$$

Proof: Let $u := \mathcal{B}(\omega)$ and $u' := \mathcal{B}(\omega')$. By definition of resolvent, there exist $t \in \mathcal{A}_{\mathrm{b}}(u)$ and $t' \in \mathcal{A}_{\mathrm{b}}(u')$ such that

$$\omega = u + \Phi^{-1}t, \qquad \omega' = u' + \Phi^{-1}t'. \tag{13}$$

Let $u = (x, \lambda)$, $u' = (x', \lambda')$ and select $\nu \in \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda)$, $\nu' \in \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda')$ so that $t = [A^\top\lambda, -(Ax - b) + \nu]^\top$ and $t' = [A^\top\lambda', -(Ax' - b) + \nu']^\top$.

$$\begin{aligned}\|\omega - \omega'\|^2_\Phi &\\ &= \|u - u'\|^2_\Phi + \|\Phi^{-1}(t - t')\|^2_\Phi + 2\langle u - u', t - t'\rangle \\ &= \|u - u'\|^2_\Phi + (t - t')^\top\Phi^{-1}(t - t') + 2\langle u - u', t - t'\rangle.\end{aligned} \tag{14}$$

We next lower bound the second and third terms. Since $\Phi^{-1} \succ 0$, we have

$$\begin{aligned}(t - t')^\top\Phi^{-1}(t - t') &\geq \lambda_{\min}(\Phi^{-1})\|t - t'\|^2 \\ &= \tfrac{1}{\lambda_{\max}(\Phi)}\|t - t'\|^2.\end{aligned} \tag{15}$$

Moreover, by construction,

$$t - t' = \begin{bmatrix} A^\top(\lambda - \lambda') \\ -A(x - x') + (\nu - \nu') \end{bmatrix}, \tag{16}$$

hence $\|t - t'\|^2 \geq \|A^\top(\lambda - \lambda')\|^2$. Combining with (15) and Lemma 2 yields $(t - t')^\top\Phi^{-1}(t - t') \geq \frac{1}{\zeta_{\alpha,\beta}}\|A^\top(\lambda - \lambda')\|^2$.

Further, using the expressions in (16)

$$\begin{aligned}\langle u - u', t - t'\rangle &= \langle x - x', A^\top(\lambda - \lambda')\rangle \\ &\quad + \langle \lambda - \lambda', -A(x - x') + (\nu - \nu')\rangle \\ &= \underbrace{\langle x - x', A^\top(\lambda - \lambda')\rangle - \langle \lambda - \lambda', A(x - x')\rangle}_{=0} \\ &\quad + \langle \lambda - \lambda', \nu - \nu'\rangle \geq 0,\end{aligned}$$

where the last inequality follows from the monotonicity of the normal cone $\mathrm{N}_{\mathbb{R}^m_{\geq 0}}$.

Plugging the obtained bounds into (14) and using the full row rank assumption (Standing Assumption 3) gives $\|\omega - \omega'\|^2_\Phi \|u - u'\|^2_\Phi + \frac{1}{\zeta_{\alpha,\beta}}\|A^\top(\lambda - \lambda')\|^2 \ \|u - u'\|^2_\Phi + \frac{\mu_A}{\zeta_{\alpha,\beta}}\|\lambda - \lambda'\|^2$. With γ_λ as in (11), this is exactly $\|\omega - \omega'\|^2_\Phi \geq \|u - u'\|^2_\Phi + \|u - u'\|^2_{\Psi_b} = \|u - u'\|^2_{\Phi+\Psi_b}$, which is (12) since $u = \mathcal{B}(\omega)$ and $u' = \mathcal{B}(\omega')$. ■

B. Partial Contractivity of the Forward Step

The contractivity properties of the forward step depend on the strong-monotonicity of the pseudogradient mapping.

