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 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 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 of other non-neighboring agents).

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 continuous-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, \( 0_n \) 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. This work was supported by NWO under research project OMEGA (613.001.702) and by the ERC under research project COSMOS (802348).
2) Euclidean spaces: Given a positive definite matrix $P \in \mathbb{R}^{n \times n}$, $H_P := (\mathbb{R}^n, \langle \cdot, \cdot \rangle_P)$ is the Euclidean space obtained by endowing $\mathbb{R}^n$ with the $P$-weighted inner product $(x, y)_P := x^T P y$, and $\| \cdot \|_P$ is the associated norm; we omit the subscripts if $P = I$. Unless otherwise stated, we always assume to work in $H = H_I$.

3) Operator-theoretic background [13]: A set-valued operator $F : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is characterized by its graph $\text{gra}(F) := \{(x, u) | x \in F(u)\}$. $\text{dom}(F) := \{x \in \mathbb{R}^q | F(x) \neq \emptyset\}$, $\text{fix}(F) := \{x \in \mathbb{R}^q | x \in F(x)\}$ and $\text{zer}(F) := \{x \in \mathbb{R}^q | 0 \in F(x)\}$ are the domain, set of fixed points and set of zeros, respectively. $F$ is strongly monotone if $\langle x - y, F(x) - F(y) \rangle \geq \|x - y\|^2$ for all $x, y \in \mathbb{R}^q$ (equivalently, $F$ is $\beta$-strongly monotone). $F$ is Lipschitz continuous if $\|F(x) - F(y)\| \leq \|x - y\|$ (equivalently, $F$ is $\beta$-Lipschitz). $\text{Id}$ is the identity operator, $I_F := (\text{Id} + F^{-1})^{-1}$ denotes the resolvent operator of $F$. For a function $\psi : \mathbb{R}^q \to \mathbb{R} \cup \{\infty\}$, $\text{dom}(\psi) := \{x \in \mathbb{R}^q | \psi(x) < \infty\}$; its subdifferential is $\partial \psi : \text{dom}(\psi) \rightrightarrows \mathbb{R}^q : x \to \{v \in \mathbb{R}^q | \psi(z) \geq \psi(x) + \langle v, z - x \rangle, \forall z \in \text{dom}(\psi)\}$. It is differentiable, $\partial \psi = \nabla \psi$. For a set $S \subseteq \mathbb{R}^n$, $\iota_S : \mathbb{R}^n \to \{0, \infty\}$ is the indicator function, i.e., $\iota_S(x) = 0$ if $x \in S$, otherwise, $\iota_S : \mathbb{R}^n \to \mathbb{R}^n : x \to \{v \in \mathbb{R}^n | \sup_{z \in S} \|v - z\| \leq 0\}$ is the normal cone of $S$. If $S$ is closed and convex, then $\partial \iota_S = \iota_N_S$ and $(\text{Id} + \iota_N_S)^{-1} = \text{proj}_S$ is the Euclidean projection onto $S$.

Definition 1 (Restricted monotonicity): An operator $F : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is restricted (strictly, $\mu$-monotone) in $H_P$ with respect to a set $\Sigma \neq \emptyset$ if $(x - x^*, u - u^*)_P \geq 0$ $(> 0, \geq \mu \|x - x^*\|_P^2$ for all $(x, u) \in \text{gra}(F), (x^*, u^*) \in \text{gra}(F)$ with $x^* \in \Sigma$. We omit the characterization in “$H_P$” whenever $P = I$.

This definition slightly generalizes that in [16, Def. 1], which only considers the zero set; note that $F$ is allowed to be set-valued on $x^* \in \Sigma$. Proximal-point algorithm: For an operator $F : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ with $\text{zer}(F) \neq \emptyset$, consider the problem of finding a point $x^* \in \text{zer}(F)$. The following iteration is called PPA:

$$x^{k+1} \in J_F(x^k) := (\text{Id} + F^{-1})^{-1} x^k. \quad (1)$$  

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

III. MATHEMATICAL SETUP

A. The game

Let $\mathcal{I} := \{1, \ldots, N\}$ 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 $x := \text{col}(\{x_i\}_{i \in \mathcal{I}}) \in \Omega$ the stacked vector of all the agents’ strategies, with $\Omega := \Omega_1 \times \cdots \times \Omega_N \subseteq \mathbb{R}^n$ 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 $x_{-i} := \text{col}(\{x_j\}_{j \in \mathcal{I} \setminus \{i\}})$. The game consists of $N$ inter-dependent optimization problems

$$\forall i \in \mathcal{I} : \arg \min_{y_i} f_i(y_i, x_{-i}). \quad (2)$$

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 $x^* = \text{col}(\{x^*_i\}_{i \in \mathcal{I}})$ such that, for all $i \in \mathcal{I}$, $x^*_i \in \arg \min_{y_i} f_i(y_i, x^*_{-i})$.

