Delft University of Technology

Nash equilibrium seeking under partial-decision information: monotonicity, smoothness and proximal-point algorithms

Bianchi, M.; Grammatico, S.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Nash equilibrium seeking under partial-decision information: Monotonicity, smoothness and proximal-point algorithms

Mattia Bianchi, Sergio Grammatico

Abstract— We consider Nash equilibrium problems in a partial-decision information scenario, where each agent can only exchange information with some neighbors, while its cost function possibly depends on the strategies of all agents. We characterize the relation between several monotonicity and smoothness assumptions postulated in the literature. Furthermore, we prove convergence of a preconditioned proximal-point algorithm, under a restricted monotonicity property that allows for a non-Lipschitz, non-continuous game mapping.

I. INTRODUCTION

Nash equilibrium (NE) seeking under partial-decision information has recently attracted considerable research interest, due to its prospect engineering applications as well as theoretical challenges. This scenario arises when, in the absence of a central coordinator, the agents in a network can only rely on the information received from some neighbors, for instance in ad-hoc-networks and sensor positioning problems [1], [2]. The technical goal is the distributed computation of an NE; the main complication is that the cost function of each agent may depend on the decision variables of other non-neighboring agents. To cope with the lack of knowledge, each agent estimates and tries to reconstruct the strategies of all the competitors [3], [4] (or an aggregation value [5], [6]) via peer-to-peer communication.

In fact, most existing methods resort to pseudogradient and consensus-type dynamics [7], [8]. Some works studied linearly convergent algorithms for games without coupling constraints [3], [9]. Other authors focused on generalized games, for example resorting to an operator-theoretic approach and forward-backward dual methods [6], [10]. All these schemes mainly suffer the following three drawbacks.

The first is that gradient-based methods typically require restrictive *monotonicity* assumptions for convergence. For instance, all the cited works postulate strong monotonicity of the game mapping. Weaker conditions are sometimes sufficient if allowing for vanishing stepsizes: strict monotonicity in the work [5], cocoercivity for generalized games in [2]. Remarkably, mere monotonicity was recently assumed in [11], via an additional diminishing Tikhonov regularization. Nonetheless, vanishing stepsizes are undesirable as they negatively affect the convergence speed. Most recently, the authors of [12] proposed a continuos-time gradient-based method for (hypo)-monotone games under a novel inverse Lipschitz assumption. The second drawback is that the

agents' cost functions must be *differentiable* with *Lipschitz* gradient [7], [10]; in turn this ensures that the pseudogradient mapping of the game is Lipschitz. As the game mapping is a global operator, implementing, in a distributed setup, the common alternatives employed in nonsmooth optimization (linesearch or adaptive steps) seems far from trivial. The third drawback is that, due to partial-decision information, the stepsizes must be chosen very small, in turn increasing the number of iterations for convergence. Importantly, this also translates in prohibitive communication cost, as the agents need to exchange data at each step.

A possible solution to remedy all three limitations is the proximal-point method [13, Th. 23.41]. Although a direct implementation in games results in double layer schemes (where the agents have to communicate virtually infinite time between iterations [14], [15]), in our recent work [16], [17] we have shown that an efficient method can be obtained via *preconditioning* – for the case of games with strongly monotone and Lipschitz mapping. The result is that, at the price of some additional *local* complexity, the number of iterations and communications for convergence to a NE can be substantially reduced.

In this paper we further leverage the properties of proximal-point algorithms (PPAs) to deal with the other two issues: monotonicity and smoothness. Our contributions are summarized as follows:

- We compare a significant group of monotonicity and smoothness assumptions employed in the partial-decision information literature; we characterize the relations between the conditions, and exemplify their restrictiveness (§IV);
- We prove convergence of our fully distributed NE seeking preconditioned proximal-point (PPP) algorithm, under the restricted monotonicity of an augmented operator. Our condition is remarkably weaker than that recently proposed in [18, Th. 2] (for a Douglas–Rachford algorithm). In particular, we do not assume strong monotonicity, nor continuity of the game mapping –which requires a different limiting argument compared to [16, Th. 2]. Interestingly, nonsmoothness only affects the *local* optimization problems of the agents (§V).

II. PRELIMINARIES

1) Notation: $[A]_{i,j}$ is the element on row i and column j of a matrix A. $\otimes$ denotes the Kronecker product. $\mathbf{0}_n \in \mathbb{R}^n$ ($\mathbf{1}_n \in \mathbb{R}^n$) is a vector with all elements equal to 0 (1); $I_n \in \mathbb{R}^{n \times n}$ is an identity matrix; we may omit the subscript if there is no ambiguity.

M. Bianchi and Sergio Grammatico are with the Delft Center for Systems and Control, TU Delft, The Netherlands. E-mail addresses: {m.bianchi, s.grammatico}@tudelft.nl.

2) Euclidean spaces: Given a positive definite matrix $\mathbb{R}^{q\times q} \ni P \succ 0$, $\mathcal{H}_P := (\mathbb{R}^q, \langle \cdot, \cdot \rangle_P)$ is the Euclidean space obtained by endowing $\mathbb{R}^q$ with the P-weighted inner product $\langle x, y \rangle_P = x^\top P y$, and $\|\cdot\|_P$ is the associated norm; we omit the subscripts if $P = I$. Unless otherwise stated, we always assume to work in $\mathcal{H} = \mathcal{H}_I$.