Lemma 4 (Contractivity of the forward step): Assume that $\alpha, \beta > 0$ and

$$\alpha < \frac{2\mu_F}{\ell^2 + 2\mu_F\beta\|A\|^2}. \tag{17}$$

Let $\mathcal{F}$ be as in (7) and

$$\Psi_f := \mathrm{diag}(\gamma_x I_n, 0_{m\times m}), \quad \gamma_x := 2\mu_F - \frac{\alpha\ell^2}{1 - \alpha\beta\|A\|^2} > 0. \tag{18}$$

Then, for all $\omega, \omega' \in \mathbb{R}^{n+m}$, it holds that

$$\|\mathcal{F}(\omega) - \mathcal{F}(\omega')\|^2_\Phi \leq \|\omega - \omega'\|^2_{\Phi-\Psi_f}. \tag{19}$$

Proof: Since $\alpha\beta\|A\|^2 < 1$ by (17), $\Phi \succ 0$ by Lemma 2. Let $\omega = (x, \lambda)$ and $\omega' = (x', \lambda')$, and $p := [F(x) - F(x'), 0]^\top$. Since $\mathcal{F} = \mathrm{Id} - \Phi^{-1}\mathcal{A}_f$ and $\mathcal{A}_f(\omega) = (F(x), 0)$, we have

$$\begin{aligned}\|\mathcal{F}(\omega) - \mathcal{F}(\omega')\|_\Phi^2 &= \|(\omega - \omega') - \Phi^{-1} p\|_\Phi^2 \\ &= \|\omega - \omega'\|_\Phi^2 - 2\langle p,\, \omega - \omega'\rangle + p^\top \Phi^{-1} p.\end{aligned} \tag{20}$$

By Standing Assumption 2, we have

$$\langle p,\, \omega - \omega'\rangle \;\geq\; \mu_F \|x - x'\|^2. \tag{21}$$

Moreover, from the Schur complement $S := \frac{1}{\beta} I_m - \alpha A A^\top \succ 0$ since $\alpha\beta\|A\|^2 < 1$. The block inversion formula gives the upper-left block of Φ^{-1} as $(\Phi^{-1})_{xx} = \alpha I_n + \alpha^2 A^\top S^{-1} A$. Since p has zero λ-component, we obtain

$$\begin{aligned}p^\top \Phi^{-1} p &= (F(x) - F(x'))^\top (\Phi^{-1})_{xx} (F(x) - F(x')) \\ &= (F(x) - F(x'))^\top \left(\alpha I_n + \alpha^2 A^\top S^{-1} A\right) \cdot \\ &\quad \cdot (F(x) - F(x')).\end{aligned} \tag{22}$$

We upper bound the spectral norm of $(\Phi^{-1})_{xx}$:

$$\begin{aligned}\lambda_{\max}\left((\Phi^{-1})_{xx}\right) &\leq \alpha + \alpha^2 \lambda_{\max}(A^\top S^{-1} A) \\ &\leq \alpha + \alpha^2\|A\|^2 \lambda_{\max}(S^{-1}) = \alpha + \alpha^2 \|A\|^2 \frac{1}{\lambda_{\min}(S)} \\ &\leq \alpha + \alpha^2\|A\|^2 \frac{1}{\frac{1}{\beta} - \alpha\|A\|^2} = \frac{\alpha}{1 - \alpha\beta\|A\|^2}.\end{aligned}$$

Using this in (22) and Lipschitz continuity (Standing Assumption 2) yields

$$\begin{aligned}p^\top \Phi^{-1} p &\leq \frac{\alpha}{1 - \alpha\beta\|A\|^2} \|(F(x) - F(x'))\|^2 \\ &\leq \frac{\alpha \ell^2}{1 - \alpha\beta\|A\|^2} \|x - x'\|^2.\end{aligned} \tag{23}$$

Plugging (21) and (23) into (20) gives the thesis. ■

Remark 3: Condition (17) on the stepsize is in line with the usual bounds for FB schemes, proportional to μ/ℓ^2 for linear convergence [16], and here adjusted to account for the coupling constraints. While achieving linear convergence with vanishing stepsizes is excluded, considering an adaptive rule similar to [26] could be a valid alternative to ensure contractivity without assuming bounds that depends on the global Lipschitz and strong monotonicity constants. □

C. Contractivity of the Full Iteration

To show contractivity of the FB iteration in (6), we use the previous two results and the following auxiliary one.