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 $\Omega_i$ is nonempty, closed and convex; the function $f_i$ 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 $\Omega$) can be found for instance in [19].

B. The communication graph

The agents can exchange information with some neighbors over an undirected communication network $G(\mathcal{I}, \mathcal{E})$. 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 $G$, with $w_{i,j} := [W]_{i,j}$ and $w_{i,j} > 0$ if $(i, j) \in \mathcal{E}, w_{i,j} = 0$ otherwise; $N_i = \{j | (i, j) \in \mathcal{E}\}$ the set of neighbors of agent $i$.

Assumption 3 (Connectivity): The communication graph $G(\mathcal{I}, \mathcal{E})$ 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: $W 1_N = 1_N, 1_N^\top W = 1_N^\top$.

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 $\Omega_i$ 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 $G$. We denote $x_i := \text{col}(\{x_{i,j}\}_{j \in \mathcal{I}}) \in \mathbb{R}^{n_i}$, where $x_{i,i} := x_i$ and $x_{i,j}$
is agent $i$’s estimate of agent $j$’s strategy, for all $j \neq i$; 
$x_{j,-i} = \text{col}(x_{j,l})_{l \in I \setminus \{i\}}$; $x = \text{col}(x_i)_{i \in I} \in \mathbb{R}_n^N$ the overall estimate vector; $x_{-i} = \text{col}(x_j)_{j \in I \setminus \{i\}}$. Let

$$\mathcal{R}_i := \left[ 0_{n_i \times n_{c_i}}, I_{n_i}, 0_{n_i \times n_{c_i}} \right],$$

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

**D. Game mapping, extended mapping, augmented operators**

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

$$0_n \in F(x^*) + N_\Omega(x^*),$$

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

$$F(x) := \text{col}((\partial x_i f_i(x_i, x_{-i}))_{i \in I})$$

(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

$$F(x) := \text{col}((\partial x_i f_i(x_i, x_{i,-i}))_{i \in I})$$

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

$$\mathcal{F}_\alpha(x) := \alpha \mathcal{R}^T F(x) + (I_{N_n} - W)x$$

$$A_{\alpha}(x) := \mathcal{F}_\alpha(x) + N_\Omega(x),$$

where $\alpha > 0$ is a design parameter, $W := W \otimes I_{n_i}$, $\Omega := \{x \in \mathbb{R}^{Nn} \mid \mathcal{R}x \in \Omega\}$. 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) $x^* = 1_N \otimes x^*$, with $x^* \in \Omega$ a NE of the game (2);
ii) $0_{N_n} \in A_{\alpha}(x^*)$.

In particular, Assumption 2 implies that $\text{zer}(A_{\alpha}) \neq \emptyset$.

**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 $F, \mathcal{R}\mathcal{T} F, \mathcal{F}_\alpha, A_{\alpha}$. 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 $\mathcal{R}^T F$ is maximally monotone.

**C2:** The operator $\mathcal{R}^T F$ is restricted monotone with respect to $\text{zer}(A_{\alpha})$.

**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 $\text{zer}(A_{\alpha})$.

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

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

**C7:** The operator $F$ is $\nu$-hypo-monotone, $\ell$-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 $\ell$-Lipschitz, for some $\ell > 0$.

**C9:** The operator $F$ is $1/\ell$ cocoercive for some $\ell > 0$.

**C10:** The operator $F$ is monotone and $\ell$-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 $\nabla f_i(x, x_{-i})$ is independent of $x_{-i}$, for all $i \in I$.

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 $[1, 0]^T$ and satisfies C2, but not C1.
Although \( \nabla_x f_i \) depends on \( x_{-i} \) in Example 1, the next lemma shows that \( C_2 \) is also not of particular interest.