3) Operator-theoretic background [13]: A set-valued operator $\mathcal{F} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is characterized by its graph $\operatorname{gra}(\mathcal{F}) := \{(x, u) \mid u \in \mathcal{F}(x)\}$. $\operatorname{dom}(\mathcal{F}) := \{x \in \mathbb{R}^q | \mathcal{F}(x) \neq \varnothing\}$, $\operatorname{fix}(\mathcal{F}) := \{x \in \mathbb{R}^q \mid x \in \mathcal{F}(x)\}$ and $\operatorname{zer}(\mathcal{F}) := \{x \in \mathbb{R}^q \mid 0 \in \mathcal{F}(x)\}$ are the domain, set of fixed points and set of zeros, respectively. $\mathcal{F}^{-1}$ denotes the inverse operator of $\mathcal{F}$, defined as $\operatorname{gra}(\mathcal{F}^{-1}) = \{(u, x) \mid (x, u) \in \operatorname{gra}(\mathcal{F})\}$. $\mathcal{F}$ is (strictly, μ-strongly, ν-hypo-) monotone in $\mathcal{H}_P$ if $\langle u - v, x - y \rangle_P \geq 0$ (> 0, $\geq \mu \|x - y\|_P^2$, $\geq -\nu \|x - y\|_P^2$) for all $(x, u), (y, v) \in \operatorname{gra}(\mathcal{F})$; we omit the indication "in $\mathcal{H}_P$" whenever $P = I$. $\mathcal{F}$ is maximally monotone if it is monotone and there is no operator $\mathcal{A}$ such that $\operatorname{gra}(\mathcal{F}) \subset \operatorname{gra}(\mathcal{A})$. A single valued operator $\mathcal{F} : \mathbb{R}^q \to \mathbb{R}^q$ is β-cocoercive if $\langle x - y, \mathcal{F}(x) - \mathcal{F}(y) \geq \beta \|\mathcal{F}(x) - \mathcal{F}(y)\|^2$ for all $x, y \in \mathbb{R}^q$ (equivalently, $\mathcal{F}^{-1}$ is β-strongly monotone); is R-inverse Lipschitz if $R\|\mathcal{F}(x) - \mathcal{F}(y)\| \geq \|x - y\|$ (equivalently, $\mathcal{F}^{-1}$ is R-Lipschitz). Id is the identity operator. $\operatorname{J}_\mathcal{F} := (\operatorname{Id} + \mathcal{F})^{-1}$ denotes the resolvent operator of $\mathcal{F}$. For a function $\psi : \mathbb{R}^q \to \mathbb{R} \cup \{\infty\}$, $\operatorname{dom}(\psi) := \{x \in \mathbb{R}^q \mid \psi(x) < \infty\}$; its subdifferential operator is $\partial \psi : \operatorname{dom}(\psi) \rightrightarrows \mathbb{R}^q : x \mapsto \{v \in \mathbb{R}^q \mid \psi(z) \geq \psi(x) + \langle v \mid z - x \rangle, \forall z \in \operatorname{dom}(\psi)\}$; if ψ is differentiable and convex, $\partial \psi = \nabla \psi$. For a set $S \subseteq \mathbb{R}^q$, $\iota_S : \mathbb{R}^q \to \{0, \infty\}$ is the indicator function, i.e., $\iota_S(x) = 0$ if $x \in S$, ∞ otherwise; $\operatorname{N}_S : S \rightrightarrows \mathbb{R}^q : x \mapsto \{v \in \mathbb{R}^q \mid \sup_{z \in S} \langle v \mid z - x \rangle \leq 0\}$ is the normal cone of S. If S is closed and convex, then $\partial \iota_S = \operatorname{N}_S$ and $(\operatorname{Id} + \operatorname{N}_S)^{-1} = \operatorname{proj}_S$ is the Euclidean projection onto S.

Definition 1 (Restricted monotonicity): An operator $\mathcal{F} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is restricted (strictly, μ-strongly) monotone in $\mathcal{H}_P$ with respect to a set $\Sigma \neq \varnothing$ if $\langle x - x^\star, u - u^\star \rangle_P \geq 0$ (> 0, $\geq \mu \|x - x^\star\|_P^2$) for all $(x, u) \in \operatorname{gra}(\mathcal{F})$, $(x^\star, u^\star) \in \operatorname{gra}(\mathcal{F})$ with $x^\star \in \Sigma$. We omit the characterization in "$\mathcal{H}_P$" whenever $P = I$.

This definition slightly generalizes that in [16, Def. 1], which only considers the zero set; note that $\mathcal{F}$ is allowed to be set-valued on $x^\star \in \Sigma$.

Proximal-point algorithm: For an operator $\mathcal{F} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ with $\operatorname{zer}(\mathcal{F}) \neq \varnothing$, we consider the problem of finding a point $x^\star \in \operatorname{zer}(\mathcal{F})$. The following iteration is called PPA:

$$(\forall k \in \mathbb{N}) \quad x^{k+1} \in \operatorname{J}_\mathcal{F}(x^k) = (\operatorname{Id} + \mathcal{F})^{-1} x^k. \tag{1}$$

Note that at each iteration (1) involves solving for x^{k+1} the regularized inclusion $\mathbf{0} \in \mathcal{F}(x^{k+1}) + x^{k+1} - x^k$. By definition, $\operatorname{fix}(\operatorname{J}_\mathcal{F}) = \operatorname{zer}(\mathcal{F})$. If $\mathcal{F}$ is maximally monotone, then $\operatorname{J}_\mathcal{F}$ is single valued and $\operatorname{dom}(\operatorname{J}_\mathcal{F}) = \mathbb{R}^q$, so (1) is uniquely defined; also, x^k converges to a point in $\operatorname{zer}(\mathcal{F})$.

III. MATHEMATICAL SETUP

A. The game

Let $\mathcal{I} := \{1, \dots, N\}$ be a set of agents, where each agent $i \in \mathcal{I}$ chooses its strategy (i.e., decision variable) x_i from its local decision set $\Omega_i \subseteq \mathbb{R}^{n_i}$. We denote by $x := \operatorname{col}((x_i)_{i \in \mathcal{I}}) \in \Omega$ the stacked vector of all the agents' strategies, with $\Omega := \Omega_1 \times \dots \times \Omega \subseteq \mathbb{R}^n$ the overall decision space and $n := \sum_{i \in \mathcal{I}} n_i$. Agent $i \in \mathcal{I}$ aims to minimize an objective function $f_i(x_i, x_{-i})$, depending both on the local variable x_i and on the strategies of the other agents $x_{-i} := \operatorname{col}((x_j)_{j \in \mathcal{I} \setminus \{i\}})$. The game consists of N inter-dependent optimization problems

$$\forall i \in \mathcal{I} : \quad \underset{y_i \in \Omega_i}{\operatorname{argmin}} \; f_i(y_i, x_{-i}). \tag{2}$$

The mathematical problem we consider is the distributed computation of a NE, a set of strategies simultaneously solving all the problems in (2).