Proposition 1: Let $\Phi \succ 0$ and let $\Psi_b \succeq 0$, $\Psi_f \succeq 0$ be symmetric matrices. Assume that for some $\gamma \in (0, 1]$ the following hold for all $\omega, \omega' \in \mathbb{R}^{n+m}$:

$$\|\mathcal{B}(\omega) - \mathcal{B}(\omega')\|_{\Phi+\Psi_b}^2 \leq \|\omega - \omega'\|_\Phi^2, \tag{24}$$

$$\|\mathcal{F}(\omega) - \mathcal{F}(\omega')\|_\Phi^2 \leq \|\omega - \omega'\|_{\Phi-\Psi_f}^2, \tag{25}$$

$$\Psi_b + \Psi_f \succeq \gamma(\Phi + \Psi_b). \tag{26}$$

Then the composite map $\mathcal{T} := \mathcal{B}\circ\mathcal{F}$ is ρ-contractive in $\|\cdot\|_{\Phi+\Psi_b}$ with $\rho = \sqrt{1-\gamma} \in [0, 1)$, namely

$$\|\mathcal{T}(\omega) - \mathcal{T}(\omega')\|_{\Phi+\Psi_b} \leq \rho\|\omega - \omega'\|_{\Phi+\Psi_b}. \tag{27}$$

Proof: This is [22, Prop. 2]; we include a proof for completeness. Let $\omega, \omega' \in \mathbb{R}^{n+m}$. Using (24) with $\omega \leftarrow \mathcal{F}(\omega)$ and $\omega' \leftarrow \mathcal{F}(\omega')$, and (25) and (26) gives

$$\begin{aligned}\|\mathcal{T}(\omega) - \mathcal{T}(\omega')\|_{\Phi+\Psi_b}^2 &= \|\mathcal{B}(\mathcal{F}(\omega)) - \mathcal{B}(\mathcal{F}(\omega'))\|_{\Phi+\Psi_b}^2 \\ &\leq \|\mathcal{F}(\omega) - \mathcal{F}(\omega')\|_\Phi^2 \leq \|\omega - \omega'\|_{\Phi-\Psi_f}^2 \\ &= \|\omega - \omega'\|_{\Phi+\Psi_b}^2 - \|\omega - \omega'\|_{\Psi_b+\Psi_f}^2 \\ &\leq (1-\gamma)\|\omega - \omega'\|_{\Phi+\Psi_b}^2,\end{aligned}$$

and taking square roots gives (27). ■

We can finally prove the main result of this letter.

Theorem 2 (Contractivity and linear convergence): Assume that $\alpha, \beta > 0$ and α as in (17). Let $\zeta_{\alpha,\beta}$ be as in (10), γ_λ as in (11), γ_x as in (18), Ψ_b as in (11), and $\Psi := \Phi + \Psi_b \succ 0$. Then the iteration (6) is contractive in $\|\cdot\|_\Psi$, with rate

$$\rho = \sqrt{1-\gamma}, \quad \gamma := \frac{\min\{\gamma_x, \gamma_\lambda\}}{\zeta_{\alpha,\beta} + \gamma_\lambda} \in (0, 1); \tag{28}$$

namely, (9) holds. Consequently, the iterates in (5) converge Q-linearly to the unique fixed point $\omega^\star$: for all $k \in \mathbb{N}$,

$$\|\omega^k - \omega^\star\|_\Psi \leq \rho^k \|\omega^0 - \omega^\star\|_\Psi.$$