**Proposition 3 (C2 is trivial):** Assume that \( F \) is single valued and continuous. Then, condition \( C_2 \) holds if and only if \( \nabla_x f_i(x^*_i, x_{-i}) \) is independent of \( x_{-i} \), for all \( i \in I \), for any \( x^* = (x^*_i, x^*_{-i}) \) NE of the game \( (2) \).

In particular, Proposition 3 implies that \( 0 \leq \nabla_x f_i(x^*_i, x_{-i}), x_i - x^*_i \) = \( \langle \nabla_x f_i(x^*_i, x_{-i}), x_i - x^*_i \rangle \) where the inequality is the first order optimality condition (as \( x^*_i \) solves \( 2 \)). This means that \( x^*_i \) is optimal for agent \( i \) regardless of \( x_{-i} \); in other terms, \( C_2 \) 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 \( C_5 \) to \( C_{10} \) are directly postulated on the game primitives \( F \) and \( R \), and are the most well-investigated (e.g., they are easy to check if \( F \) is a linear operator \([13], [21], [12] \)). Conditions \( C_5 \) to \( C_8 \) imply uniqueness of the equilibrium; methods with linear convergence were proposed under \( C_5, C_6, C_7, C_8 \). Although \( C_5 \) is weaker than \( C_6 \) 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 \( C_6, C_7, C_8 \). Actually, it is straightforward to check that \( C_6 \) implies that there is \( \alpha > 0 \) such that \( F_{x_i} \) is Lipschitz and restricted strongly-monotone with respect to the whole consensus subspace \( E := \{ y \in \mathbb{R}^{n} | y = I_N \otimes y, y \in \mathbb{R}^n \} \cap \text{zer}(A_\alpha) \) \([10], \text{Lem. 3} \), a much more restrictive condition that \( C_4, C_{10} \) allow for multiple NEs; yet – as for \( C_8 \) – 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**

Conditions \( C_3 \) and \( C_4 \) are more abstract and often replaced by more easily checked sufficient conditions. For example, restricted monotonicity of \( F_{x_i} \) with respect to the consensus space \( E \) can be easily checked without knowledge of the solutions, and implies \( C_4 \).

Despite this complication, \( C_3 \) and \( C_4 \) are of great interest, especially for nonsmooth games, as exemplified next. The following examples also show that \( C_3 \) is significantly more restrictive than \( C_4 \).

**Example 2:** Consider the game defined by \( N = 2, n = 2, \Omega = \mathbb{R}^n, F(x) = \bar{F}(x) + F(x), \) with \( F(x) = \text{col}(x_1^3, 0) \) and \( \bar{F}(x) = [\frac{1}{2} \ 0] x^2 + [\frac{1}{2}] \). As \( \bar{F} \) is monotone and \( F \) is strongly monotone, the game admits a unique NE. Conditions \( C_5 \) to \( C_{10} \) are violated, as they require Lipschitz continuity of \( F \); \( C_2 \) also fails (as the best response of agent 2 is \( -0.5x^1 - 2 \) by Proposition 3). However, \( C_4 \) holds. To show this, consider the components of the extended game mapping \( E \) and \( F \) corresponding to \( \bar{F} \) and \( F \); \( R^\top \bar{F} \) is monotone, while \( \alpha R^\top F + (I - W) \) can be made restricted monotone with respect to the consensus subspace by choosing \( \alpha > 0 \) small enough \([10], \text{Lem. 3} \). We can check numerically that \( C_3 \) also holds for some \( W \) (in particular, because \( \alpha R^\top F + (I - W) \) 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 \( F(x) = \text{col}(x_1^3(x_2 + 1), 0) \) and \( \bar{F}(x) = [\frac{1}{2} \ 0] x \). The game admits a NE \( x^* = 0 \). As \( F \) is restricted strongly monotone with respect to \( x^* \), the equilibrium must be unique. As for Example 2, it is easy to prove that \( C_4 \) holds, because \( R^\top \bar{F} \) is restricted monotone with respect to \( 0 \) (by Proposition 3). Yet, \( F \) is not monotone; therefore \( C_3 \) cannot hold (nor can \( C_1 \) to \( C_2, C_5 \) to \( C_{10} \)).