Definition 2: A Nash equilibrium is a set of strategies $x^\star = \operatorname{col}((x_i^\star)_{i \in \mathcal{I}})$ such that, for all $i \in \mathcal{I}$, $x_i^\star \in \operatorname{argmin}_{y_i \in \Omega_i} f_i(y_i, x_{-i}^\star)$

We restrict our attention to convex games. The following are standard regularity conditions [5, Asm. 1], [10, Asm. 1].

Assumption 1 (Convexity): For each $i \in \mathcal{I}$, the set Ω_i is nonempty, closed and convex; the fuction f_i is continuous and the function $f_i(\cdot, x_{-i})$ is convex for any x_{-i}.

Furthermore, we assume existence of a solution.

Assumption 2 (Existence): The game in (2) admits at least one Nash equilibrium.

Sufficient conditions for existence of a NE (e.g., compactness of Ω) can be found for instance in [19].

B. The communication network

The agents can exchange information with some neighbors over an undirected communication network $\mathcal{G}(\mathcal{I}, \mathcal{E})$. The unordered pair (i, j) belongs to the set of edges $\mathcal{E}$ if and only if agent i and j can mutually exchange information. We denote: $W \in \mathbb{R}^{N \times N}$ the weight matrix of $\mathcal{G}$, with $w_{i,j} := [W]_{i,j}$ and $w_{i,j} > 0$ if $(i, j) \in \mathcal{E}$, $w_{i,j} = 0$ otherwise; $\mathcal{N}_i = \{j \mid (i, j) \in \mathcal{E}\}$ the set of neighbors of agent i.

Assumption 3 (Connectivity): The communication graph $\mathcal{G}(\mathcal{I}, \mathcal{E})$ is undirected and connected. The weight matrix W satisfies the following conditions:
 (i) *Symmetry:* $W = W^\top$;
 (ii) *Self loops:* $w_{i,i} > 0$ for all $i \in \mathcal{I}$;
(iii) *Double stochasticity:* $W \mathbf{1}_N = \mathbf{1}_N, \mathbf{1}^\top W = \mathbf{1}^\top$.

The requirements (ii)-(iii) in Assumption 3 are intended to ease the notation; for instance, they can be satisfied by assigning Metropolis weights [9, §2].

C. The partial-decision information scenario

We consider the so-called partial-decision information model, where agent $i \in \mathcal{I}$ can only access its own feasible set Ω_i and an analytic expression of its private cost f_i, but cannot access the strategies of all the competitors x_{-i}. Therefore, each agent i is unable to evaluate the actual value of $f_i(x_i, x_{-i})$. Instead, each agent keeps an estimate of all other agents' actions [4], [5], [8], and aims at reconstructing the actual values, only based information exchanged locally with neighbors over the communication graph $\mathcal{G}$. We denote $\boldsymbol{x}_i = \operatorname{col}((\boldsymbol{x}_{i,j})_{j \in \mathcal{I}}) \in \mathbb{R}^n$, where $\boldsymbol{x}_{i,i} := x_i$ and $\boldsymbol{x}_{i,j}$

is agent i's estimate of agent j's strategy, for all $j \neq i$; $\boldsymbol{x}_{j,-i} = \mathrm{col}((\boldsymbol{x}_{j,l})_{l \in \mathcal{I} \setminus \{i\}})$; $\boldsymbol{x} = \mathrm{col}((\boldsymbol{x}_i)_{i \in \mathcal{I}}) \in \mathbb{R}^{Nn}$ the overall estimate vector; $\boldsymbol{x}_{-i} = \mathrm{col}((x_j)_{j \in \mathcal{I} \setminus \{i\}})$. Let

$$\mathcal{R}_i := \begin{bmatrix} \mathbf{0}_{n_i \times n_{<i}} & I_{n_i} & \mathbf{0}_{n_i \times n_{>i}} \end{bmatrix}, \tag{3}$$

where $n_{<i} := \sum_{j<i, j \in \mathcal{I}} n_j$, $n_{>i} := \sum_{j>i, j \in \mathcal{I}} n_j$. In simple terms, $\mathcal{R}_i$ selects the i-th n_i-dimensional component from an n-dimensional vector, i.e., $\mathcal{R}_i \boldsymbol{x}_i = \boldsymbol{x}_{i,i} = x_i$. Let also $\mathcal{R} := \mathrm{diag}((\mathcal{R}_i)_{i \in \mathcal{I}})$, so that $x = \mathcal{R}\boldsymbol{x}$.

D. Game mapping, extended mapping, augmented operators

Under Assumption 1, a strategy $x^\star$ is a NE of the game in (2) if and only if

$$\mathbf{0}_n \in F(x^\star) + \mathrm{N}_\Omega(x^\star), \tag{4}$$

where $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the *game mapping*

$$F(x) := \mathrm{col}((\partial_{x_i} f_i(x_i, x_{-i}))_{i \in \mathcal{I}}) \tag{5}$$

(in fact, (4) are the first order optimality conditions of each convex problems in (2)). Typically, distributed NE seeking methods require some monotonicity assumption on F. Since we deal with the partial-decision information scenario, it is also useful to introduce the *extended game mapping*

$$\boldsymbol{F}(\boldsymbol{x}) := \mathrm{col}((\partial_{x_i} f_i(x_i, \boldsymbol{x}_{i,-i}))_{i \in \mathcal{I}}) \tag{6}$$

where the subdifferentials are computed on the *estimates*, and the *augmented* operators

$$\mathcal{F}_\alpha(\boldsymbol{x}) := \alpha \mathcal{R}^\top \boldsymbol{F}(\boldsymbol{x}) + (I_{Nn} - \boldsymbol{W})\boldsymbol{x} \tag{7}$$

$$\mathcal{A}_\alpha(\boldsymbol{x}) := \mathcal{F}_\alpha(\boldsymbol{x}) + \mathrm{N}_{\boldsymbol{\Omega}}(\boldsymbol{x}), \tag{8}$$

where $\alpha > 0$ is a design parameter, $\boldsymbol{W} := W \otimes I_n$, $\boldsymbol{\Omega} := \{\boldsymbol{x} \in \mathbb{R}^{Nn} \mid \mathcal{R}\boldsymbol{x} \in \Omega\}$. The following well-known result (e.g., [3, Prop. 1]) provides an extension of the inclusion (4) to the estimate space.