Proof: The stepsize condition implies $\alpha\beta\|A\|^2 < 1$, hence $\Phi \succ 0$ by Lemma 2. Lemma 3 shows (24) with Ψ_b as in (11) while Lemma 4 shows (25) with $\Psi_f = \mathrm{diag}(\gamma_x I_n, 0)$. It remains to verify (26). Since $\Psi_b + \Psi_f \succeq \min\{\gamma_x, \gamma_\lambda\} I_{n+m}$ and $\lambda_{\max}(\Phi + \Psi_b) \leq \lambda_{\max}(\Phi) + \lambda_{\max}(\Psi_b) \leq \zeta_{\alpha,\beta} + \gamma_\lambda$, we have $I_{n+m} \succeq \frac{1}{\zeta_{\alpha,\beta}+\gamma_\lambda}(\Phi + \Psi_b)$, therefore $\Psi_b + \Psi_f \succeq \frac{\min\{\gamma_x,\gamma_\lambda\}}{\zeta_{\alpha,\beta}+\gamma_\lambda}(\Phi+\Psi_b) = \gamma(\Phi + \Psi_b)$, which is (26) with γ given in (28). Proposition 1 then yields contractivity of $\mathcal{T} = \mathcal{B}\circ\mathcal{F}$ in $\|\cdot\|_{\Phi+\Psi_b}$ with rate $\rho = \sqrt{1-\gamma}$. Finally, contractivity implies existence of a unique fixed point $\omega^\star$ and Q-linear convergence of ω^k to $\omega^\star$. ■

IV. Stochastic Case

In this section, we extend the result of Theorem 2 to a stochastic game, where the agents optimize an expected value function; namely, in the game (1),

$$J_i(x_i, x_{-i}) = \mathbb{E}_{\mathbb{P}}[f_i(x_i, x_{-i}, \xi_i(\varpi))]. \tag{29}$$

Here, $f_i : \mathbb{R}^n \times \mathbb{R}^{d_i} \to \mathbb{R}$ is a measurable function that depends also on some uncertainty ξ_i; this is a random variable $\xi_i : \Xi \to \mathbb{R}^{d_i}$ in the probability space $(\Xi, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}_{\mathbb{P}}$ is its expectation, which we assume to be well defined for all $x \in \mathbb{R}^n$. With some usual abuse of notation [14], we will write ξ_i instead of $\xi_i(\varpi)$ and we omit the subscript $\mathbb{P}$. We still postulate Standing Assumptions 1-3. Furthermore, we make the following assumption.

Assumption 1 (Regularity): The function $f_i(\cdot, x_-i, \xi_i)$ is differentiable, for any $i \in \mathcal{N}$, $x_{-i} \in \mathbb{R}^{n-n_i}$, and $\xi_i \in \mathbb{R}^{d_i}$.

Accordingly, the pseudogradient mapping appearing in the KKT conditions in (2) can be rewritten as $F(x) = \mathrm{col}((\mathbb{E}[\nabla_{x_i} f_i(x_i, x_{-i}, \xi_i)])_{i\in\mathcal{N}})$ and satisfies Standing Assumption 2; hence, a unique v-GNE $x^\star$ exists for the game (1). However, since $\nabla_{x_i} J_i$ cannot be computed due to the unknown uncertainty, in the algorithm a sample average stochastic approximation is used. In particular, the stochastic algorithm we consider reads as

$$\begin{aligned} x^{k+1} &= x^k - \alpha \left[\hat{F}(x^k, \xi^k) + A^\top \lambda^k\right] \\ \lambda^{k+1} &= \mathrm{proj}_{\mathbb{R}_{\geq 0}} \left(\lambda^k + \beta \left[2Ax^{k+1} - Ax^k - b\right]\right), \end{aligned} \tag{30}$$

where $\hat{F}(x^k, \xi) = \mathrm{col}(\hat{F}_1(x^k, \xi_1^k), \ldots, \hat{F}_N(x^k, \xi_N^k))$ and

$$\hat{F}_i(x, \xi_i^k) = \frac{1}{S_k} \sum_{s=1}^{S_k} \nabla_{x_i} f_i(x_i^k, x_{-i}^k, \xi_{i,s}^k), \tag{31}$$

where S_k is the number of samples used at step k, and the vector ξ_i^k collects the S_k samples of the independent and identically distributed (i.i.d) random realizations $\xi_{i,s}^k$. Let $\varepsilon^k := F(x^k) - \hat{F}(x^k, \xi^k)$ be the stochastic error, and $\mathcal{M} = \{\mathcal{M}_k\}_{k\in\mathbb{N}}$ be a filtration, i.e., a family of σ-algebras such that $\mathcal{M}_k = \sigma(x_0, x_1, \ldots, x_k)$ for all $k \geq 0$.