**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 \( |x| \) is the absolute value. The game admits a unique NE in \( 0 \); moreover, \( F \) is set valued, as \( f_1 \) is not differentiable in the local variable. It can be checked that \( C_4 \) holds. Yet, \( F \) is not monotone, thus \( C_3 \) 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 \( x_{i, \cdots} \), 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 argmin 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 \( A_\alpha \). However, the computation of \( (I + A_\alpha)^{-1} \) 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 \( A_\alpha \) with a positive definite matrix

\[
\Phi := \hat{I}_n + W. \tag{9}
\]

**Lemma 2:** Algorithm 1 can be written as

\[
x_{k+1} = \frac{1}{2} (x_k + \sum_{j=1}^N w_{ij} x_j^k) \tag{10}
\]
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 argmin is only approximated at each iteration), coupling constraint, acceleration terms [16]. It also immediately shows that the fixed points of Algorithm 1 coincide with \( \text{zer}(A_n) = \text{zer}(\Phi^{-1}A_n) \) (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 \((x^k)\) generated by Algorithm 1 converges to a point \( x^* = 1_N \otimes x^* \), where \( x^* \) 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 \( J_{A_i^{-1}A_n} \) is continuous, even if \( A_n \) 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 \( x^k \)'s are initialized in \( \Omega^N \) (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 \( q_i(x_i) \) (besides the indicator function \( 1_{\Omega_i} \)), 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 PPA 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 \( I - W \) is maximally monotone. Hence, for any \( \alpha \geq 0, F_\alpha = \alpha \Delta^{-1}F + (I - W) \) is the sum of two maximally monotone operators; moreover, \( \text{dom}(I - W) = \mathbb{R}^{Nn} \), so the conclusion follows by [13, Cor. 25.5].

2) **Proof of Proposition 2:** \( \Rightarrow \): 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 \( \nabla x_i f_i(x_i, x_{-i}) \rangle < \langle \nabla x_i f_i(x_i, x'_{-i}) \rangle \). By continuity, there exists \( \epsilon > 0 \) such that \( \nabla x_i f_i(x_i + \epsilon e_i, x_{-i}) \rangle < \langle \nabla x_i f_i(x_i, x'_{-i}) \rangle \), where \( e_i \in \mathbb{R}^{n_i} \) is the \( i \)-th vector of the canonical basis. The monotonicity in C1, applied to a pair of estimate vectors \( (x_i, x_{-i}), (x'_i, x'_{-i}), \) for any \( x_{-i} \) and \( x_i = (x_i + \epsilon e_i, x_{-i}), x'_i = (x_i, x'_{-i}) \), gives

\[
0 \leq \langle \nabla x_i f_i(x_i + \epsilon e_i, x_{-i}) - \nabla x_i f_i(x_i, x'_{-i}), \epsilon e_i \rangle = \epsilon \langle \nabla x_i f_i(x_i + \epsilon e_i, x_{-i}) - \nabla x_i f_i(x_i, x'_{-i}) \rangle < 0
\]

which is a contradiction. Analogously it can be shown that \( \langle \nabla x_i f_i(x_i, x_{-i}) \rangle > \langle \nabla x_i f_i(x'_i, x'_{-i}) \rangle \) leads to a contradiction. Hence \( \nabla x_i f_i(x'_i, x_{-i}) = \nabla x_i f_i(x_i, x_{-i}) \).

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 \( x^* \) and \( x_{-i} \) such that \( \nabla x_i f_i(x^*_i, x_{-i}) \rangle < \langle \nabla x_i f_i(x^*_i, x_{-i}) \rangle \). By continuity, there exists \( \epsilon > 0 \) such that \( \nabla x_i f_i(x^*_i + \epsilon e_i, x_{-i}) \rangle < \langle \nabla x_i f_i(x^*_i, x_{-i}) \rangle \). Restricted monotonicity in C2, applied to a pair of estimate vectors \( (x_i, x_{-i}), (x^*_i, x_{-i}), \) for any \( x_{-i} \) and \( x_i = (x^*_i + \epsilon e_i, x_{-i}) \), gives

\[
0 \leq \langle \nabla x_i f_i(x^*_i + \epsilon e_i, x_{-i}) - \nabla x_i f_i(x^*_i, x_{-i}), \epsilon e_i \rangle = \epsilon \langle \nabla x_i f_i(x^*_i + \epsilon e_i, x_{-i}) - \nabla x_i f_i(x^*_i, x_{-i}) \rangle < 0
\]