Lemma 1: The following statements are equivalent:
i) $\boldsymbol{x}^\star = \mathbf{1}_N \otimes x^\star$, with $x^\star \in \Omega$ a NE of the game (2);
ii) $\mathbf{0}_{Nn} \in \mathcal{A}_\alpha(\boldsymbol{x}^\star)$.
In particular, Assumption 2 implies that $\mathrm{zer}(\mathcal{A}_\alpha) \neq \varnothing$.

IV. TOWARDS A TAXONOMY OF ASSUMPTIONS

In recent years, distributed NE seeking under partial-decision information has been studied under a variety of conditions on the operators $F, \mathcal{R}^\top \boldsymbol{F}, \mathcal{F}_\alpha, \mathcal{A}_\alpha$. Some of the assumptions postulated have not been exemplified, nor it is evident how restrictive they are –in theory and in practice. Towards a solution of this issue, we start by considering the following, representative, conditions.

C1: The operator $\mathcal{R}^\top \boldsymbol{F}$ is maximally monotone.

C2: The operator $\mathcal{R}^\top \boldsymbol{F}$ is restricted monotone with respect to $\mathrm{zer}(\mathcal{A}_\alpha)$.

C3: There exists $\alpha \geq 0$ such that the operator $\mathcal{F}_\alpha$ is maximally monotone.

C4: There exists $\alpha \geq 0$ such that the operator $\mathcal{F}_\alpha$ is restricted monotone with respect to $\mathrm{zer}(\mathcal{A}_\alpha)$.

C5: The operator F is μ-restricted strongly monotone with respect to the set of NEs and ℓ-Lipschitz, for some $\mu > 0$, $\ell > 0$.

	ref	extra asm.	step sizes
C1	[4], [12]		Continuous time
C3	[18], [20]		Fixed
C5	[3]		Fixed
C6	[16], [10], [8]		Fixed
C7	[12]	$\Omega = \mathbb{R}^n$	Continuous time
C8	[5]	Ω compact	Vanishing
C9	[2]	Ω compact	Vanishing
C10	[11]	Ω compact	Vanishing

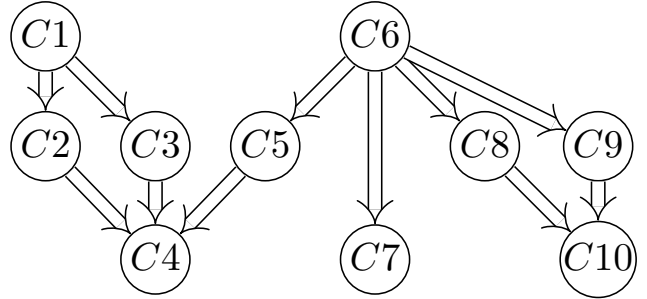Table I. Technical assumptions in the literature.

Fig. 1: Relations between technical assumptions in monotone games under partial-decision information.

C6: The operator F is μ-strongly monotone and ℓ-Lipschitz, for some $\mu > 0$, $\ell > 0$.

C7: The operator F is ν-hypomonotone, ℓ-Lipschitz, and R-inverse Lipschitz, for some $\nu \geq 0$, $\ell > 0$, $R > 0$, $R\nu < 1$.

C8: The operator F is strictly monotone and ℓ-Lipschitz, for some $\ell > 0$.

C9: The operator F is $\frac{1}{\ell}$ cocoercive for some $\ell > 0$.

C10: The operator F is monotone and ℓ-Lipschitz, for some $\ell > 0$.

Although C6 is the most common technical assumption, all these conditions have been formulated in the literature (see Table I), except for C2 (which is a natural relaxations of C1) and C4 (which we will use to show convergence of our algorithm). The following result characterizes the relation between them.

Proposition 1: The implications in Figure 1 hold true.

It can be also shown by counter examples that no other implication exists between the conditions in C1-C10.

A. Conditions on the extended pseudogradient

We next prove, under the commonly used assumption that F is single valued, that C1 is very restrictive.

Proposition 2 (C1 is trivial): Assume that F is single valued and continuous. Then, condition C1 holds if and only if $\nabla f_i(\cdot, x_{-i})$ is independent of x_{-i}, for all $i \in \mathcal{I}$.

As the actions x_{-i} are not affecting the optimization problem of agent i (beside possibly for a separable component), there appears to be no reason for agent i to keep estimates (hence, for a partial-decision information setup).

Example 1: The game defined by $N = 2$, $n = 2$, $\Omega = \mathbb{R}^n$, $f_1(x) = (x_1 - 1)^2(x_2{}^2 + 1)$, $f_2(x) = x_2{}^2(x_1{}^2 + 1)$ has a unique NE in $\begin{bmatrix} 1 & 0 \end{bmatrix}^\top$ and satisfies C2, but not C1.

5082

Although $\nabla_{x_i} f_i$ depends on x_{-i} in Example 1, the next lemma shows that C2 is also not of particular interest.

Proposition 3 (C2 is trivial): Assume that F is single valued and continuous. Then, condition C2 holds if and only if $\nabla_{x_i} f_i(x_i^\star, x_{-i})$ is independent of x_{-i}, for all $i \in \mathcal{I}$, for any $x^\star = (x_i^\star, x_{-i}^\star)$ NE of the game (2).

In particular, Proposition 3 implies that $0 \leq \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star), x_i - x_i^\star\rangle = \langle \nabla_{x_i} f_i(x_i^\star, x_{-i}), x_i - x_i^\star\rangle$ where the inequality is the first order optimality condition (as $x_i^\star$ solves (2)). This means that $x_i^\star$ is optimal for agent i regardless of x_{-i}; in other terms, C2 implies that the Nash equilibria are uniquely composed by dominant strategies (as in Example 1). This is also a trivial case, as the agents do not need to communicate to compute a NE. Although the condition in Proposition 3 might be violated if F is not continuous, this can only happen at discontinuity points, which is quite a pathological case.