Assumption 2 (Variance reduction): There exist constants $\nu_1, \nu_2 > 0$ such that, for any $k \geq 0$, $\mathbb{E}[\|\varepsilon^k\|^2 \mid \mathcal{M}_k] \leq \frac{\nu_1\|x^k - x^\star\|^2 + \nu_2}{S_k}$ almost surely (a.s.).

Remark 4: Assumptions 1-2 are common [11], [18], and weaker than many conditions imposed in the related literature–e.g., Assumption 2 does not require unbiased estimates nor uniform bounds for the mean-squared error. □

Let us note that iteration (30) can be written as (see [25]) $\omega^{k+1} = \mathcal{B}(\hat{\mathcal{F}}(\omega^k, \xi^k))$, where

$$\hat{\mathcal{F}}(\omega^k, \xi^k) = \omega^k - \Phi^{-1} \begin{bmatrix} \hat{F}(x^k, \xi^k) \\ \mathbf{0}_m. \end{bmatrix}. \tag{32}$$

We can prove linear convergence of the stochastic step (30).

Theorem 3: Let Assumptions 1-2 hold. Consider the iteration in (30), and assume that the stepsizes α, β are chosen as in Theorem 2. Let Ψ and ρ be as in Theorem 2. Denote $v^k := \mathbb{E}[\|\omega^k - \omega^\star\|_\Psi^2]$. If $S_k \to \infty$ when $k \to \infty$, then v^k converges to zero. Furthermore, if $S_k \geq q^{-(k+1)}$ for some $q \in (0, 1)$ and all $k \geq 0$, then convergence is linear: for any $1 > \bar{\rho} > \rho$, there is $\bar{C} > 0$ such that

$$v^k \leq \bar{C} \ \max\{\bar{\rho}^2, q\}^k \ v^0. \tag{33}$$

Proof: We start by establishing the recursion (34) below, for the mean-squared error, which is the cornerstone of our derivation. First, we have, for any $\delta > 0$,,

$$\begin{aligned} \|\omega^{k+1} - \omega^\star\|_\Psi^2 &= \|\mathcal{B}(\hat{\mathcal{F}}(\omega^k, \xi^k)) - \omega^\star\|_\Psi^2 \\ &= \|\mathcal{B}(\hat{\mathcal{F}}(\omega^k, \xi^k)) - \omega^\star \pm \mathcal{B}(\mathcal{F}(\omega^k))\|_\Psi^2 \\ &\overset{(i)}{\leq} (1+\delta)\|\mathcal{B}(\mathcal{F}(\omega^k)) - \omega^\star\|_\Psi^2 \\ &\quad + \left(1 + \tfrac{1}{\delta}\right) \|\mathcal{B}(\mathcal{F}(\omega^k)) - \mathcal{B}(\hat{\mathcal{F}}(\omega^k, \xi^k))\|_\Psi^2 \\ &\overset{(ii)}{\leq} (1+\delta)\rho^2\|\omega^k - \omega^\star\|_\Psi^2 \\ &\quad + \left(1 + \tfrac{1}{\delta}\right) \|\mathcal{F}(\omega^k) - \hat{\mathcal{F}}(\omega^k, \xi^k)\|_\Phi^2 \\ &= (1+\delta)\rho^2\|\omega^k - \omega^\star\|_\Psi^2 \\ &\quad + \left(1 + \tfrac{1}{\delta}\right) \|\mathrm{col}(F(x^k) - \hat{F}(x^k, \xi^k), \mathbf{0}_m)\|_{\Phi^{-1}}^2 \\ &\overset{(iii)}{\leq} (1+\delta)\rho^2\|\omega^k - \omega^\star\|_\Psi^2 + \left(1 + \tfrac{1}{\delta}\right) \tau\|\varepsilon^k\|^2 \end{aligned}$$