which is a contradiction. Analogously it can be shown that \( \langle \nabla x_i f_i(x^*_i, x_{-i}) \rangle > \langle \nabla x_i f_i(x^*_i, x_{-i}) \rangle \) leads to a contradiction. Hence \( \nabla x_i f_i(x^*_i, x_{-i}) = \nabla x_i f_i(x^*_i, x_{-i}) \).
⇒: For any $i \in I$, $x_i, x_{-i}, \text{NE } x^\star$, by assumption and convexity, $\langle f_i(x_i, x_{-i}) - \nabla x f_i(x_i, x_{-i}^\star), x_i - x_i^\star \rangle = \langle \nabla_x f_i(x_i, x_{-i}) - \nabla_x f_i(x_i^\star, x_{-i}), x_i - x_i^\star \rangle \geq 0$.

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 $f(y, y)$ is $\mu$-strongly convex for any $y \in \mathbb{R}^n$, $\mu > 0$. Let $X \subseteq \mathbb{R}^n$ be a convex closed set. Then the (single valued, full domain) mapping $y \mapsto g(y) = \min_{x \in X} f(x, y)$ is continuous.

Proof: We show that for any given sequence $(y_k)_{k \in \mathbb{N}}$ with $y_k \to y^*$ (converging, hence bounded), $x_k := g(y_k) \to g(y^*) =: x^*$, 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

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

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

$$M > 1, \quad M > 2 \frac{\|d\|}{\mu} + 1. \quad (11)$$

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

$$l_1 \leq f(x_1, y) \leq \frac{M-1}{M} f(x_0, y) + \frac{1}{M} f(x_2, y) - \frac{1}{2} \mu \frac{M-1}{M} \|x_2 - x_1\|^2$$

$$= \frac{M-1}{M} f(x_0, y) + \frac{1}{M} f(x_2, y) - \frac{1}{2} \mu (M-1).$$

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 $\|x - x_1\| > M$, $f(x, y) < f(x_0, y)$. In turn, for all $y \in Y$, $\|y\| < \|x_0\| + M$, 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 $K = (k_1, k_2, . . .) \subseteq \mathbb{N}$ be a diverging subsequence such that $x_{k_n} \to x^*$. Since $f(x_{k_n}, y_{k_n}) \leq f(x_{k_n}, y^*)$ for all $x_n \in X$, by continuity of $f$, we have $f(x^*, y^*) \leq f(x^*, y)$ for all $x \in X$. Since the minimizer must be unique by strong convexity, we have $x^* = x^*$. In particular, this shows that $x^*$ is the unique accumulation point of $x_k$; therefore, $x_k \to x^*$.

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

**Lemma 4:** The operator $J_{\Phi^{-1}A_0}$ is continuous.

Proof: For each $i \in I$, the mapping $\tilde{x}_i \mapsto \arg\min_{y \in Y} f_i(y, \tilde{x}_{-i}) + \frac{1}{2} \|y - \tilde{x}_i\|^2$ is continuous by Lemma 3. The result follows by Lemma 2 and the explicit form of $J_{\Phi^{-1}A_0}$ 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 $A_0$ is restricted monotone with respect to $\text{zer}(A_0)$ (because $F_\alpha$ is so by assumption, and by monotonicity of the normal cone [13, Th. 20.25]), i.e., for all $(x, u), (x^*, u^*) \in \text{gra}(A_0)$, with $x^* \in \text{zer}(A_0)$

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

which shows that $\Phi^{-1}A_0$ is restricted monotone with respect to $\text{zer}(A_0)$ in $H_\Phi$. Therefore, by Lemma 2 and by applying [16, Th. 1(i)], we infer that the sequence $(x_k^\star)$ is bounded, hence it admits at least one cluster point, say $\bar{x}$. By [16, Th. 1(ii)], $J_{\Phi^{-1}A_0}(x_k^\star) - x_k^\star \to 0$; therefore, by continuity in Lemma 4, it must be $\bar{x} \in \text{fix}(J_{\Phi^{-1}A_0}) = \text{zer}(A_0)$. The conclusion follows by [16, Th. 1(iii)].