B. Conditions on the game primitives

Conditions C5 to C10 are directly postulated on the game mapping F and are the most well-investigated (e.g., they are easy to check if F is a linear operator [13], [21], [12]). Conditions C5 to C8 imply uniqueness of the equilibrium; methods with linear convergence were proposed under C5, C6 [3], [9], but not C7, C8. Although C5 is weaker than C6 in theory, it is difficult to check without knowledge of the solutions; we have included it because it causes very limited complications in the convergence analysis with respect to C6. C6 actually implies that there is $\alpha > 0$ such that $\mathcal{F}_\alpha$ is Lipschitz and restricted strongly-monotone with respect to the whole consensus subspace $\boldsymbol{E} := \{\boldsymbol{y} \in \mathbb{R}^{Nn} \mid \boldsymbol{y} = \mathbf{1}_N \otimes y, \ y \in \mathbb{R}^n\} \supset \text{zer}(\mathcal{A}_\alpha)$ [10, Lem. 3], a much more restrictive condition that C4. C10 and C9 allow for multiple NEs; yet – as for C8 – the related methods require not only compact feasible sets (possibly reasonable in practice) but also vanishing steps, which affect the convergence speed.

C. Conditions on the augmented operator

C3 and C4 are more abstract and often replaced by more easily checked sufficient conditions. For example, restricted monotonicity of $\mathcal{F}_\alpha$ with respect to the consensus space $\boldsymbol{E}$ can be always checked without knowledge of the solutions, and implies C4.

Despite this complication, C3 and C4 are of great interest, especially for nonsmooth games, as exemplified next. The following examples also show that C3 is significantly more restrictive than C4.

Example 2: Consider the game defined by $N = 2$, $n = 2$, $\Omega = \mathbb{R}^n$, $F(x) = \bar{F}(x) + \hat{F}(x)$, with $\bar{F}(x) = \text{col}(x_1^3, 0)$ and $\hat{F}(x) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x + \begin{bmatrix} 5 \\ 4 \end{bmatrix}$. As $\bar{F}$ is monotone and $\hat{F}$ is strongly monotone, the game admits a unique NE. Conditions C5–C10 are violated, as they require Lipschitz continuity of F; C2 also fails (as the best response of agent 2 is $-0.5x_1 - 2$ and by Proposition 3). However, C4 holds: to show this, consider the components of the extended game mapping $\bar{\boldsymbol{F}}$ and $\hat{\boldsymbol{F}}$ corresponding to $\bar{F}$ and $\hat{F}$; $\mathcal{R}^\top \bar{\boldsymbol{F}}$ is monotone, while $\alpha \mathcal{R}^\top \hat{\boldsymbol{F}} + (I - \boldsymbol{W})$ can be made restricted monotone with

---

Algorithm 1 Fully-distributed PPP algorithm

$$\tilde{\boldsymbol{x}}_i^k = \tfrac{1}{2}\left(\boldsymbol{x}_i^k + \sum_{j=1}^N w_{i,j} \boldsymbol{x}_j^k\right)$$
$$\boldsymbol{x}_{i,-i}^{k+1} = \tilde{\boldsymbol{x}}_{i,-i}^k$$
$$x_i^{k+1} = \underset{y \in \Omega_i}{\arg\min} \ \left(f_i(y, \tilde{\boldsymbol{x}}_{i,-i}^k) + \tfrac{1}{\alpha}\|y - \tilde{\boldsymbol{x}}_{i,i}^k\|^2\right)$$

---

respect to the consensus subspace by choosing $\alpha > 0$ small enough [10, Lem. 3]. We can check numerically that C3 also holds for some W (in particular, because $\alpha \mathcal{R}^\top \hat{\boldsymbol{F}} + (I - \boldsymbol{W})$ can be made monotone, although there is no analytical result available to check this a priori).

Example 3 (Non-monotone game): Consider Example 2 but with $\bar{F}(x) = \text{col}(x_1^3(x_2^4 + 1), 0)$ and $\hat{F}(x) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} x$. The game admits a NE $x^\star = \boldsymbol{0}$. As F is restricted strongly monotone with respect to $x^\star$, the equilibrium must be unique. As for Example 2, it is easy to prove that C4 holds, because $\mathcal{R}^\top \bar{\boldsymbol{F}}$ is restricted monotone with respect to $\boldsymbol{0}$ (by Proposition 3). Yet, F is not monotone: therefore C3 cannot hold (nor can C1–C2, C5-C10).

Example 4 (Set-valued F): Consider the game defined by $N = 2$, $n = 2$, $\Omega = \mathbb{R}^n$, $f_1(x) = x_1^2 - |x_1|x_2$, $f_2(x) = x_2^2 + x_2 x_1$, where $|\cdot|$ is the absolute value. The game admits a unique NE in $\boldsymbol{0}$; moreover, F is set valued, as f_1 is not differentiable in the local variable. It can be checked that C4 holds. Yet, F is not monotone, thus C3 fails.

V. THE PPP ALGORITHM

In this section we consider the fully-distributed proximal-point NE seeking method shown in Algorithm 1. The iteration coincides with that studied in [17], although the terms have been rearranged. The algorithm includes a consensus phase, where the agents exchange and mix their variable vectors. The local actions are then updated according to a proximal-best response with stepsize $\alpha > 0$ – importantly, the cost function of each agent i evaluated in the estimates $\boldsymbol{x}_{i,-i}$, and not on the real competitor's actions x_{-i}. Note that the algorithm is always well (uniquely) defined, as the update of x_i is the $\arg\min$ of a strongly convex function (by convexity of $f_i(\cdot, x_{-i})$ in Assumption 1).

Algorithm 1 can be formulated as a proximal-point method applied to the operator $\mathcal{A}_\alpha$. However, the computation of $(\text{Id} + \mathcal{A}_\alpha)^{-1}$ cannot be performed in a distributed way (more precisely, it would require the collaborative solution of a regularized game at each iteration, resulting in a scheme with nested layers of communication, see [14]). We have shown in [17], [16] that this complication can be tackled by preconditioning the operator $\mathcal{A}_\alpha$ with a positive definite matrix

$$\Phi := I_{Nn} + \boldsymbol{W}. \tag{9}$$

Lemma 2 ([17, Lem. 2]): Algorithm 1 can be written as

$$\boldsymbol{x}^{k+1} = (\text{Id} + \Phi^{-1}\mathcal{A}_\alpha)^{-1}(\boldsymbol{x}^k). \tag{10}$$

This operator-theoretic interpretation is very powerful, as it seamlessly allows to study convergence of analogous proximal-best response schemes even in the presence of inexact updates (i.e., the $\arg\min$ is only approximated at each iteration), coupling constraint, acceleration terms [16]. It also immediately shows that the fixed points of Algorithm 1 coincide with $\operatorname{zer}(\mathcal{A}_\alpha) = \operatorname{zer}(\Phi^{-1}\mathcal{A}_\alpha)$ (i.e., they are estimates at consensus at a Nash equilibrium).