where (i) follows by Cauchy–Schwarz and Young's inequality after expanding the square; (ii) follows by applying Theorem 2 to the first term and Lemma 3 to the second; (iii) is obtained as in Lemma 4, with $\tau := \frac{\alpha}{1-\alpha\beta\|A\|^2}$. By taking the conditional expectation on both sides,

$$\begin{aligned} &\mathbb{E}[\|\omega^{k+1} - \omega^\star\|_\Psi^2 \mid \mathcal{M}_k] \\ &\leq (1+\delta)\rho^2\mathbb{E}[\|\omega^k - \omega^\star\|_\Psi^2 \mid \mathcal{M}_k] + \left(1 + \tfrac{1}{\delta}\right) \mathbb{E}[\|\varepsilon^k\|^2 \mid \mathcal{M}_k] \\ &\leq (1+\delta)\rho^2\|\omega^k - \omega^\star\|_\Psi^2 + \left(1 + \tfrac{1}{\delta}\right) \tau \frac{\nu_1\|x^k - x^\star\|^2 + \nu_2}{S_k}. \end{aligned} \tag{34}$$

almost surely, where the last inequality follows because ω^k is $\mathcal{M}_k$-measurable and by Assumption 2.

Therefore, with $v^k := \mathbb{E}[\|\omega^k - \omega^\star\|_\Psi^2]$, by taking the (unconditional) expectation on both sides of (34), we have

$$v^{k+1} \leq \left[(1+\delta)\rho^2 + \left(1 + \tfrac{1}{\delta}\right) \tfrac{\nu_1\tau}{S_k}\|\Psi^{-1}\|\right] v^k + \left(1 + \tfrac{1}{\delta}\right) \frac{\tau\nu_2}{S_k},$$

where we used that $\|x^k - x^\star\|^2 \leq \|\omega^k - \omega^\star\|^2 \leq \|\Psi^{-1}\|\|\omega^k - \omega^\star\|_\Psi^2$, and hence the same holds for their expected values. Now, choose δ such that $(1+\delta)\rho^2 < \bar{\rho}^2 < 1$. If $S_k \to \infty$, the gain $(1+\delta)\rho^2 + \left(1 + \frac{1}{\delta}\right) \frac{\tau\nu_1}{S_k}\|\Psi^{-1}\| < \bar{\rho}^2$ for all k large enough, and $\left(1 + \frac{1}{\delta}\right) \frac{\tau\nu_2}{S_k} \to 0$. Then, convergence of v_k to zero follows immediately. To show linear convergence (and to compute an expression for $\bar{C}$) apply [18, Lemma 2.3]. ■

Remark 5: Theorem 3 is analogous to [17, Theorem 1], [18, Theorem 1], in that it establishes linear convergence in the mean-squared error sense, under analogous stochasticity assumptions. However, our contractivity analysis and proof strategy, derived from the previous section, are different, as we consider *generalized* games and have to account for the coupling constraints that affect the backward operator. □

V. Numerical Simulations

We consider an energy grid with $N = 20$ companies sharing $m = 7$ markets, as in [1, Fig. 1], see also [11]. Each company i chooses the quantities $x_i \in \mathbb{R}^{n_i}$ of energy to deliver to the n_i markets it is connected with. Each market j has a maximal capacity b_j, randomly drawn from $[0.5, 1]$, resulting in the coupling constraints $Ax \leq b$, where $A = \begin{bmatrix} A_1 & A_2 & \ldots & A_n \end{bmatrix}$, and $A_i \in \{0, 1\}^{m\times n_i}$ specifies the markets company i participates in. The generation cost incurred by company i is given by $c_i(x_i) = \pi_i \sum_{j=1}^{n_i}([x_i]_j)^2 + q_i^\top x_i$, where the constants π_i and q_i are randomly drawn from $[1, 8]$ and $[0.1, 0.6]$, respectively. The companies have also a revenue from selling in each market; the unitary price is $P(x) = \bar{P} - DA\boldsymbol{x}$, where $\bar{P} = \mathrm{col}(\bar{P}_1, \ldots, \bar{P}_7)$ is constant and randomly drawn from $[2, 4]$ while D indicates the supply and is randomly drawn from $[0.5, 1]$. Finally, the agents have local generation constraints of the form $0 \leq x_i \leq \theta_i$, θ_i with entries randomly drawn from $[1, 1.5]$. We model these local constraints by adding a (smooth, convex) barrier function r to the agents' cost [15]. Specifically, let

$$[r(y)]_j = \begin{cases} -\log y_j, & y_j \geq 1/\gamma, \\ -\gamma y_j + 1 + \log\gamma, & y_j < 1/\gamma \end{cases}$$

with y_j the j-th component of the variable y, and $\gamma \gg 0$ ($\gamma = 20$ in our simulations). Overall, the cost for agent i is $J_i(x) = c_i(x_i) + P(x)A_i x_i + r(x_i) + r(\theta_i - x_i)$. In the stochastic setting, the supply is the random quantity, i.e., $P(x, \xi) = \bar{P} - D(\xi)A\boldsymbol{x}$ and the entries of $D(\xi)$ are taken with a normal distribution,

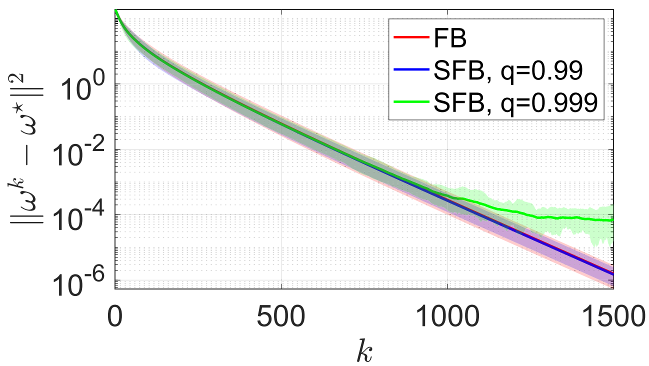

Fig. 1. Linear convergence of FB and SFB.

with mean D and finite variance. Standing Assumptions 1-3 and Assumption 1 are verified.

We simulate the FB algorithm in (5) and its stochastic counterpart in (30) with α, β and S_k taken according to Theorems 2 and 3. We report our results in Figure 1, where the thick lines indicate the average performance over 20 runs. In the stochastic case, a larger q requires less samples per iteration, but eventually limits the contractivity rate, as predicted by Theorem 3.

VI. Extensions and Outlook

We have established the contractivity of a forward-backward algorithm for equilibrium seeking in strongly monotone generalized games with full-rank inequality constraints. Contractivity not only implies Q-linear convergence, but also ensures important robustness and modularity properties (e.g., the composition of contractive maps is contractive). In this letter, we have exploited contractivity to provide novel linear convergence guarantees in stochastic generalized games. However, Theorem 2 paves the way for several other extensions. Fully-distributed algorithms and tracking guarantees for games with time-varying costs can be addressed as in [15]. Contractivity can also be exploited to study iterations with inexact updates or delays. As our analysis exploits contractivity in a (non-diagonally) weighted space, it is not clear how to account for projections for local constraints in the game; this is the main drawback of our analysis. Similarly, considering problems with time-varying coupling constraints is problematic, because the contractivity result is expressed in terms of a norm that depends on the constraint matrix A. These are important issues for future investigation.

References

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

[2] J. Lei and U. V. Shanbhag, "Stochastic Nash equilibrium problems: Models, analysis, and algorithms," *IEEE Control Syst. Mag.*, vol. 42, no. 4, pp. 103–124, Aug. 2022.

[3] O. Bilenne, B. Franci, P. Jacquot, N. Oudjane, M. Staudigl, and C. Wan, "A privacy-preserving decentralized algorithm for distribution locational marginal prices," in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, Dec. 2022, pp. 4143–4148.