The following theorem is the main result of the paper. It extends the convergence results in [16, Th. 3], formulated under C6, to the case of restricted monotone – possibly nonsmooth – games (i.e., C4).

Theorem 1: Let Assumption 1–3 hold, and assume that C4 holds for some $\alpha > 0$. Then, the sequence $(\boldsymbol{x}^k)$ generated by Algorithm 1 converges to a point $\boldsymbol{x}^\star = \mathbf{1}_N \otimes x^\star$, where $x^\star$ is a Nash equilibrium of the game in (2).

Remark 1: In [17] we have proven (linear) convergence of Algorithm 1 assuming C6. Under the weaker C4, Theorem 1 leverages the general results for the proximal-point algorithm of restricted (merely) monotone games [16]. With respect to [16] and to the Douglas-Rachford algorithm in [18], we use a different limiting argument in our proof, which does not require F to be Lipschitz continuous (or even continuous). The core idea is to show that the operator $\mathrm{J}_{\Phi^{-1}\mathcal{A}_\alpha}$ is continuous, even if $\mathcal{A}_\alpha$ might not (nor is maximally monotone). For instance, Theorem 1 can be applied to the games in Examples 2 to 4, while [16, Th. 2], [12, Th. 2] cannot. Examples 3 and 4 also show a significant gap between C4 and the stronger condition C3, employed in [18, Th. 3].

We conclude this section by sketching some technical extensions of our results. To start, our arguments in Theorem 1 can be readily adapted to the algorithms – for generalized games – studied in [16], to show convergence under C4. Moreover, our convergence result would hold assuming the definition of restricted monotonicity proposed in [16, Def. 1], slightly less restrictive than our Definition 1. We also note that we assumed monotonicity properties of F (and similarly for the other game operators) to hold over all $\mathbb{R}^n$; however, the conditions can be relaxed to hold only over the feasible set, if the estimates $\boldsymbol{x}$'s are initialized in Ω^N (since the update in Algorithm 1 guarantees invariance for this set). The costs in (2) can be modified to include a more general (discontinuous) proper, convex, closed function $g_i(x_i)$ (besides the indicator function ι_{Ω_i}), without technical complications. Much more intriguing is the case of discontinuity in the part of the cost coupled with the other agents (i.e., violating Assumption 1): although our convergence arguments do not hold in this case, it would be interesting to verify whether C3 could be satisfied to apply standard PPA results.

VI. CONCLUSION AND OUTLINE

Besides their efficiency, proximal-point algorithms have the advantage of only requiring mild monotonicity and smoothness conditions. We have compared and analyzed several assumptions in NE seeking under partial-decision information, and proved the convergence of a fully distributed PPP method under one of the weakest.

Future work should investigate linear rates in absence of (restricted) strong monotonicity. One promising option is to leverage inverse Lipschitz properties, which can ensure contractivity of certain resolvents. Proving convergence in merely monotone regime, under fixed step sizes, is also a challenging open problem.

APPENDIX

1) Proof of Proposition 1: $C1 \Rightarrow C2$, $C3 \Rightarrow C4$, $C6 \Rightarrow C5$, $C6 \Rightarrow C8$, $C8 \Rightarrow C10$: By definition.

$C1 \Rightarrow C3$: As $(I - W)$ is a positive semidefinite matrix, the operator $I - \boldsymbol{W}$ is maximally monotone. Hence, for any $\alpha \geq 0$, $\mathcal{F}_\alpha = \alpha \mathcal{R}^\top \boldsymbol{F} + (I - \boldsymbol{W})$ is the sum of two maximally monotone operators; moreover, $\operatorname{dom}(I - \boldsymbol{W}) = \mathbb{R}^{Nn}$, so the conclusion follows by [13, Cor. 25.5].

$C2 \Rightarrow C4$: $\mathcal{F}_\alpha$ is the sum of a restricted monotone operator and a monotone operator, hence restricted monotone.

$C5 \Rightarrow C6$: See, for instance, [16, Lem. 3].

$C6 \Rightarrow C7$: It follows by definition and [12, Prop. 3].

$C6 \Rightarrow C9$: See, e.g., [12, Prop. 5].

$C9 \Rightarrow C10$: It follows by definition of cocoercivity and the Cauchy–Schwartz inequality. ∎

2) Proof of Proposition 2: "$\Leftarrow$": For the sake of contradiction, assume that, for some $i \in \mathcal{I}$, there exist $l \in \{1, 2, \ldots, n_i\}$, $x_i \in \mathbb{R}^{n_i}$ and a pair of vectors x_{-i} and x'_{-i} such that $[\nabla_{x_i} f_i(x_i, x_{-i})]_l < [\nabla_{x_i} f_i(x_i, x'_{-i})]_l$. By continuity, there exists $\epsilon > 0$ such that $[\nabla_{x_i} f_i(x_i + \epsilon e_l, x_{-i})]_l < [\nabla_{x_i} f_i(x_i, x'_{-i})]_l$, where $e_l \in \mathbb{R}^n_i$ is the l-th vector of the canonical basis. The monotonicity in C1, applied to a pair of estimate vectors $(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$, $(\boldsymbol{x}'_i, \boldsymbol{x}_{-i})$, for any $\boldsymbol{x}_{-i}$ and $\boldsymbol{x}_i = (x_i + \epsilon e_l, x_{-i})$, $\boldsymbol{x}'_i = (x_i, x'_{-i})$, gives

$$0 \leq \langle \nabla_{x_i} f_i(x_i + \epsilon e_l, x_{-i}) - \nabla_{x_i} f_i(x_i, x'_{-i}), \epsilon e_l \rangle$$
$$= \epsilon[\nabla_{x_i} f_i(x_i + \epsilon e_l, x_{-i}) - \nabla_{x_i} f_i(x_i, x'_{-i})]_l \qquad < 0$$

which is a contradiction. Because x_{-i}, x'_{-i} are arbitrary, we conclude that, for all $i \in \mathcal{I}$, for all x_i, and for all x_{-i}, x'_{-i}, $\nabla_{x_i} f_i(x_i, x_{-i}) = \nabla_{x_i} f_i(x_i, x'_{-i})$.

"$\Rightarrow$": By assumption, for any $i \in \mathcal{I}$, x_i, x'_i, x_{-i}, x'_{-i},

$$\langle \nabla_{x_i} f_i(x_i, x_{-i}) - \nabla_{x_i} f_i(x'_i, x'_{-i}), x_i - x'_i \rangle$$
$$= \langle \nabla_{x_i} f_i(x_i, x'_{-i}) - \nabla_{x_i} f_i(x'_i, x'_{-i}), x_i - x_i \rangle \quad \geq 0,$$

where the inequality is convexity of f_i in the first argument (Assumption 1). Stacking the inequalities for $i \in \mathcal{I}$ retrieves the monotonicity of $\mathcal{R}^\top \boldsymbol{F}$. ∎

3) Proof of Proposition 3: "$\Leftarrow$": For contradiction, assume that there exist $i \in \mathcal{I}$, $l \in \{1, 2, \ldots, n_i\}$, an NE $x^\star$ and x_{-i} such that $[\nabla_{x_i} f_i(x_i^\star, x_{-i})]_l < [\nabla_{x_i} f_i(x_i^\star, x_{-i}^\star)]_l$. By continuity, there exists $\epsilon > 0$ such that $[\nabla_{x_i} f_i(x_i^\star + \epsilon e_l, x_{-i})]_l < [\nabla_{x_i} f_i(x_i^\star, x_{-i}^\star)]_l$. Restricted monotonicity in C2, applied to a pair of estimate vectors $(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$, $(\boldsymbol{x}^\star, \boldsymbol{x}_{-i})$, for any $\boldsymbol{x}_{-i}$ and $\boldsymbol{x}_i = (x_i^\star + \epsilon e_l, x_{-i})$, gives

$$0 \leq \langle \nabla_{x_i} f_i(x_i^\star + \epsilon e_l, x_{-i}) - \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star), \epsilon e_l \rangle$$
$$= \epsilon[\nabla_{x_i} f_i(x_i^\star + \epsilon e_l, x_{-i}) - \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star)]_l \qquad < 0$$

which is a contradiction. Analogously it can be shown that $[\nabla_{x_i} f_i(x_i^\star, x_{-i})]_l > [\nabla_{x_i} f_i(x_i^\star, x_{-i}^\star)]_l$ leads to a contradiction. Hence $\nabla_{x_i} f_i(x_i^\star, x_{-i}) = \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star)$.

"⇒": For any $i \in \mathcal{I}$, x_i, x_{-i}, NE $x^\star$, by assumption and convexity, $\langle f_i(x_i, x_{-i}) - \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star), x_i - x_i^\star \rangle = \langle \nabla_{x_i} f_i(x_i, x_{-i}) - \nabla_{x_i} f_i(x_i^\star, x_{-i}^\star), x_i - x_i^\star \rangle \geq 0$. ∎

4) Proof of Theorem 1: We start by an auxiliary result.

Lemma 3: Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} : (x, y) \mapsto f(x, y)$ be a continuous function, and assume that $f(\cdot, y)$ is μ-strongly convex for any $y \in \mathbb{R}^m$, $\mu > 0$. Let $X \subseteq \mathbb{R}^n$ be a convex closed set. Then the (single valued, full domain) mapping $y \mapsto g(y) = \operatorname{argmin}_{x \in X} f(x, y)$ is continuous.

Proof: We show that, for any given sequence $(y_k)_{k \in \mathbb{N}}$ with $y^k \to y^\star$ (converging, hence bounded), $x^k := g(y^k) \to g(y^\star) =: x^\star$, which is the definition of continuity.

First, we show that $(x^k)_{k \in \mathbb{N}}$ is bounded. Let Y be a compact set containing $(y^k)_{k \in \mathbb{N}}$. Let $x_0 \in X$ and

$$l_0 := \max_{y \in Y} f(x_0, y), \qquad l_1 := \min_{x \in \partial B(x_0, 1), y \in Y} f(x, y)$$

where $\partial B(x_0, 1) = \{x \in \mathbb{R}^n \mid \|x - x_0\| = 1\}$ is the boundary of the unit ball centered at x_0; the min and max are achieved because the domains are compact. Let $d \in \mathbb{R}^n$ be any unitary vector, i.e., $\|d\| = 1$; $x_1 := x_0 + d \in \partial B(x_0, 1)$; $x_2 = x_0 + Md$, for some scalar such that

$$M > 1, \qquad M > 2\frac{l_0 - l_1}{\mu} + 1. \tag{11}$$

Then, $x_1 = \frac{M-1}{M} x_0 + \frac{1}{M} x_2$. By definition of strong convexity, we have, for all $y \in Y$

$$\begin{aligned}
l_1 &\leq f(x_1, y) \\
&\leq \frac{M-1}{M} f(x_0, y) + \frac{1}{M} f(x_2, y) - \frac{1}{2}\mu \frac{M-1}{M} \frac{1}{M} \|x_0 - x_2\|^2 \\
&= \frac{M-1}{M} f(x_0, y) + \frac{1}{M} f(x_2, y) - \frac{1}{2}\mu(M - 1).
\end{aligned}$$

Assume for contradiction that there exists $y \in Y$ such that $f(x_2, y) \leq f(x_0, y)$. Then, since $f(x_0, y) \leq l_0$, the previous inequality implies $l_1 - l_0 \leq -\frac{1}{2}\mu(M - 1)$, which contradicts (11). Since d is arbitrary, we conclude that, for any $y \in Y$, for all x such that $\|x_0 - x\| > M$, $f(x_0, y) < f(x, y)$. In turn, for all $y \in Y$, $\|g(y)\| < \|x_0\| + M$, i.e., g is uniformly bounded over Y; thus $(x_k)_{k \in \mathbb{N}}$ is bounded.

Hence $(x_k)_{k \in \mathbb{N}}$ admits an accumulation point, say x'. Let $\bar{K} = (\bar{k}_1, \bar{k}_2, \dots) \subseteq \mathbb{N}$ be a diverging subsequence such that $x^{\bar{k}_n} \to x'$. Since $f(x^{\bar{k}_n}, y^{\bar{k}_n}) \leq f(x, y^{\bar{k}_n})$ for all $x \in X$, by continuity of f, we have $f(x', y^\star) \leq f(x, y^\star)$ for all $x \in X$. Since the minimizer must be unique by strong convexity, we have $x' = x^\star$. In particular, this shows that $x^\star$ is the unique accumulation point of x^k: therefore, $x^k \to x^\star$. ∎

The proof of Theorem 1 is based on the following result.

Lemma 4: The operator $J_{\Phi^{-1}\mathcal{A}_\alpha}$ is continuous.

Proof: For each $i \in \mathcal{I}$, the mapping $\tilde{x}_i \mapsto \operatorname{argmin}_{y \in \Omega_i}(f_i(y, \tilde{x}_{i,-i}) + \frac{1}{\alpha}\|y - \tilde{x}_{i,i}\|^2)$ is continuous by Lemma 3. The result follows by Lemma 2 and the explicit form of $J_{\Phi^{-1}\mathcal{A}_\alpha}$ in Algorithm 1. ∎

We are now in a position to apply the results on proximal-point algorithm for restricted monotone operators in [16]. First, note that the operator $\mathcal{A}_\alpha$ is restricted monotone with respect to $\operatorname{zer}(\mathcal{A}_\alpha)$ (because $\mathcal{F}_\alpha$ is so by assumption, and by monotonicity of the normal cone [13, Th. 20.25]), i.e., for all $(\boldsymbol{x}, \boldsymbol{u}), (x^\star, \boldsymbol{u}^\star) \in \operatorname{gra}(\mathcal{A}_\alpha)$, with $x^\star \in \operatorname{zer}(\mathcal{A}_\alpha)$

$$0 \leq \langle \boldsymbol{u} - \boldsymbol{u}^\star, \boldsymbol{x} - \boldsymbol{x}^\star \rangle = \langle \Phi^{-1}\boldsymbol{u} - \Phi^{-1}\boldsymbol{u}^\star, \boldsymbol{x} - \boldsymbol{x}^\star \rangle_\Phi,$$

which shows that $\Phi^{-1}\mathcal{A}_\alpha$ is restricted monotone with respect to $\operatorname{zer}(\mathcal{A}_\alpha)$ in $\mathcal{H}_\Phi$. Therefore, by Lemma 2 and by applying [16, Th. 1(i)], we infer that the sequence $(\boldsymbol{x}^k)$ is bounded, hence it admits at least one cluster point, say $\bar{\boldsymbol{x}}$. By [16, Th. 1(ii)], $J_{\Phi^{-1}\mathcal{A}_\alpha}(\boldsymbol{x}^k) - \boldsymbol{x}^k \to 0$; therefore, by continuity in Lemma 4, it must be $\bar{\boldsymbol{x}} \in \operatorname{fix}(J_{\Phi^{-1}\mathcal{A}_\alpha}) = \operatorname{zer}(\mathcal{A}_\alpha)$. The conclusion follows by [16, Th. 1(iii)]. ∎

REFERENCES

[1] H. Dürr, M. S. Stanković, and K. H. Johansson, "Distributed positioning of autonomous mobile sensors with application to coverage control," in *Proceedings of the 2011 American Control Conference*, 2011.

[2] G. Belgioioso, A. Nedic, and S. Grammatico, "Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks," *IEEE Transactions on Automatic Control*, vol. 66, 2021.

[3] T. Tatarenko, W. Shi, and A. Nedic, "Geometric convergence of gradient play algorithms for distributed Nash equilibrium seeking," *IEEE Transactions on Automatic Control*, vol. 66, 2021.

[4] D. Gadjov and L. Pavel, "A passivity-based approach to Nash equilibrium seeking over networks," *IEEE Transactions on Automatic Control*, vol. 64, 2019.

[5] J. Koshal, A. Nedić, and U. V. Shanbhag, "Distributed algorithms for aggregative games on graphs," *Operations Research*, vol. 64, 2016.

[6] D. Gadjov and L. Pavel, "Single-timescale distributed GNE seeking for aggregative games over networks via forward-backward operator splitting," *IEEE Transactions on Automatic Control*, vol. 66, 2021.

[7] M. Ye and G. Hu, "Distributed Nash equilibrium seeking by a consensus based approach," *IEEE Transactions on Automatic Control*, vol. 62, 2017.

[8] C. De Persis and S. Grammatico, "Distributed averaging integral Nash equilibrium seeking on networks," *Automatica*, vol. 110, 2019.

[9] M. Bianchi and S. Grammatico, "Fully distributed Nash equilibrium seeking over time-varying communication networks with linear convergence rate," *IEEE Control Systems Letters*, vol. 5, no. 2, 2021.

[10] L. Pavel, "Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, 2020.

[11] J. Lei, U. V. Shanbhag, and J. Chen, "Distributed computation of Nash equilibria for monotone aggregative games via iterative regularization," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020.

[12] D. Gadjov and L. Pavel, "On the exact convergence to Nash equilibrium in hypomonotone regimes under full and partial-decision information," *IEEE Transactions on Automatic Control*, DOI:10.1109/TAC.2022.3204543. 2022.

[13] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer New York, 2017.

[14] G. Scutari, F. Facchinei, J. S. Pang, and D. P. Palomar, "Real and complex monotone communication games," *IEEE Transactions on Information Theory*, vol. 60, 2014.

[15] P. Yi and L. Pavel, "Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms," *IEEE Transactions on Control of Network Systems*, vol. 6, 2019.

[16] M. Bianchi, G. Belgioioso, and S. Grammatico, "Fast generalized Nash equilibrium seeking under partial-decision information," *Automatica*, vol. 136, 2022.

[17] ——, "A fully-distributed proximal-point algorithm for Nash equilibrium seeking with linear convergence rate," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020.

[18] Y. Huang and J. Hu, "Distributed solution of GNEP over networks via the Douglas-Rachford splitting method," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021.

[19] F. Facchinei and J.-S. Pang, *Nash equilibria: the variational approach*. Cambridge University Press, 2009.

[20] Y. Huang and J. Hu, "Distributed computation of stochastic GNE with partial information: An augmented best-response approach," 2021. [Online]. Available: https://arxiv.org/abs/2109.12290

[21] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer New York ed., 2003.