[4] F. Fabiani, "Pursuing robust decisions in uncertain traffic equilibrium problems," in *Proc. 60th IEEE Conf. Decis. Control (CDC)*, Dec. 2021, pp. 5116–5121.

[5] F. Fabiani and B. Franci, "A stochastic generalized Nash equilibrium model for platforms competition in the ride-hail market," in *Proc. IEEE 61st Conf. Decis. Control (CDC)*, Dec. 2022, pp. 4455–4460.

[6] 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.

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

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

[9] M. Bianchi and S. Grammatico, "The END: Estimation network design for games under partial-decision information," *IEEE Trans. Control Netw. Syst.*, vol. 11, no. 4, pp. 2200–2212, Dec. 2024.

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

[11] B. Franci and S. Grammatico, "A distributed forward–backward algorithm for stochastic generalized Nash equilibrium seeking," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5467–5473, Nov. 2021.

[12] B. Franci and S. Grammatico, "Stochastic generalized Nash equilibrium-seeking in merely monotone games," *IEEE Trans. Autom. Control*, vol. 67, no. 8, pp. 3905–3919, Aug. 2022.

[13] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Ann. Oper. Res.*, vol. 175, pp. 177–211, Mar. 2010.

[14] U. Ravat and U. V. Shanbhag, "On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games," *SIAM J. Optim.*, vol. 21, no. 3, pp. 1168–1199, Jul. 2011.

[15] M. Bianchi, E. Benenati, and S. Grammatico, "Linear convergence in time-varying generalized Nash equilibrium problems," in *Proc. 62nd IEEE Conf. Decis. Control (CDC)*, Dec. 2023, pp. 7220–7226.

[16] T. Tatarenko, W. Shi, and A. Nedic, "Geometric convergence of gradient play algorithms for distributed Nash equilibrium seeking," *IEEE Trans. Autom. Control*, vol. 66, no. 11, pp. 5342–5353, Nov. 2021.

[17] J. Lei and U. V. Shanbhag, "Linearly convergent variable sample-size schemes for stochastic Nash games: Best-response schemes and distributed gradient-response schemes," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 3547–3552.

[18] J. Lei and U. V. Shanbhag, "Distributed variable sample-size gradient-response and best-response schemes for stochastic Nash equilibrium problems," *SIAM J. Optim.*, vol. 32, no. 2, pp. 573–603, Jun. 2022.

[19] F. Facchinei and J. Pang, "Nash equilibria: The variational approach," in *Convex Optimization in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2009, pp. 443–493.

[20] U. V. Shanbhag, J.-S. Pang, and S. Sen, "Inexact best-response schemes for stochastic Nash games: Linear convergence and iteration complexity analysis," in *Proc. IEEE 55th Conf. Decis. Control (CDC)*, Dec. 2016, pp. 3591–3596.

[21] H. Tao, A. Iannelli, M. Marschner, M. Staudigl, U. V. Shanbhag, and S. Cui, "Complexity guarantees for risk-neutral generalized Nash equilibrium problems," 2025, *arXiv:2506.11409*.

[22] C. Dirren, M. Bianchi, P. D. Grontas, J. Lygeros, and F. Dorfler, "Contractivity and linear convergence in bilinear saddle-point problems: An operator-theoretic approach," in *Proc. 28th Int. Conf. Artif. Intell. Statist.*, vol. 258, 2025, pp. 1954–1962.

[23] 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, Jan. 2012.

[24] F. Facchinei and J. Pang, *Finite-Dimensional Variational Inequalities Complementarity Problems*. Cham, Switzerland: Springer, 2007.

[25] G. Belgioioso and S. Grammatico, "Projected-gradient algorithms for generalized equilibrium seeking in aggregative games arepreconditioned forward-backward methods," in *Proc. Eur. Control Conf. (ECC)*, Jun. 2018, pp. 2188–2193.

[26] Y. Malitsky, "Golden ratio algorithms for variational inequalities," *Math. Program.*, vol. 184, nos. 1–2, pp. 383–410, Nov. 2020.

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement