Fast generalized Nash equilibrium seeking under partial-decision information

Bianchi, Mattia; Belgioioso, Giuseppe; Grammatico, Sergio

DOI
10.1016/j.automatica.2021.110080

Publication date
2022

Document Version
Final published version

Published in
Automatica

Automatica 136 (2022) 110080

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Fast generalized Nash equilibrium seeking under partial-decision information☆

Mattia Bianchi [a,*], Giuseppe Belgioioso [b], Sergio Grammatico [a]

[a] Delft Center for Systems and Control, Delft University of Technology, The Netherlands
[b] Automatic Control Laboratory, Swiss Federal Institute of Technology (ETH) Zürich, Switzerland

ARTICLE INFO

Article history:
Received 25 October 2020
Received in revised form 11 June 2021
Accepted 20 October 2021
Available online xxxx

Keywords:
Nash equilibrium seeking
Proximal-point method
Distributed algorithms
Multi-agent systems

ABSTRACT

We address the generalized Nash equilibrium seeking problem in a partial-decision information scenario, where each agent can only exchange information with some neighbors, although its cost function possibly depends on the strategies of all agents. The few existing methods build on projected pseudo-gradient dynamics, and require either double-layer iterations or conservative conditions on the step sizes. To overcome both these flaws and improve efficiency, we design the first fully-distributed single-layer algorithms based on proximal best-response. Our schemes are fixed-step and allow for inexact updates, which is crucial for reducing the computational complexity. Under standard assumptions on the game primitives, we establish convergence to a variational equilibrium (with linear rate for games without coupling constraints) by recasting our algorithms as proximal-point methods, opportunely preconditioned to distribute the computation among the agents. Since our analysis hinges on a restricted monotonicity property, we also provide new general results that significantly extend the domain of applicability of proximal-point methods. Besides, our operator-theoretic approach favors the implementation of provably correct acceleration schemes that can further improve the convergence speed. Finally, the potential of our algorithms is demonstrated numerically, revealing much faster convergence with respect to projected pseudo-gradient methods and validating our theoretical findings.

1. Introduction

Generalized games model the interaction between self-interested decision makers, or agents, that aim at optimizing their individual, yet inter-dependent, objective functions, subject to shared constraints. This competitive scenario has received increasing attention with the spreading of networked systems, due to the numerous engineering applications, including demand-side management in the smart grid (Saad, Han, Poor, & Basar, 2012), charging/discharging of electric vehicles (Grammatico, 2017), demand response in competitive markets (Li, Chen, & Dahleh, 2015), and radio communication (Facchinei & Pang, 2009). From a game-theoretic perspective, the challenge is to assign the agents behavioral rules that eventually ensure the attainment of a satisfactory equilibrium.

A recent part of the literature focuses in fact on designing *distributed* algorithms to seek a generalized Nash equilibrium (GNE), a decision set from which no agent has interest to unilaterally deviate (Belgioioso & Grammatico, 2018; Chen, Ming, & Hong, 2021; Facchinei & Kanzow, 2010; Yi & Pavel, 2019; Yu, van der Schaar, & Sayed, 2017). In these works, the computational effort is partitioned among the agents, but assuming that each of them has access to the decision of all the competitors (or to an aggregation value, in the case of aggregative games). Such a hypothesis, referred to as *full-decision information*, generally requires the presence of a central coordinator that communicates with all the agents, which is impractical in some cases (Frihauf, Krstic, & Basar, 2012; Swenson, Kar, & Xavier, 2015). One example is the Nash–Cournot competition model described in Koshal, Nedić, and Shanbhag (2016), where the profit of each of a group of firms depends not only on its own production, but also on the total supply, a quantity not directly accessible by any of the firms. Instead, in this paper we consider the so-called *partial-decision information* scenario, where each agent estimates the actions of all the competitors by relying only on the information exchanged with some neighbors over a communication network. Thus, the goal is to design *fully-distributed* (namely, center-free) algorithms, based exclusively on peer-to-peer communication.

☆ This work is supported by the NWO, Netherlands under project OMEGA (613.001.702) and by the ERC under project COSMOS (802348). The material in this paper was partially presented at the 59th IEEE Conference on Decision and Control, December 14–18, 2020, Jeju Island, Republic of Korea. This paper was recommended for publication in revised form by Associate Editor Gurdal Arslan under the direction of Editor Ian R. Petersen.

* Corresponding author.
E-mail addresses: M.Bianchi@tudelft.nl (M. Bianchi), gbelgioioso@ethz.ch (G. Belgioioso), s.grammatico@tudelft.nl (S. Grammatico).

https://doi.org/10.1016/j.automatica.2021.110080

The partial-decision information setup has only been introduced very recently. Most results consider non-generalized games (i.e., games without shared constraints) (Koshal et al., 2016; Salehisadaghiani & Pavel, 2016; Salehisadaghiani, Shi, & Pavel, 2019; Tatarenko, Shi, & Nedić, 2020). Even fewer algorithms can cope with the presence of coupling constraints (Belgioioso, Nedić, & Grammatico, 2021; Gadjov & Pavel, 2021; Pavel, 2020), despite this extension arises naturally in most resource allocation problems (Facchinei & Kanzow, 2010, §2), e.g., due to shared capacity limitations. All the cited formulations resort to (projected) gradient and consensus-type dynamics, and are single-layer (i.e., they require a fixed finite number of communications per iteration). The main drawback is that, due to the partial-decision information assumption, theoretical guarantees are obtained only for small (or vanishing) step sizes, which significantly affect the speed of convergence. The only alternative available in literature consists of double-layer algorithms (Lei & Shanbhag, 2018; Parise, Gentile, & Lygeros, 2020), where the agents must communicate multiple (virtually infinite) times to reach consensus, before each update. An extensive communication requirement is however a performance bottleneck, as the communication time can overwhelm the time spent on local useful processing — in fact, this is a common problem in parallel computing (Ivkin et al., 2019). Let alone the time lost in the transmission, sending large volumes of data on wireless networks results in a dramatically increased energetic cost.

Contributions: To improve speed and efficiency, we design the first fully-distributed single-layer GNE seeking algorithms based on proximal best-response. For the sake of generality and mathematical elegance, we take here an operator-theoretic approach (Belgioioso & Grammatico, 2017; Yi & Pavel, 2019), and reformulate the GNE problem as that of finding a zero of a monotone operator. The advantage is that several fixed-point iterations are known to solve monotone inclusions (Bauschke & Combettes, 2017, §26), thus providing a unifying framework to design algorithms and study their convergence. For instance, the methods in Belgioioso et al. (2021), Gadjov and Pavel (2021) and Pavel (2020), were developed based on the (preconditioned) forward–backward (FB) splitting (Bauschke & Combettes, 2017, §26.5). To enhance the convergence speed, we instead employ a proximal-point algorithm (PPA) (Bauschke & Combettes, 2017, Th. 28.1), which typically can tolerate much larger step sizes. Nonetheless, the design of distributed GNE seeking PPAs was elusive until now, because a direct implementation results in double-layer algorithms (Scutari, Facchinei, Pang, & Palomar, 2014; Yi & Pavel, 2019). The novelties of this work are summarized as follows:

- We propose the first PPA to compute a zero of a *restricted* monotone operator, which significantly generalizes classical results for maximally monotone operators. Differently from other recent extensions (El Farouq, 2001; Moudafi, 2020), we also allow for set-valued resolvents and inexact updates, and we do not assume pseudomonotonicity or hypomonotonicity. This is a fundamental result of independent interest, which we exploit to prove convergence of our algorithms (Section 4.2);
- We introduce a novel primal–dual proximal best-response GNE seeking algorithm, which is the first non-gradient-based scheme for the partial-decision information setup. We derive our method as a PPA, where we design a novel preconditioning matrix to distribute the computation and obtain a single-layer iteration. Under strong monotonicity and Lipschitz continuity of the game mapping, we prove global convergence with fixed step sizes, by exploiting restricted monotonicity properties. Convergence is retained even if the proximal best-response is computed inexactly (with summable errors), which is crucial for practical implementation. Differently from Pavel (2020, Alg. 1), the step sizes can be chosen independently of a certain restricted strong monotonicity constant. In turn, not only we allow for much larger steps, but parametric dependence is also improved: for instance, the bounds do not vanish when the number of agents grows, and the resulting convergence rate for non-generalized games is superior. Moreover our scheme requires only one communication per iteration, instead of two (Section 4.3, Section 5.1);
- We apply some acceleration schemes (Iutzeler & Hendrickx, 2019) to our preconditioned PPA (PPPA) and provide new theoretical convergence guarantees. We observe numerically that the iterations needed to converge can be halved (Section 5);
- We tailor our method to efficiently solve aggregative games, by letting each agent keep and exchange an estimate of the aggregative value only, instead of an estimate of all the other agents' actions (Section 6);
- Via numerical simulations, we show that our PPPAs significantly outperform the pseudo-gradient methods in Gadjov and Pavel (2019) and Pavel (2020) (the only other known fully-distributed, single-layer, fixed-step GNE seeking schemes), not only in terms of number of iterations needed to converge (hence with a considerable reduction of the communication burden), but also in terms of total computational cost (despite each agent must locally solve a strongly convex optimization problem, rather than a projection, at each step) (Section 7).

Some preliminary results of this paper appeared in Bianchi, Belgioioso, and Grammatico (2020), where we study only one special case for games without coupling constraints and with exact computation of the resolvent (and where we do not consider aggregative games or acceleration schemes), see Section 5.1.

Basic notation: $\mathbb{N}$ is the set of natural numbers, including 0. $\mathbb{R}$ ($\mathbb{R}_{\geq 0}$) is the set of (nonnegative) real numbers. $\mathbf{0}_q \in \mathbb{R}^q$ ($\mathbf{1}_q \in \mathbb{R}^q$) is a vector with all elements equal to 0 (1); $I_q \in \mathbb{R}^{q\times q}$ is an identity matrix; the subscripts may be omitted when there is no ambiguity. For a matrix $A \in \mathbb{R}^{p\times q}$, $[A]_{i,j}$ is the element on row i and column j; $\operatorname{null}(A) := \{x \in \mathbb{R}^q \mid Ax = \mathbf{0}_n\}$ and $\operatorname{range}(A) := \{v \in \mathbb{R}^p \mid v = Ax, x \in \mathbb{R}^q\}$; $\|A\|_\infty$ is the maximum of the absolute row sums of A. If $A = A^\top \in \mathbb{R}^{q\times q}$, $\lambda_{\min}(A) =: \lambda_1(A) \leq \cdots \leq \lambda_q(A) =: \lambda_{\max}(A)$ denote its eigenvalues. $\operatorname{diag}(A_1, \ldots, A_N)$ is the block diagonal matrix with $A_1, \ldots, A_N$ on its diagonal. Given N vectors $x_1, \ldots, x_N$, $\operatorname{col}(x_1, \ldots, x_N) := [x_1^\top \ldots x_N^\top]^\top$. $\otimes$ denotes the Kronecker product. ℓ^1 is the set of absolutely summable sequences.

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

Operator-theoretic background: A set-valued operator $\mathcal{F} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is characterized by its graph $\operatorname{gra}(\mathcal{F}) := \{(x, u) \mid u \in \mathcal{F}(x)\}$. $\operatorname{dom}(\mathcal{F}) := \{x \in \mathbb{R}^q \mid \mathcal{F}(x) \neq \varnothing\}$, $\operatorname{fix}(\mathcal{F}) := \{x \in \mathbb{R}^q \mid x \in \mathcal{F}(x)\}$ 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 (μ-strongly) monotone in $\mathcal{H}_P$ if $\langle u - v \mid x - y \rangle_P \geq 0$ ($\geq \mu\|x - y\|_P^2$) for all $(x, u), (y, v) \in \operatorname{gra}(\mathcal{F})$; we omit the indication "in $\mathcal{H}_P$" whenever $P = I$. Id is the identity operator. $\mathrm{J}_{\mathcal{F}} := (\mathrm{Id} + \mathcal{F})^{-1}$ 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; $\mathrm{N}_S : S \rightrightarrows \mathbb{R}^q : x \mapsto \{v \in \mathbb{R}^q \mid \sup_{z\in S} \langle v \mid z - x \rangle \leq 0\}$ is the normal cone operator of S. If S is closed and convex, then $\partial\iota_S = \mathrm{N}_S$ and $(\mathrm{Id} + \mathrm{N}_S)^{-1} = \mathrm{proj}_S$ is the Euclidean projection onto S. Given

$\mathcal{F} : S \to \mathbb{R}^q$, the variational inequality VI($\mathcal{F}, S$) is the problem of finding $x^* \in S$ such that $\langle \mathcal{F}(x^*) \mid x - x^* \rangle \geq 0$, for all $x \in S$ (or, equivalently, x^* such that $\mathbf{0} \in \mathcal{F}(x^*) + \mathrm{N}_S(x^*)$). We denote the solution set of VI($\mathcal{F}, S$) by SOL($\mathcal{F}, S$).

2. Mathematical setup

We consider a set of agents, $\mathcal{I} := \{1, \dots, N\}$, where each agent $i \in \mathcal{I}$ shall choose its decision variable (i.e., strategy) x_i from its local decision set $\Omega_i \subseteq \mathbb{R}^{n_i}$. Let $x := \mathrm{col}((x_i)_{i\in\mathcal{I}}) \in \Omega$ denote the stacked vector of all the agents' decisions, $\Omega := \Omega_1 \times \ldots \times \Omega_N \subseteq \mathbb{R}^n$ the overall action space and $n := \sum_{i=1}^N n_i$. The goal of each agent $i \in \mathcal{I}$ is to minimize its objective function $J_i(x_i, x_{-i})$, which depends on both the local variable x_i and on the decision variables of the other agents $x_{-i} := \mathrm{col}((x_j)_{j\in\mathcal{I}\setminus\{i\}})$. Furthermore, the feasible decisions of each agent depends on the action of the other agents via coupling constraints, which we assume affine: most of the literature focuses on this case (Belgioioso et al., 2021; Parise et al., 2020), which in fact accounts for the vast majority of practical applications (Facchinei & Kanzow, 2010, §3.2). Specifically, the overall feasible set is $\mathcal{X} := \Omega \cap \{x \in \mathbb{R}^n \mid Ax \leq b\}$, where $A := [A_1, \dots, A_N]$ and $b := \sum_{i=1}^N b_i$, $A_i \in \mathbb{R}^{m\times n_i}$ and $b_i \in \mathbb{R}^m$ being local data of agent i. The game is then represented by the inter-dependent optimization problems:

$$\forall i \in \mathcal{I}: \ \underset{y_i\in\mathbb{R}^{n_i}}{\text{minimize}} \ J_i(y_i, x_{-i}) \quad \text{s.t. } (y_i, x_{-i}) \in \mathcal{X}. \tag{1}$$

The technical problem we consider here is the computation of a GNE, namely a set of decisions that simultaneously solve all the optimization problems in (1).

Definition 1. A collective strategy $x^* = \mathrm{col}\left((x_i^*)_{i\in\mathcal{I}}\right)$ is a generalized Nash equilibrium if $J_i(x_i^*, x_{-i}^*) \leq \inf\left\{J_i(y_i, x_{-i}^*) \mid (y_i, x_{-i}^*) \in \mathcal{X}\right\}$ for all $i \in \mathcal{I}$. □

Next, we postulate some common regularity and convexity assumptions for the constraint sets and cost functions, as in, e.g., Koshal et al. (2016, Asm. 1) and Pavel (2020, Asm. 1).

Standing Assumption 1. For each $i \in \mathcal{I}$, the set Ω_i is closed and convex; $\mathcal{X}$ is non-empty and satisfies Slater's constraint qualification; J_i is continuous and $J_i(\cdot, x_{-i})$ is convex and continuously differentiable for every x_{-i}. □

As per standard practice (Parise et al., 2020; Yi & Pavel, 2019), among all the possible GNEs, we focus on the subclass of variational GNEs (v-GNEs) (Facchinei & Kanzow, 2010, Def. 3.11), which are more economically justifiable, as well as computationally tractable (Kulkarni & Shanbhag, 2012). The v-GNEs are so called because they coincide with the solutions to the variational inequality VI($F, \mathcal{X}$), where F is the *pseudo-gradient* mapping of the game:

$$F(x) := \mathrm{col}\left(\left(\nabla_{x_i} J_i(x_i, x_{-i})\right)_{i\in\mathcal{I}}\right). \tag{2}$$

Under Standing Assumption 1, x^* is a v-GNE of the game in (1) if and only if there exists a dual variable $\lambda^* \in \mathbb{R}^m$ such that the following Karush–Kuhn–Tucker (KKT) conditions are satisfied (Facchinei & Kanzow, 2010, Th. 4.8):

$$\begin{aligned} &\mathbf{0}_n \in F(x^*) + A^\top \lambda^* + \mathrm{N}_\Omega(x^*) \\ &\mathbf{0}_m \in -(Ax^* - b) + \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*). \end{aligned} \tag{3}$$

Standing Assumption 2. The pseudo-gradient mapping F in (2) is μ-strongly monotone and θ_0-Lipschitz continuous, for some μ, $\theta_0 > 0$. □

The strong monotonicity of F is sufficient to ensure existence and uniqueness of a v-GNE (Facchinei & Pang, 2007, Th. 2.3.3); it was always assumed for GNE seeking under partial-decision information with fixed step sizes (Tatarenko et al., 2020, Asm. 2; Pavel, 2020, Asm. 3) (while it is sometimes replaced by strict monotonicity or cocoercivity, under vanishing steps and compactness of $\mathcal{X}$ (Koshal et al., 2016, Asm. 2; Pang & Hu, 2020, Asm. 3; Belgioioso et al., 2021, Asm. 5).

3. Fully-distributed equilibrium seeking

In this section, we present our baseline algorithm to seek a v-GNE of the game in (1) in a fully-distributed way. Specifically, each agent i only knows its own cost function J_i and feasible set Ω_i, and the portion of the coupling constraints (A_i, b_i). Moreover, agent i does not have full knowledge of x_{-i}, and only relies on the information exchanged locally 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: $W = [w_{i,j}]_{i,j\in\mathcal{I}} \in \mathbb{R}^{N\times N}$ the symmetric weight matrix of $\mathcal{G}$, with $w_{i,j} > 0$ if $(i, j) \in \mathcal{E}$, $w_{i,j} = 0$ otherwise, and the convention $w_{ii} = 0$ for all $i \in \mathcal{I}$; $L := D - W$ the Laplacian matrix of $\mathcal{G}$, with degree matrix $D := \mathrm{diag}((d_i)_{i\in\mathcal{I}})$, and $d_i := \sum_{j=1}^N w_{i,j}$ for all $i \in \mathcal{I}$; $\mathcal{N}_i = \{j \mid (i, j) \in \mathcal{E}\}$ the set of neighbors of agent i. Moreover, we label the edges $(e_\ell)_{\ell\in\{1,\dots,E\}}$, where E is the cardinality of $\mathcal{E}$, and we assign to each edge e_ℓ an arbitrary orientation. We denote the weighted incidence matrix as $V \in \mathbb{R}^{E\times N}$, where $[V]_{\ell,i} = \sqrt{(w_{i,j})}$ if $e_\ell = (i, j)$ and i is the output vertex of e_ℓ, $[V]_{\ell,i} = -\sqrt{(w_{i,j})}$ if $e_\ell = (i, j)$ and i is the input vertex of e_ℓ, $[V]_{\ell,i} = 0$ otherwise. It holds that $L = V^\top V$; moreover, $\mathrm{null}(V) = \mathrm{null}(L) = \{\kappa \mathbf{1}_N, \kappa \in \mathbb{R}\}$ under the following connectedness assumption (Godsil & Royle, 2013, Ch. 8).

Standing Assumption 3. The communication graph $\mathcal{G}(\mathcal{I}, \mathcal{E})$ is undirected and connected. □

In the partial-decision information, to cope with the lack of knowledge, each agent keeps an estimate of all other agents' actions (Pavel, 2020; Tatarenko et al., 2020; Ye & Hu, 2017). We denote $\boldsymbol{x}_i := \mathrm{col}((\boldsymbol{x}_{i,j})_{j\in\mathcal{I}}) \in \mathbb{R}^n$, where $\boldsymbol{x}_{i,i} := x_i$ and $\boldsymbol{x}_{i,j}$ is agent i's estimate of agent j's action, for all $j \neq i$; let also $\boldsymbol{x}_{j,-i} := \mathrm{col}((\boldsymbol{x}_{j,\ell})_{\ell\in\mathcal{I}\setminus\{i\}})$. Moreover, we let each agent keep an estimate $\lambda_i \in \mathbb{R}^m_{\geq 0}$ of the dual variable, and an auxiliary variable $z_i \in \mathbb{R}^m$.

Our proposed dynamics are summarized in Algorithm 1, where the global parameter $\alpha > 0$ and the positive step sizes τ_i, δ_i, $\nu_{(i,j)} = \nu_{(j,i)}$, for all $i \in \mathcal{I}$ and $(i, j) \in \mathcal{E}$, have to be chosen appropriately (see Section 4). Each agent i updates its action x_i similarly to a proximal best-response, but with two extra terms that are meant to penalize and correct the disagreement among the estimates and the coupling constraints violation. Most importantly, the agents evaluate their cost functions in their local estimates, not on the actual collective strategy. In steady state, the agents should agree on their estimates, i.e., $\boldsymbol{x}_i = \boldsymbol{x}_j$, $\lambda_i = \lambda_j$, for all $i, j \in \mathcal{I}$. This motivates the presence of consensus terms for both primal and dual variables. From a control-theoretic perspective, the updates of each z_i can be seen as integrator dynamics driven by the disagreement of the variables λ_j's. This integral action is meant to permit the distributed asymptotic satisfaction of the coupling constraints, despite the computation of each λ_i only involves the local block (A_i, b_i) – differently from typical centralized dual ascent iterations. We postpone a formal derivation of Algorithm 1 to Section 4.

Algorithm 1 Fully-distributed v-GNE seeking via PPPA

Initialization: · For all $i \in \mathcal{I}$, set $x_i^0 \in \Omega_i$, $\boldsymbol{x}_{i,-i}^0 \in \mathbb{R}^{n-n_i}$, $z_i^0 = \mathbf{0}_m$, $\lambda_i^0 \in \mathbb{R}^m_{\geq 0}$.

For all $k \in \mathbb{N}$: · Communication: The agents exchange the variables $\{x_i^k, \boldsymbol{x}_{i,-i}^k, \lambda_i^k\}$ with their neighbors.

· Local variables update: each agent $i \in \mathcal{I}$ computes

$$\boldsymbol{x}_{i,-i}^{k+1} = \tfrac{1}{1+\tau_i d_i}\big(\boldsymbol{x}_{i,-i}^k + \tau_i \textstyle\sum_{j\in\mathcal{N}_i} w_{i,j}\boldsymbol{x}_{j,-i}^k\big)$$

$$x_i^{k+1} = \operatorname*{argmin}_{y\in\Omega_i}\Big(J_i(y, \boldsymbol{x}_{i,-i}^{k+1}) + \tfrac{1}{2\alpha\tau_i}\big\|y - x_i^k\big\|^2 + \tfrac{d_i}{2\alpha}\big\|y - \tfrac{1}{d_i}\textstyle\sum_{j\in\mathcal{N}_i} w_{i,j}\boldsymbol{x}_{j,i}^k\big\|^2 + \tfrac{1}{\alpha}(A_i^\top\lambda_i^k)^\top y\Big)$$

$$z_i^{k+1} = z_i^k + \textstyle\sum_{j\in\mathcal{N}_i} v_{(i,j)} w_{i,j}(\lambda_i^k - \lambda_j^k)$$

$$\lambda_i^{k+1} = \operatorname{proj}_{\mathbb{R}^m_{\geq 0}}\big(\lambda_i^k + \delta_i\big(A_i(2x_i^{k+1} - x_i^k) - b_i - (2z_i^{k+1} - z_i^k)\big)\big).$$

Remark 1. The functions $J_i(\cdot, \boldsymbol{x}_{i,-i})$ are strongly convex, for all $\boldsymbol{x}_{i,-i}$, $i \in \mathcal{I}$, as a consequence of Standing Assumption 2. Hence, the argmin operator in Algorithm 1 is single-valued, and the algorithm is well defined. □

Remark 2. In Algorithm 1, each agent has to *locally* solve an optimization problem, at every iteration. Not only these subproblems are fully-decentralized (i.e., they do not require extra communication), but they are also of low dimension (n_i). This is a major departure from the procedure proposed in the PPAs of Scutari et al. (2014, Alg. 2), Yi and Pavel (2019, Alg. 2), where the agents have to collaboratively solve a subgame (of dimension n) before each update. □

4. Convergence analysis

4.1. Definitions and preliminary results

We denote $\boldsymbol{x} := \operatorname{col}((\boldsymbol{x}_i)_{i\in\mathcal{I}}) \in \mathbb{R}^{Nn}$. Besides, let us define, as in Pavel (2020, Eq. 13, 14), for all $i \in \mathcal{I}$,

$$\mathcal{R}_i := \begin{bmatrix}\mathbf{0}_{n_i\times n_{<i}} & I_{n_i} & \mathbf{0}_{n_i\times n_{>i}}\end{bmatrix} \in \mathbb{R}^{n_i\times n}, \tag{4a}$$

$$\mathcal{S}_i := \begin{bmatrix} I_{n_{<i}} & \mathbf{0}_{n_{<i}\times n_i} & \mathbf{0}_{n_{<i}\times n_{>i}} \\ \mathbf{0}_{n_{>i}\times n_{<i}} & \mathbf{0}_{n_{>i}\times n_i} & I_{n_{>i}}\end{bmatrix} \in \mathbb{R}^{n_{-i}\times n} \tag{4b}$$

where $n_{<i} := \sum_{j<i, j\in\mathcal{I}} n_j$, $n_{>i} := \sum_{j>i, j\in\mathcal{I}} n_j$ and $n_{-i} := n - n_i$. In simple terms, $\mathcal{R}_i$ selects the i-th n_i-dimensional component from an n-dimensional vector, while $\mathcal{S}_i$ removes it. Thus, $\mathcal{R}_i\boldsymbol{x} = \boldsymbol{x}_{i,i} = x_i$ and $\mathcal{S}_i\boldsymbol{x} = \boldsymbol{x}_{i,-i}$. Let $\mathcal{R} := \operatorname{diag}((\mathcal{R}_i)_{i\in\mathcal{I}})$, $\mathcal{S} := \operatorname{diag}((\mathcal{S}_i)_{i\in\mathcal{I}})$. It follows that $x = \mathcal{R}\boldsymbol{x}$ and $\operatorname{col}((\boldsymbol{x}_{i,-i})_{i\in\mathcal{I}}) = \mathcal{S}\boldsymbol{x} \in \mathbb{R}^{(N-1)n}$. Moreover, $\boldsymbol{x} = \mathcal{R}^\top x + \mathcal{S}^\top\mathcal{S}\boldsymbol{x}$. We define the *extended pseudo-gradient* mapping $\boldsymbol{F} : \mathbb{R}^{Nn} \to \mathbb{R}^n$ as

$$\boldsymbol{F}(\boldsymbol{x}) := \operatorname{col}\Big(\big(\nabla_{x_i} J_i(x_i, \boldsymbol{x}_{i,-i})\big)_{i\in\mathcal{I}}\Big), \tag{5}$$

and the operators

$$\boldsymbol{F}_{\mathrm{a}}(\boldsymbol{x}) := \alpha\mathcal{R}^\top\boldsymbol{F}(\boldsymbol{x}) + (\boldsymbol{D}_n - \boldsymbol{W}_n)\boldsymbol{x}, \tag{6}$$

$$\mathcal{A}(\boldsymbol{\omega}) := \underbrace{\begin{bmatrix}\boldsymbol{F}_{\mathrm{a}}(\boldsymbol{x}) \\ \mathbf{0}_{Em} \\ \boldsymbol{b}\end{bmatrix} + \begin{bmatrix}\mathcal{R}^\top\boldsymbol{A}^\top\boldsymbol{\lambda} \\ -\boldsymbol{V}_m\boldsymbol{\lambda} \\ \boldsymbol{V}_m^\top\boldsymbol{v} - \boldsymbol{A}\mathcal{R}\boldsymbol{x}\end{bmatrix}}_{:=\mathcal{A}_1(\boldsymbol{\omega})} + \begin{bmatrix}\mathrm{N}_{\Omega}(\boldsymbol{x}) \\ \mathbf{0}_{Em} \\ \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\boldsymbol{\lambda})\end{bmatrix} \tag{7}$$

where $\alpha > 0$ is a design constant, $\boldsymbol{\omega} := \operatorname{col}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\lambda})$, $\boldsymbol{v} := \operatorname{col}((v_\ell)_{\ell\in\{1,\dots,E\}}) \in \mathbb{R}^{Em}$, $\boldsymbol{\lambda} := \operatorname{col}((\lambda_i)_{i\in\mathcal{I}}) \in \mathbb{R}^{Nm}$, $\boldsymbol{A} := \operatorname{diag}((A_i)_{i\in\mathcal{I}})$, $\boldsymbol{W}_n := W \otimes I_n$, $\boldsymbol{D}_n := D \otimes I_n$, $\boldsymbol{V}_m := V \otimes I_m$, and $\boldsymbol{\Omega} := \{\boldsymbol{x} \in \mathbb{R}^{nN} \mid \mathcal{R}\boldsymbol{x} \in \Omega\}$.

The following lemma relates the unique v-GNE of the game in (1) to the zeros of the operator $\mathcal{A}$. The proof is analogous to Pavel (2020, Th. 1) or Lemma 10 in Appendix B, and hence it is omitted.

Lemma 1. *Let $\mathcal{A}$ be as in (7). It holds that $\operatorname{zer}(\mathcal{A}) \neq \varnothing$. Moreover, let $\boldsymbol{x}^* \in \mathbb{R}^{Nn}$, $\boldsymbol{\lambda}^* \in \mathbb{R}^{Nm}$; then, the following statements are equivalent:*

(i) There exists $\boldsymbol{v}^*$ such that $\operatorname{col}(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \operatorname{zer}(\mathcal{A})$.

(ii) $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$ and $\boldsymbol{\lambda}^* = \mathbf{1}_N \otimes \lambda^*$, where the pair (x^*, λ^*) satisfies the Karush–Kuhn–Tucker (KKT) conditions in (3), hence x^* is the v-GNE of the game in (1). □

Effectively, Lemma 1 provides an extension of the KKT conditions in (3) and allows us to recast the GNE problem as that of computing a zero of the operator $\mathcal{A}$, for which a number of iterative algorithms are available (Bauschke & Combettes, 2017, §26-28). In fact, in Section 4.3, we show that Algorithm 1 can be recast as a PPA (Bauschke & Combettes, 2017, Th. 23.41).

Nonetheless, technical difficulties arise in the analysis because of the partial-decision information setup. Specifically, in (5), each partial gradient $\nabla_{x_i} J_i(x_i, \boldsymbol{x}_{i,-i})$ is evaluated on the local estimate $\boldsymbol{x}_{i,-i}$, and not on the actual value x_{-i}. Only when the estimates $\boldsymbol{x}$ are at consensus, i.e., $\boldsymbol{x} = \mathbf{1}_N \otimes x$ (namely, the estimate of each agents coincide with the actual value of x), we have that $\boldsymbol{F}(\boldsymbol{x}) = F(x)$. As a result, the operator $\mathcal{R}^\top\boldsymbol{F}$ (and consequently the operator $\mathcal{A}$) is not monotone in general[1], not even under the strong monotonicity of the game mapping F in Standing Assumption 2. Instead, analogously to the approaches in Gadjov and Pavel (2019), Pavel (2020) and Salehisadaghiani et al. (2019), our analysis is based on a *restricted* monotonicity property.

Definition 2. An operator $\mathcal{F} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ is restricted (μ-strongly) monotone in $\mathcal{H}_P$ if $\operatorname{zer}(\mathcal{F}) \neq \varnothing$ and $\langle \omega - \omega^* \mid u\rangle_P \geq 0$ ($\geq \mu\|\omega - \omega^*\|_P^2$) for all $(\omega, u) \in \operatorname{gra}(\mathcal{F})$, $\omega^* \in \operatorname{zer}(\mathcal{F})$ (we omit the characterization "in $\mathcal{H}_P$" whenever $P = I$). □

Definition 2 differs from that in Pavel (2020, Lem. 3), as we only consider properties with respect to the zero set and we need to include set-valued operators. The definition comprises the nonemptiness of the zero set and it does not exclude an operator that is multi-valued on its zeros. The next lemmas show that restricted monotonicity of $\mathcal{A}$ can be guaranteed for any game satisfying Standing Assumptions 1–3, without additional hypotheses.

Lemma 2 (*Bianchi & Grammatico, 2020, Lemma 3*). *The mapping $\boldsymbol{F}$ in (5) is θ-Lipschitz continuous, for some $\theta \in [\mu, \theta_0]$.* □

Lemma 3. *Let* $\alpha_{\max} := \frac{4\mu\lambda_2(L)}{(\theta_0+\theta)^2 + 4\mu\theta}$,

$$M := \alpha\begin{bmatrix}\frac{\mu}{N} & -\frac{\theta_0+\theta}{2\sqrt{N}} \\ -\frac{\theta_0+\theta}{2\sqrt{N}} & \frac{\lambda_2(L)}{\alpha} - \theta\end{bmatrix}, \qquad \mu_{\boldsymbol{F}_a} := \lambda_{\min}(M). \tag{8}$$

[1] It can be shown that $\mathcal{R}^\top\boldsymbol{F}$ is monotone only if the mappings $\nabla_{x_i} J_i(x)$'s do not depend on x_{-i} (in which case, there is no need for a partial-decision information assumption).

If $\alpha \in (0, \alpha_{\max}]$, then $\mu_{F_a} \geq 0$ and the operator $\mathcal{A}$ in (7) is restricted monotone. □

Proof. The operator $\mathcal{A}$ in (7) is the sum of three operators. The third is monotone by properties of normal cones (Bauschke & Combettes, 2017, Th. 20.25); the second is a linear skew-symmetric operator, hence monotone (Bauschke & Combettes, 2017, Ex. 20.35). Let $\boldsymbol{\omega}^* = \operatorname{col}(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \operatorname{zer}(\mathcal{A})$, where $\operatorname{zer}(\mathcal{A}) \neq \varnothing$ by Lemma 1. By Lemma 1, $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$, with x^* the v-GNE of the game in (1); hence by Pavel (2020, Lemma 3), for any $\alpha \in (0, \alpha_{\max}]$, it holds that $M \succcurlyeq 0$ and that, for all $\boldsymbol{x} \in \mathbb{R}^{Nn}$

$$\langle \boldsymbol{x} - \boldsymbol{x}^* \mid \boldsymbol{F}_a(\boldsymbol{x}) - \boldsymbol{F}_a(\boldsymbol{x}^*) \rangle \geq \mu_{F_a} \|\boldsymbol{x} - \boldsymbol{x}^*\|^2. \tag{9}$$

Therefore, for all $(\boldsymbol{\omega}, \boldsymbol{u}) \in \operatorname{gra}(\mathcal{A})$, with $\boldsymbol{\omega} = \operatorname{col}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{\lambda})$, it holds that $\langle \boldsymbol{\omega} - \boldsymbol{\omega}^* \mid \boldsymbol{u} - \mathbf{0} \rangle \geq \mu_{F_a} \|\boldsymbol{x} - \boldsymbol{x}^*\|^2 \geq 0$. ■

4.2. PPA for restricted monotone operators

In the remainder of this section, we show that Algorithm 1 is an instance of the PPA, applied to seek a zero of the (suitably preconditioned) operator $\mathcal{A}$ in (7). Then, we show its convergence based on the restricted monotonicity result in Lemma 3.

Informally speaking, in proximal-point methods, a problem is decomposed into a sequence of regularized subproblems, which are possibly better conditioned and easier to solve. Let $\mathcal{B} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ be maximally monotone (Bauschke & Combettes, 2017, Def. 20.20) in a space $\mathcal{H}_P$, and $\mathrm{J}_{\mathcal{B}} = (\mathrm{Id} + \mathcal{B})^{-1}$ its resolvent. Then, $\operatorname{dom}(\mathrm{J}_{\mathcal{B}}) = \mathbb{R}^q$ and $\mathrm{J}_{\mathcal{B}}$ is single-valued; moreover, if $\operatorname{zer}(\mathcal{B}) \neq \varnothing$, then the sequence $(\omega^k)_{k\in\mathbb{N}}$ generated by the PPA,

$$(\forall k \in \mathbb{N}) \quad \omega^{k+1} = \mathrm{J}_{\mathcal{B}}(\omega^k), \quad \omega^0 \in \mathbb{R}^q, \tag{10}$$

converges to a point in $\operatorname{zer}(\mathcal{B}) = \operatorname{fix}(\mathrm{J}_{\mathcal{B}})$ (Bauschke & Combettes, 2017, Th. 23.41). Note that performing the update in (10) is equivalent to solving for ω^{k+1} the (regularized) inclusion

$$\mathbf{0} \in \mathcal{B}(\omega^{k+1}) + \omega^{k+1} - \omega^k. \tag{11}$$

Unfortunately, many operator-theoretic properties are not guaranteed if $\mathcal{B}$ is only restricted monotone. In fact, $\mathrm{J}_{\mathcal{B}}$ might not be defined everywhere or single-valued.

Example 1. Let $\mathcal{B} : \mathbb{R} \to \mathbb{R}$, with $\mathcal{B}(\omega) = 9 - 2\omega$ if $\omega \in [3, 4)$, $\mathcal{B}(\omega) = \omega$ otherwise. Then, $\operatorname{zer}(\mathcal{B}) = \{0\}$ and $\mathcal{B}$ is restricted strongly monotone. However, $\mathrm{J}_{\mathcal{B}}(\omega) = \{\frac{\omega}{2}, 9 - \omega\}$ if $\omega \in [5, 6)$ and $\mathrm{J}_{\mathcal{B}}(\omega) = \varnothing$ if $\omega \in (6, 8)$. □

Nonetheless, some important properties carry on to the restricted monotone case, as we prove next.

Lemma 4. *Let $\mathcal{B} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ be restricted monotone in $\mathcal{H}_P$. Then, $\mathrm{J}_{\mathcal{B}}$ is firmly quasinonexpansive in $\mathcal{H}_P$: for any $(\omega, u) \in \operatorname{gra}(\mathrm{J}_{\mathcal{B}})$, $\omega^* \in \operatorname{zer}(\mathcal{B}) = \operatorname{fix}(\mathrm{J}_{\mathcal{B}})$, it holds that*

$$\langle \omega - u \mid \omega - \omega^* \rangle_P - \|u - \omega\|_P^2 = \langle \omega - u \mid u - \omega^* \rangle_P \geq 0. \tag{12}$$

Moreover, $\mathrm{J}_{\mathcal{B}}(\omega^*) = \{\omega^*\}$. □

Proof. By definition of resolvent, $\omega^* \in \mathrm{J}_{\mathcal{B}}(\omega^*) \Leftrightarrow \omega^* + \mathcal{B}\omega^* \ni \omega^* \Leftrightarrow \mathbf{0} \in \mathcal{B}(\omega^*)$; also, for any $(\omega, u) \in \operatorname{gra}(\mathrm{J}_{\mathcal{B}})$, $\omega - u \in \mathcal{B}(u)$. Hence, the inequality in (12) is the restricted monotonicity of $\mathcal{B}$; the elementary equality follows by expanding the terms. Finally, by taking $\omega = \omega^*$ in (12), we infer that $\mathrm{J}_{\mathcal{B}}$ is single-valued on $\operatorname{fix}(\mathrm{J}_{\mathcal{B}})$. ■

Next, by leveraging Lemma 4, we extend classical results for the PPA (Combettes, 2001, Th. 5.6) to the case of a restricted monotone operator (possibly with multi-valued resolvent).

Theorem 1. *Let $\mathcal{B} : \mathbb{R}^q \rightrightarrows \mathbb{R}^q$ be restricted monotone in $\mathcal{H}_P$, and $C := \operatorname{zer}(\mathcal{B}) \neq \varnothing$. Let $(\gamma^k)_{k\in\mathbb{N}}$ be a sequence in $[0, 2]$, and $(e^k)_{k\in\mathbb{N}}$ a sequence in $\mathbb{R}^q$ such that $(\gamma^k \|e^k\|_P)_{k\in\mathbb{N}} \in \ell^1$. Let $\omega^0 \in \mathbb{R}^q$ and let $(\omega^k)_{k\in\mathbb{N}}$ be any sequence such that:*

$$(\forall k \in \mathbb{N}) \quad \omega^{k+1} = \omega^k + \gamma^k(u^k - \omega^k + e^k),\ u^k \in \mathrm{J}_{\mathcal{B}}(\omega^k). \tag{13}$$

Then, the following statements hold:

(i) $(\forall \omega^* \in C)(\forall k \in \mathbb{N}) \quad \|\omega^{k+1} - \omega^*\|_P \leq \|\omega^k - \omega^*\|_P + \gamma^k \|e^k\|_P$.

(ii) $\left(\gamma^k(2 - \gamma^k)\|u^k - \omega^k\|_P^2\right)_{k\in\mathbb{N}} \in \ell^1$.

(iii) Assume that every cluster point of $(\omega^k)_{k\in\mathbb{N}}$ belongs to C. Then, $(\omega^k)_{k\in\mathbb{N}}$ converges to a point in C.

(iv) Assume that $\mathcal{B}$ is $\mu_{\mathcal{B}}$-strongly restricted monotone in $\mathcal{H}_P$. Then, $C = \{\omega^*\}$ and $\|\omega^{k+1} - \omega^*\|_P \leq \rho^k \|\omega^k - \omega^*\|_P + \gamma^k \|e^k\|_P$ for all $k \in \mathbb{N}$, where $\rho^k = \max(1 - \frac{\gamma^k \mu_{\mathcal{B}}}{1+\mu_{\mathcal{B}}}, \gamma^k - 1)$. □

Proof. See Appendix A. ■

Remark 3. The condition $\operatorname{dom}(\mathrm{J}_{\mathcal{B}}) = \mathbb{R}^q$ is sufficient (but not necessary) for the existence of a sequence $(\omega^k)_{k\in\mathbb{N}}$ that satisfies (13), which can be constructed choosing arbitrarily $u_k \in \mathrm{J}_{\mathcal{B}}(\omega^k)$, for all $k \in \mathbb{N}$. □

Example 2. Consider the $\mathrm{VI}(\Psi, S)$, where $S \subset \mathbb{R}^q$ is compact and convex, and $\Psi : \mathbb{R}^q \to \mathbb{R}^q$ is continuous and pseudomonotone in the sense of Karamardian (i.e., for all $\omega, \omega' \in \mathbb{R}^q$, the implication $\langle \Psi(\omega), \omega' - \omega \rangle \geq 0 \Rightarrow \langle \Psi(\omega'), \omega' - \omega \rangle \geq 0$ holds). It holds that $\mathrm{SOL}(\Psi, S) = \operatorname{zer}(\mathcal{B}) \neq \varnothing$, where $\mathcal{B} = \Psi + \mathrm{N}_S$ (Facchinei & Pang, 2007, Prop. 2.2.3). Moreover $\mathcal{B}$ is restricted monotone. To show this, consider any $\omega^* \in \operatorname{zer}(\mathcal{B})$ and $(\omega, u) \in \operatorname{gra}(\mathcal{B})$, so $u = \Psi(\omega) + u'$, for some u' such that $(\omega, u') \in \operatorname{gra}(\mathrm{N}_S)$. Then, $\langle u \mid \omega - \omega^* \rangle = \langle \Psi(\omega) \mid \omega - \omega^* \rangle + \langle u' - \mathbf{0} \mid \omega - \omega^* \rangle \geq 0$, where we used that $\langle \Psi(\omega) \mid \omega - \omega^* \rangle \geq 0$, by pseudomonotonicity and because $\langle \Psi(\omega^*) \mid \omega - \omega^* \rangle \geq 0$ by definition of VI, and $\langle u' - \mathbf{0} \mid \omega - \omega^* \rangle \geq 0$ because $(\omega^*, \mathbf{0}) \in \operatorname{gra}(\mathrm{N}_S)$ and monotonicity of the normal cone.

We note that $\operatorname{dom}(\mathrm{J}_{\mathcal{B}}) = \mathbb{R}^q$ by Facchinei and Pang (2007, Prop. 2.2.3). Let us consider any sequence $(\omega^k)_{k\in\mathbb{N}}$ such that, for all $k \in \mathbb{N}$, $\omega^{k+1} = u^k$, $u^k \in \mathrm{J}_{\mathcal{B}}(\omega^k)$, (or equivalently (11) or $\omega^{k+1} \in \mathrm{SOL}(\Psi + \mathrm{Id} - \omega^k, S)$). By Theorem 1 (with $\gamma^k = 1$, $e^k = \mathbf{0}$), $(\omega^k)_{k\in\mathbb{N}}$ is bounded, hence it admits at least one cluster point, say $\bar{\omega}$; by Theorem 1(ii) $\|u^k - \omega^k\| \to 0$. However, by definition of VI, for any $\omega \in S$, $\langle \Psi(u^k) + u^k - \omega^k \mid \omega - u^k \rangle \geq 0$. By passing to the limit (on a subsequence) and by continuity, we obtain $\langle \Psi(\bar{\omega}) \mid \omega - \bar{\omega} \rangle \geq 0$, which shows that $\bar{\omega} \in \mathrm{SOL}(\Psi, S)$. Therefore $(\omega^k)_{k\in\mathbb{N}}$ converges to a solution to $\mathrm{VI}(\Psi, S)$ by Theorem 1(iii). This extends the results in El Farouq (2001, §4.2), where hypomonotonicity of Ψ is assumed and where a small-enough step size is chosen to ensure that $\mathrm{J}_{\mathcal{B}}$ is single-valued (besides, pseudomonotonicity of Ψ is sufficient, but not necessary, for the restricted monotonicity of $\mathcal{B}$, and Theorem 1 would also allow to take into account iterations with errors, cf. El Farouq (2001, §4.2)). □

4.3. Derivation and convergence

Next, we show how that Algorithm 1 is obtained by applying the iteration in (13) to the operator $\Phi^{-1}\mathcal{A}$, where

$$\Phi := \begin{bmatrix} \bar{\tau}^{-1} + \boldsymbol{W}_n & \mathbf{0} & -\mathcal{R}^\top \boldsymbol{A}^\top \\ \mathbf{0} & \bar{\nu}^{-1} & \boldsymbol{V}_m \\ -\boldsymbol{A}\mathcal{R} & \boldsymbol{V}_m^\top & \bar{\delta}^{-1} \end{bmatrix} \tag{14}$$

is called *preconditioning* matrix. The step sizes $\bar{\tau} := \operatorname{diag}((\tau_i I_n)_{i\in\mathcal{I}})$, $\bar{\nu} := \operatorname{diag}((\nu_{(i,j)} I_m)_{(i,j)\in\mathcal{E}})$, $\bar{\delta} := \operatorname{diag}((\delta_i I_m)_{i\in\mathcal{I}})$, have to be chosen such that $\Phi \succ 0$. In this case, it also holds that $\operatorname{zer}(\Phi^{-1}\mathcal{A}) = \operatorname{zer}(\mathcal{A})$. Sufficient conditions that ensure $\Phi \succ 0$ are given in the next lemma, which follows by the Gershgorin's circle theorem.

Lemma 5. *The matrix Φ in (14) is positive definite if $\nu_{(i,j)}^{-1} > 2\sqrt{(w_{i,j})}$ for all $(i,j) \in \mathcal{E}$ and $\tau_i^{-1} > d_i + \|A_i^\top\|_\infty$, $\delta_i^{-1} > \|A_i\|_\infty + \sum_{j=1}^N \sqrt{(w_{i,j})}$ for all $i \in \mathcal{I}$.* □

In the following, we always assume that the step sizes in Algorithm 1 are chosen such that $\Phi \succ 0$. Then, we are able to formulate the following result.

Lemma 6. *Algorithm 1 is equivalent to the iteration*

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{\omega}^{k+1} \in \mathrm{J}_{\Phi^{-1}\mathcal{A}}(\boldsymbol{\omega}^k), \tag{15}$$

with $\mathcal{A}$ as in (7), Φ as in (14): for any initial condition $\boldsymbol{\omega}^0 = \mathrm{col}(\boldsymbol{x}^0, \boldsymbol{v}^0 = \mathbf{0}_{Em}, \boldsymbol{\lambda}^0)$, the sequence $(\boldsymbol{x}^k, \boldsymbol{V}_m^\top \boldsymbol{v}^k, \boldsymbol{\lambda}^k)_{k\in\mathbb{N}}$ generated by (15) coincides with the sequence $(\boldsymbol{x}^k, \boldsymbol{z}^k, \boldsymbol{\lambda}^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 with initial conditions $(\boldsymbol{x}^0, \boldsymbol{z}^0 = \mathbf{0}_{Nm}, \boldsymbol{\lambda}^0)$. □

Proof. By definition of inverse operator, we have that

$$\begin{aligned}
&\boldsymbol{\omega}^{k+1} \in (\mathrm{Id} + \Phi^{-1}\mathcal{A})^{-1}(\boldsymbol{\omega}^k)\\
\Leftrightarrow\ &\mathbf{0} \in \Phi^{-1}\mathcal{A}(\boldsymbol{\omega}^{k+1}) - \boldsymbol{\omega}^k + \boldsymbol{\omega}^{k+1}\\
\Leftrightarrow\ &\mathbf{0} \in \Phi(\boldsymbol{\omega}^{k+1} - \boldsymbol{\omega}^k) + \mathcal{A}(\boldsymbol{\omega}^{k+1}) \qquad (16)\\
\Leftrightarrow\ &\begin{cases}
\mathbf{0} \in \bar{\tau}^{-1}(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) + \cancel{\boldsymbol{W}_{\bar{n}}\boldsymbol{x}^{k+1}} - \boldsymbol{W}_n\boldsymbol{x}^k + \boldsymbol{D}_n\boldsymbol{x}^{k+1}\\
\qquad - \cancel{\mathcal{R}^\top\boldsymbol{A}^\top\boldsymbol{\lambda}^{k+1}} + \mathcal{R}^\top\boldsymbol{A}^\top\boldsymbol{\lambda}^k + \alpha\mathcal{R}^\top\boldsymbol{F}(\boldsymbol{x}^{k+1})\\
\qquad - \cancel{\boldsymbol{W}_{\bar{n}}\boldsymbol{x}^{k+1}} + \mathcal{R}^\top\cancel{\boldsymbol{A}^\top\boldsymbol{\lambda}^{k+1}} + \mathrm{N}_\Omega(\boldsymbol{x}^{k+1})\\
\mathbf{0} \in \bar{\nu}^{-1}(\boldsymbol{v}^{k+1} - \boldsymbol{v}^k) + \cancel{\boldsymbol{V}_{\bar{m}}\boldsymbol{\lambda}^{k+1}} - \boldsymbol{V}_m\boldsymbol{\lambda}^k - \cancel{\boldsymbol{V}_{\bar{m}}\boldsymbol{\lambda}^{k+1}}\\
\mathbf{0} \in \bar{\delta}^{-1}(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \mathrm{N}_{\mathbb{R}^{mN}_{\geq 0}}(\boldsymbol{\lambda}^{k+1}) + \boldsymbol{b}\\
\qquad - \boldsymbol{A}\mathcal{R}(2\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) + \boldsymbol{V}_m^\top(2\boldsymbol{v}^{k+1} - \boldsymbol{v}^k)
\end{cases} \qquad (17)
\end{aligned}$$

In turn, the first inclusion in (17) can be split in two by left-multiplying both sides with $\mathcal{R}$ and $\mathcal{S}$. By $\mathcal{S}\mathrm{N}_\Omega = \mathbf{0}_{(N-1)n}$, $\mathcal{R}\mathcal{R}^\top = I_n$ and $\mathcal{S}\mathcal{R}^\top = \mathbf{0}_{(N-1)n\times n}$, we get

$$\begin{aligned}
&\begin{cases}
\mathbf{0} \in \mathcal{S}((I + \bar{\tau}\boldsymbol{D}_n)\boldsymbol{x}^{k+1} - \boldsymbol{x}^k - \bar{\tau}\boldsymbol{W}_n\boldsymbol{x}^k)\\
\mathbf{0} \in \mathcal{R}((I + \bar{\tau}\boldsymbol{D}_n)\boldsymbol{x}^{k+1} - \boldsymbol{x}^k - \bar{\tau}\boldsymbol{W}_n\boldsymbol{x}^k)\\
\qquad + \mathrm{N}_\Omega(x^{k+1}) + \alpha\bar{\tau}\boldsymbol{F}((x^{k+1}, \mathcal{S}\boldsymbol{x}^{k+1})) + \bar{\tau}\boldsymbol{A}^\top\boldsymbol{\lambda}^k
\end{cases}\\
\Leftrightarrow_{\forall i\in\mathcal{I}}\ &\begin{cases}
\boldsymbol{x}_{i,-i}^{k+1} = \frac{1}{1+\tau_i d_i}(\boldsymbol{x}_{i,-i}^k + \tau_i\sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,-i}^k)\\
\mathbf{0}_{n_i} \in \partial_{x_i^{k+1}}\Big(J_i(x_i^{k+1}, \boldsymbol{x}_{i,-i}^{k+1}) + \frac{1}{2\alpha\tau_i}\big\|x_i^{k+1} - x_i^k\big\|^2\\
\qquad + \frac{1}{2\alpha d_i}\big\|d_i x_i^{k+1} - \sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,i}^k\big\|^2\\
\qquad + \iota_{\Omega_i}(x_i^{k+1}) + \frac{1}{\alpha}(A_i^\top\lambda_i^k)^\top x_i^{k+1}\Big).
\end{cases}
\end{aligned}$$

Therefore, since the zeros of the subdifferential of a (strongly) convex function coincide with the minima (unique minimum) (Bauschke & Combettes, 2017, Th. 16.3), (17) can be rewritten as

$$\begin{aligned}
&\forall i \in \mathcal{I}: \begin{cases}
\boldsymbol{x}_{i,-i}^{k+1} = \frac{1}{1+\tau_i d_i}(\boldsymbol{x}_{i,-i}^k + \tau_i\sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,-i}^k)\\
x_i^{k+1} = \underset{y\in\Omega_i}{\mathrm{argmin}}\Big(J_i(y, \boldsymbol{x}_{i,-i}^{k+1}) + \frac{1}{2\alpha\tau_i}\big\|y - x_i^k\big\|^2\\
\qquad + \frac{1}{2\alpha d_i}\big\|d_i y - \sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,i}^k\big\|^2\\
\qquad + \frac{1}{\alpha}(A_i^\top\lambda_i^k)^\top y\Big)
\end{cases}\\
&\boldsymbol{v}^{k+1} = \boldsymbol{v}^k + \bar{\nu}\boldsymbol{V}_m\boldsymbol{\lambda}^k\\
&\boldsymbol{\lambda}^{k+1} = \mathrm{proj}_{\mathbb{R}^{mN}_{\geq 0}}\Big(\boldsymbol{\lambda}^k + \bar{\delta}\Big(\boldsymbol{A}\mathcal{R}(2\boldsymbol{x}^{k+1} - \boldsymbol{x}^k) - \boldsymbol{b}\\
&\qquad\qquad - \boldsymbol{V}_m^\top(2\boldsymbol{v}^{k+1} - \boldsymbol{v}^k)\Big)\Big). \qquad (18)
\end{aligned}$$

The conclusion follows by defining $\boldsymbol{z}^k := \boldsymbol{V}_m^\top\boldsymbol{v}^k$, where $\boldsymbol{z}^k = \mathrm{col}((z_i)_{i\in\mathcal{I}}) \in \mathbb{R}^{Nm}$ and $z_i^k \in \mathbb{R}^m$ is a local variables kept by agent i, provided that $\boldsymbol{z}^0 = \boldsymbol{V}_m^\top\boldsymbol{v}^0$. The latter is ensured by $\boldsymbol{z}^0 = \mathbf{0}_{Nm}$, as in Algorithm 1. ■

Remark 4. The preconditioning matrix Φ is designed to make the system in (17) block triangular, i.e., to remove the term $\boldsymbol{W}_n\boldsymbol{x}^{k+1}$ and $\mathcal{R}^\top\boldsymbol{A}^\top\boldsymbol{\lambda}^{k+1}$ from the first inclusion, and the terms $\boldsymbol{V}_m\boldsymbol{\lambda}^{k+1}$ from the second one: in this way, $\boldsymbol{x}_i^{k+1}$ and $\boldsymbol{z}^{k+1}$ do not depend on $\boldsymbol{x}_j^{k+1}$, for $i \neq j$, or on $\boldsymbol{\lambda}^{k+1}$. This ensures that the resulting iteration can be computed by the agents in a fully-distributed fashion (differently from the non-preconditioned resolvent $\mathrm{J}_{\mathcal{A}}$). Furthermore, the change of variable $\boldsymbol{z} = \boldsymbol{V}_m^\top\boldsymbol{v}$ reduces the number of auxiliary variables and decouples the dual update in (18) from the graph structure. □

Remark 5. By Lemma 6, Remark 1 and by the expression of $\mathrm{J}_{\Phi^{-1}\mathcal{A}}$ in (18), we conclude that $\mathrm{dom}(\mathrm{J}_{\Phi^{-1}\mathcal{A}}) = \mathbb{R}^{Nn+Em+Nm}$ and that $\mathrm{J}_{\Phi^{-1}\mathcal{A}}$ is single-valued. □

In order to apply Theorem 1 to the iteration in (15), we still need the following lemma.

Lemma 7. *Let $\alpha \in (0, \alpha_{\max}]$, $\alpha_{\max}$ as in Lemma 3. Then $\Phi^{-1}\mathcal{A}$ is restricted monotone in $\mathcal{H}_\Phi$.* □

Proof. Let $(\boldsymbol{\omega}, \boldsymbol{u}) \in \mathrm{gra}(\Phi^{-1}\mathcal{A})$, $\boldsymbol{\omega}^* \in \mathrm{zer}(\Phi^{-1}\mathcal{A})$. Then, $(\boldsymbol{\omega}, \Phi\boldsymbol{u}) \in \mathrm{gra}(\mathcal{A})$ and $\boldsymbol{\omega}^* \in \mathrm{zer}(\mathcal{A})$. Therefore, by Lemma 3, we conclude that $\langle \boldsymbol{u} \mid \boldsymbol{\omega} - \boldsymbol{\omega}^*\rangle_\Phi = \langle \Phi\boldsymbol{u} \mid \boldsymbol{\omega} - \boldsymbol{\omega}^*\rangle \geq 0$. ■

Theorem 2. *Let $\alpha \in (0, \alpha_{\max}]$, with $\alpha_{\max}$ as in Lemma 3, and let the step sizes $\bar{\tau}, \bar{\nu}, \bar{\delta}$ be as in Lemma 5. Then, the sequence $(\boldsymbol{x}^k, \boldsymbol{z}^k, \boldsymbol{\lambda}^k)_{k\in\mathbb{N}}$ generated by Algorithm 1 converges to some equilibrium $(\boldsymbol{x}^*, \boldsymbol{z}^*, \boldsymbol{\lambda}^*)$, where $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$ and x^* is the v-GNE of the game in (1).* □

Proof. By Lemma 6, we can equivalently study the convergence of the iteration in (15). In turn, (15) can be rewritten as (13) with $\gamma^k = 1$, $e^k = \mathbf{0}$, for all $k \in \mathbb{N}$. For later reference, let us define $\boldsymbol{u}^k = \mathrm{J}_{\Phi^{-1}\mathcal{A}}(\boldsymbol{\omega}^k)$ (here $\boldsymbol{u}^k = \boldsymbol{\omega}^{k+1}$). $\Phi^{-1}\mathcal{A}$ is restricted monotone in $\mathcal{H}_\Phi$ by Lemma 7. By Theorem 1(i), the sequence $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ is bounded, hence it admits at least one cluster point, say $\bar{\boldsymbol{\omega}}$. By (16) and (7), it holds, for any $\boldsymbol{\omega} \in \boldsymbol{\Omega} \times \mathbb{R}^{Em} \times \mathbb{R}^{Nm}_{\geq 0}$, that $\langle \mathcal{A}_1(\boldsymbol{u}^k) + \Phi(\boldsymbol{u}^k - \boldsymbol{\omega}^k) \mid \boldsymbol{\omega} - \boldsymbol{u}^k\rangle \geq 0$, with $\mathcal{A}_1$ as in (7). By Theorem 1(ii), $\boldsymbol{u}^k - \boldsymbol{\omega}^k \to \mathbf{0}$. Therefore, by continuity of $\mathcal{A}_1$, taking the limit on a diverging subsequence $(l_k)_{k\in\mathbb{N}}$ such that $(\boldsymbol{\omega}^{l_k})_{k\in\mathbb{N}} \to \bar{\boldsymbol{\omega}}$, we have that for all $\boldsymbol{\omega} \in \boldsymbol{\Omega} \times \mathbb{R}^{Em} \times \mathbb{R}^{Nm}_{\geq 0}$, $\langle \mathcal{A}_1(\bar{\boldsymbol{\omega}}) \mid \boldsymbol{\omega} - \bar{\boldsymbol{\omega}}\rangle \geq 0$, which shows that $\bar{\boldsymbol{\omega}} \in \mathrm{zer}(\mathcal{A}) = \mathrm{fix}(\mathrm{J}_{\Phi^{-1}\mathcal{A}})$. Hence $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ converges to an equilibrium of (15) by Theorem 1(iii). The conclusion follows by Lemma 1. ■

Remark 6. While the choice of step sizes in Lemma 5 is decentralized, computing the bound $\alpha_{\max}$ for the common parameter α in Algorithm 1 requires some global information on the graph $\mathcal{G}$ (i.e., the algebraic connectivity) and on the game mapping (the strong monotonicity and Lipschitz constants). □

Remark 7. If $\boldsymbol{x}^0 \in \Omega^N$, then $\boldsymbol{x}^k \in \Omega^N$ for all $k \in \mathbb{N}$ (by convexity and the updates in Algorithm 1), and Assumption 2 can be relaxed to hold only on Ω. □

Remark 8 (*Inexact Updates*)**.** The local optimization problems in Algorithm 1 are strongly convex, hence they can be efficiently solved by several iterative algorithms (with linear rate). While computing the exact solutions $\bar{x}_i^k$ would require an infinite number of iterations, the convergence in Theorem 2 still holds if x_i is updated with an approximation $\widehat{x}_i^k$ of $\bar{x}_i^k$, provided that the errors $e_i^k := \bar{x}_i^k - \widehat{x}_i^k$ are norm summable, i.e., $(\|e_i^k\|)_{k\in\mathbb{N}} \in \ell^1$, for all $i \in \mathcal{I}$ (the same proof applies, since the condition on e^k in Theorem 1 would be satisfied, by equivalence of norms). For example, assume that $\widehat{x}_i^k$ is computed via a finite number $j_i^k \geq 1$ of steps of the projected gradient method, warm-started at x_i^k, with

Algorithm 2 Fully-distributed v-GNE seeking via accelerated PPPA

Initialization:
- Choose acceleration:
  - Overrelaxation: set $\gamma > 0$, $\zeta = 0$, $\eta = 0$;
  - Inertia: set $\gamma = 0$, $\zeta > 0$, $\eta = 0$;
  - Alternated inertia: set $\gamma = 0$, $\zeta = 0$, $\eta > 0$;
- For all $i \in \mathcal{I}$, set $x_i^{-1} = x_i^0 \in \Omega_i$, $\boldsymbol{x}_{i,-i}^{-1} = \boldsymbol{x}_{i,-i}^0 \in \mathbb{R}^{n-n_i}$, $z_i^{-1} = z_i^0 = \mathbf{0}_m$, $\lambda_i^{-1} = \lambda_i^0 \in \mathbb{R}^m_{\geq 0}$.

For all $k > 0$:
- (Alternated) inertial step: set $\tilde{\eta}^k = 0$ if k is even, $\tilde{\eta}^k = \eta$ otherwise; each agent $i \in \mathcal{I}$ computes

$$\tilde{\boldsymbol{x}}_{i,-i}^k = \boldsymbol{x}_{i,-i}^k + (\zeta + \tilde{\eta}^k)(\boldsymbol{x}_{i,-i}^k - \boldsymbol{x}_{i,-i}^{k-1}) \qquad \tilde{x}_i^k = x_i^k + (\zeta + \tilde{\eta}^k)(x_i^k - x_i^{k-1})$$
$$\tilde{z}_i^k = z_i^k + (\zeta + \tilde{\eta}^k)(z_i^k - z_i^{k-1}) \qquad \tilde{\lambda}_i^k = \lambda_i^k + (\zeta + \tilde{\eta}^k)(\lambda_i^k - \lambda_i^{k-1})$$

- Communication: The agents exchange the variables $\{\tilde{x}_i^k, \tilde{\boldsymbol{x}}_{i,-i}^k, \tilde{\lambda}_i^k\}$ with their neighbors.
- Resolvent computation: each agent $i \in \mathcal{I}$ computes

$$\check{\boldsymbol{x}}_{i,-i}^{k+1} = \tfrac{1}{1+\tau_i d_i}\big(\tilde{\boldsymbol{x}}_{i,-i}^k + \tau_i \textstyle\sum_{j=1}^N w_{i,j}\tilde{\boldsymbol{x}}_{j,-i}^k\big)$$
$$\check{x}_i^{k+1} = \operatorname*{argmin}_{y \in \Omega_i}\Big(J_i(y, \check{\boldsymbol{x}}_{i,-i}^{k+1}) + \tfrac{1}{2\alpha\tau_i}\big\|y - \tilde{x}_i^k\big\|^2 + \tfrac{1}{2\alpha d_i}\big\|d_i y - \textstyle\sum_{j=1}^N w_{i,j}\tilde{\boldsymbol{x}}_{j,i}^k\big\|^2 + \tfrac{1}{\alpha}(A_i^\top \tilde{\lambda}_i^k)^\top y\Big)$$
$$\check{z}_i^{k+1} = \tilde{z}_i^k + \textstyle\sum_{j=1}^N \nu_{(i,j)} w_{i,j}(\tilde{\lambda}_i^k - \tilde{\lambda}_j^k)$$
$$\check{\lambda}_i^{k+1} = \mathrm{proj}_{\mathbb{R}^m_{\geq 0}}\big(\tilde{\lambda}_i^k + \delta_i\big(A_i(2\check{x}_i^{k+1} - \tilde{x}_i^k) - b_i - (2\check{z}_i^{k+1} - \tilde{z}_i^k)\big)\big).$$

- Relaxation step: each agent $i \in \mathcal{I}$ computes

$$\boldsymbol{x}_{i,-i}^{k+1} = \gamma\check{\boldsymbol{x}}_{i,-i}^{k+1} + (1-\gamma)\,\boldsymbol{x}_{i,-i}^k \qquad x_i^{k+1} = \gamma\check{x}_i^{k+1} + (1-\gamma)\,x_i^k$$
$$z_i^{k+1} = \gamma\check{z}_i^{k+1} + (1-\gamma)\,z_i^k \qquad \lambda_i^{k+1} = \gamma\check{\lambda}_i^{k+1} + (1-\gamma)\,\lambda_i^k$$

(small enough) fixed step. Then, each agent can independently ensure that $\|e_i^k\| \leq \varepsilon_i^k$, for some $(\varepsilon_i^k)_{k\in\mathbb{N}} \in \ell^1$, by simply choosing

$$j_i^k \geq \log\Big(\varepsilon_i^k(1-\rho_i)/\|x_i^k - \widehat{x}_i^{k,1}\|\Big)\,/\,\log(\rho_i), \tag{19}$$

where $\widehat{x}_i^{k,1}$ is the approximation obtained after one gradient step and $\rho_i \in (0,1)$ is the contractivity parameter of the gradient descent.[2] We finally remark that $\bar{x}_i^k$ must be estimated with increasing accuracy. In practice, however, when x_i^k is converging, $\|x_i^{k+1} - x_i^k\| \to 0$. Hence x_i^k is a good initial guess for $\bar{x}_i^k$, and the computation of x_i^{k+1} often requires few gradient steps, see also Section 7. □

5. Accelerations

Lemma 6 shows that Algorithm 1 can be recast (modulo the change of variables $\boldsymbol{z} = \boldsymbol{V}_m^\top \boldsymbol{v}$) as

$$\boldsymbol{\omega}^{k+1} = T(\boldsymbol{\omega}^k), \tag{20}$$

where $T := \mathrm{J}_{\Phi^{-1}\mathcal{A}}$. This compact operator representation allows for some modifications of Algorithm 1, that can increase its convergence speed. In particular, we consider three popular accelerations schemes (Iutzeler & Hendrickx, 2019), which have been extensively studied for the case of firmly nonexpansive operators (Bauschke & Combettes, 2017, Def. 4.1), and also found application in games under full-decision information (Belgioioso & Grammatico, 2020; Scutari et al., 2014). Here we provide convergence guarantees for the partial-decision information setup, where T is only firmly quasinonexpansive. Our fully distributed accelerated algorithms are illustrated in Algorithm 2. In the following, we assume that $\alpha \in (0, \alpha_{\max}]$, $\alpha_{\max}$ as in Lemma 3, and that the step sizes $\bar{\tau}$, $\bar{\nu}$, $\bar{\delta}$ are chosen as in Lemma 5.

[2] ρ_i can be taken independent of k: since $\nabla J_i(\cdot, \boldsymbol{x}_{i,-i})$ is μ_i strongly monotone and θ_i Lipschitz, for some $\mu_i \geq \mu$, $\theta_i \leq \theta$ and for all $\boldsymbol{x}_{i,-i}$, the factor $\rho_i = \frac{\theta_i - \mu_i}{\theta_i + \mu_i + 1/(\alpha\tau_i) + d_i/\alpha}$ is ensured by the step $2/(\theta_i + \mu_i + 1/(\alpha\tau_i) + d_i/\alpha)$.

Proposition 1 (*Overrelaxation*). *Let $\gamma \in [1, 2)$. Then, for any $\boldsymbol{\omega}^0$, the sequence $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ generated by*

$$(\forall k \in \mathbb{N}), \quad \boldsymbol{\omega}^{k+1} = \boldsymbol{\omega}^k + \gamma(T(\boldsymbol{\omega}^k) - \boldsymbol{\omega}^k), \tag{21}$$

converges to an equilibrium $(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\mathcal{A})$, where $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$ and x^* is the v-GNE of the game in (1). □

Proof. The iteration in (21) is in the form (13), with $\gamma^k = \gamma$, $e^k = \mathbf{0}$, for all $k \in \mathbb{N}$. Then, the conclusion follows analogously to Theorem 2. ■

Proposition 2 (*Inertia*). *Let $\zeta \in [0, \frac{1}{3})$. Then, for any $\boldsymbol{\omega}^{-1} := \boldsymbol{\omega}^0$, the sequence $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ generated by*

$$(\forall k \in \mathbb{N}), \quad \boldsymbol{\omega}^{k+1} = T(\boldsymbol{\omega}^k + \zeta(\boldsymbol{\omega}^k - \boldsymbol{\omega}^{k-1})), \tag{22}$$

converges to an equilibrium $(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\mathcal{A})$, where $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$ and x^* is the v-GNE of the game in (1). □

Proof (sketch). By following all the steps in the proof of Boţ, Csetnek, and Hendrich (2015, Th. 5) (which can be done by recalling that an operator T is firmly (quasi)nonexpansive if and only if the operator $2T - \mathrm{Id}$ is (quasi)nonexpansive (Bauschke & Combettes, 2017, Prop. 4.2, 4.4), it can be shown that, if $\zeta \in [0, \frac{1}{3})$, then $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ is bounded and $\boldsymbol{\omega}^{k+1} - \boldsymbol{\omega}^k \to 0$. Then, the proof follows analogously to Theorem 2. ■

Proposition 3 (*Alternated Inertia*). *Let $\eta \in [0, 1]$. Then, for any $\boldsymbol{\omega}^0$, the sequence $(\boldsymbol{\omega}^k)_{k\in\mathbb{N}}$ generated by*

$$\begin{cases} \boldsymbol{\omega}^{k+1} = T(\boldsymbol{\omega}^k) & \text{if } k \text{ is even,} \\ \boldsymbol{\omega}^{k+1} = T(\boldsymbol{\omega}^k + \eta(\boldsymbol{\omega}^k - \boldsymbol{\omega}^{k-1})) & \text{if } k \text{ is odd,} \end{cases} \tag{23}$$

converges to an equilibrium $(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\mathcal{A})$, where $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$ and x^* is the v-GNE of the game in (1). □

Proof. For all $k \in \mathbb{N}$, $\boldsymbol{\omega}^{2k+2} = T(T(\boldsymbol{\omega}^{2k}) + \eta(T(\boldsymbol{\omega}^{2k}) - \boldsymbol{\omega}^{2k}))$, which is the same two-steps update obtained in (13) with $\gamma^{2k} = 1 + \eta$,

$\gamma^{2k+1} = 1$ (and $\mathcal{B} := \Phi^{-1}\mathcal{A}$, $e^k = \mathbf{0}$). Therefore the convergence of the sequence $(\boldsymbol{\omega}^{2k})_{k\in\mathbb{N}}$ to an equilibrium $(\boldsymbol{x}^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \operatorname{zer}(\mathcal{A})$ follows analogously to Theorem 2 (with a minor modification for the case $\eta = 1$). The convergence of the sequence $(\boldsymbol{\omega}^{2k+1})_{k\in\mathbb{N}}$ then follows by Theorem 1(i). ■

We note that, by Theorem 1, the convergence results in Propositions 1 and 3 hold also in the case of summable errors on the updates, as in Remark 8. Analogously to our analysis, provably convergent acceleration schemes could also be obtained for the FB algorithm in Pavel (2020): however, an advantage of our PPA is that the bounds on the inertial/relaxation parameters are fixed and independent on (unknown) problem parameters.

5.1. On the convergence rate

We conclude this section with a discussion on the convergence rate of Algorithms 1 and 2. First, even under Standing Assumption 2, the KKT operator on the right-hand side of (3) is generally not strongly monotone. Similarly, the operator $\mathcal{A}$ in (7) is not strongly monotone and Algorithm 1 can have multiple fixed points. Therefore, one should not expect linear convergence. By Lemma 6 and the proof of Theorem 1, we can derive the following ergodic rate for the fixed-point residual in Algorithm 1:

$$\tfrac{1}{k}\textstyle\sum_{i=0}^{k} \|\boldsymbol{\omega}^{k+1} - \boldsymbol{\omega}^k\|^2 \le O(1/k).$$

This rate also holds for the iterations in (21), (22), (23); for the case of general operator splittings (and differently from optimization algorithms), tighter rates for accelerated schemes are only known for particular cases, and most works focus on mere convergence (Boţ et al., 2015; Iutzeler & Hendrickx, 2019). Yet, the practice shows that relaxation and inertia often result in improved speed, see Belgioioso and Grammatico (2020) or Section 7.

The same residual rate $O(1/k)$ can also be shown for the pseudo-gradient method in Pavel (2020, Alg. 1). However, a major difference from Lemma 5 is that the upper bounds for the step sizes in Pavel (2020, Th. 2) are proportional to the constant $\mu_{\boldsymbol{F}_a}$ in (8), which is typically very small (up to scaling of the whole operator $\boldsymbol{F}_a$), see Bianchi and Grammatico (2021) or also Section 7.1. Most importantly, $\mu_{\boldsymbol{F}_a}$ vanishes as the number of agents increases (fixed the other parameters). In contrast, our algorithms allows for much larger steps, which can be chosen independently of the number of agents. This is a structural advantage of the PPA, whose convergence does not depend on the cocoercivity constants of the operators involved. Indeed, step sizes must be taken into account if convergence is evaluated in terms of residuals.

We finally note that linear convergence can be achieved via PPPA for games without coupling constraints. For instance, Algorithm 3 corresponds to the overrelaxed method in Algorithm 2, and can be derived, as in Lemma 6, by taking $\mathcal{B} = \Phi_{\text{NE}}^{-1}\mathcal{A}_{\text{NE}}(\boldsymbol{x})$ in (13), where $\mathcal{A}_{\text{NE}}(\boldsymbol{x}) := \boldsymbol{F}_a(\boldsymbol{x}) + \mathrm{N}_{\Omega}(\boldsymbol{x})$ and $\Phi_{\text{NE}} := \bar{\tau}^{-1} + \boldsymbol{W}_n$ are obtained by removing the dual variables from $\mathcal{A}$, Φ. By (9), as in Lemma 7, it can be shown that $\mathcal{A}_{\text{NE}}$ is restricted $\frac{\mu_{\boldsymbol{F}_a}}{\|\Phi_{\text{NE}}\|}$-strongly monotone in $\mathcal{H}_{\Phi_{\text{NE}}}$. Thus, recursively applying Theorem 1(iv), we can infer the following result, which appeared in Bianchi et al. (2020) only limited to $\gamma = 1$.

Theorem 3. *Let $\tau_i^{-1} > d_i$ for all $i \in \mathcal{I}$, let $\gamma \in (0, 2)$, and let $\alpha \in (0, \alpha_{\max}]$, with $\alpha_{\max}$ as in Lemma 3. Then, the sequence $(\boldsymbol{x}^k)_{k\in\mathbb{N}}$ generated by Algorithm 3 converges to $\boldsymbol{x}^* = \mathbf{1}_N \otimes x^*$, where x^* is the unique Nash equilibrium of the game in (1), with linear rate:*

$$(\forall k \in \mathbb{N}) \quad \|\boldsymbol{x}^k - \boldsymbol{x}^*\|_{\Phi_{\text{NE}}} \le (\rho_\gamma)^k \|\boldsymbol{x}^0 - \boldsymbol{x}^*\|_{\Phi_{\text{NE}}},$$

where $\rho_\gamma := \max(1 - \frac{\gamma\mu_{\boldsymbol{F}_a}}{\|\Phi_{\text{NE}}\|+\mu_{\boldsymbol{F}_a}}, \gamma - 1)$, $\mu_{\boldsymbol{F}_a}$ as in (8). □

Algorithm 3 Fully-distributed NE seeking via PPPA

$$\check{\boldsymbol{x}}_{i,-i}^{k+1} = \tfrac{1}{1+\tau_i d_i}\left(\boldsymbol{x}_{i,-i}^k + \tau_i \textstyle\sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,-i}^k\right)$$

$$\check{x}_i^{k+1} = \operatorname*{argmin}_{y\in\Omega_i} \Big(J_i(y, \check{\boldsymbol{x}}_{i,-i}^{k+1}) + \tfrac{1}{2\alpha\tau_i}\|y - x_i^k\|^2 + \tfrac{1}{2\alpha d_i}\|d_i y - \textstyle\sum_{j=1}^N w_{i,j}\boldsymbol{x}_{j,i}^k\|^2\Big)$$

$$\boldsymbol{x}_i^{k+1} = \boldsymbol{x}_i^k + \gamma(\check{\boldsymbol{x}}_i^{k+1} - \boldsymbol{x}_i^k)$$

Table 1
Comparison between our PPPA and projected pseudo-gradient methods.

	FB (Pavel, 2020, Alg. 1)	PPPA
Step sizes	$O\left(\frac{\mu_{\boldsymbol{F}_a}}{\theta_{\boldsymbol{F}_a}{}^2 + \mu_{\boldsymbol{F}_a}}\right)$	$O(1)$
Linear rate ρ (no coupling constraints)	$(1 - \kappa_{\boldsymbol{F}_a}{}^2)^{\frac{1}{2}}$	$1 - \kappa_{\boldsymbol{F}_a}$

The best theoretical rate $\rho_{\bar{\gamma}} = 1 - 2\mu_{\boldsymbol{F}_a}/(\|\Phi_{\text{NE}}\| + 2\mu_{\boldsymbol{F}_a})$ is obtained for $\bar{\gamma} = 1 + \|\Phi_{\text{NE}}\|/(\|\Phi_{\text{NE}}\| + 2\mu_{\boldsymbol{F}_a})$. We observed in Bianchi et al. (2020, §5) that this rate compares favorably with that of the state-of-the-art algorithms — please refer to Bianchi et al. (2020), also for numerical results. For instance, in the absence of coupling constraints, the FB algorithm in Pavel (2020, Alg. 1) reduces to Tatarenko et al. (2020, Alg. 1), whose optimal linear rate $O((1 - \kappa_{\boldsymbol{F}_a}{}^2)^{k/2})$ depends *quadratically* on the quantity $\kappa_{\boldsymbol{F}_a} := \mu_{\boldsymbol{F}_a}/\theta_{\boldsymbol{F}_a} < 1$ (Tatarenko et al., 2020, Th. 7), where $\theta_{\boldsymbol{F}_a} := 2\max((d_i)_{i\in\mathcal{I}}) + \alpha\theta$. Instead, $\rho_{\bar{\gamma}} \le 1 - \kappa_{\boldsymbol{F}_a}$, for large enough τ_i's (since $\|\Phi_{\text{NE}}\| + 2\mu_{\boldsymbol{F}_a} \le \max((d_i + \tau_i^{-1})_{i\in\mathcal{I}}) + 2\alpha\theta$), as shown in Table 1.

6. Aggregative games

In this section we focus on the particularly relevant class of (average) aggregative games, which arises in a variety of engineering applications, e.g., network congestion control and demand-side management (Grammatico, 2017). In aggregative games, $n_i = \bar{n} > 0$ for all $i \in \mathcal{I}$ (hence $n = N\bar{n}$) and the cost function of each agent depends only on its local decision and on the value of the average strategy $\operatorname{avg}(x) := \frac{1}{N}\sum_{i\in\mathcal{I}} x_i$. Therefore, for each $i \in \mathcal{I}$, there is a function $f_i : \mathbb{R}^{\bar{n}} \times \mathbb{R}^{\bar{n}} \to \mathbb{R}$ such that the original cost function J_i in (1) can be written as

$$J_i(x_i, x_{-i}) =: f_i(x_i, \operatorname{avg}(x)). \tag{25}$$

Since an aggregative game is only a particular instance of the game in (1), all the considerations on the existence and uniqueness of a v-GNE and the equivalence with the KKT conditions in (3) are still valid.

Moreover, Algorithms 1 could still be used to compute a v-GNE. This would require each agent to keep (and exchange) an estimate of all other agents' action, i.e., a vector of $(N-1)\bar{n}$ components. In practice, however, the cost of each agent is only a function of the aggregative value $\operatorname{avg}(x)$, whose dimension $\bar{n}$ is independent of the number N of agents. To reduce communication and computation burden, in this section we introduce a PPPA specifically tailored to seek a v-GNE in aggregative games, that is scalable with the number of agents. The proposed iteration is shown in Algorithm 4, where the parameters α, β, and τ_i, δ_i for all $i \in \mathcal{I}$, $v_{(i,j)}$ for all $(i,j) \in \mathcal{E}$ have to be chosen appropriately, and we denote

$$\tilde{\boldsymbol{F}}_i(x_i, \xi_i) := \nabla_{x_i} f_i(x_i, \xi_i) + \tfrac{1}{N}\nabla_{\xi_i} f_i(x_i, \xi_i). \tag{26}$$

We note that $\tilde{\boldsymbol{F}}_i(x_i, \operatorname{avg}(x)) = \nabla_{x_i} J_i(x_i, x_{-i}) = \nabla_{x_i} f_i(x_i, \operatorname{avg}(x))$.

Algorithm 4 Fully-distributed v-GNE seeking in aggregative games via PPPA

Initialization: · For all $i \in \mathcal{I}$, set $x_i^0 \in \Omega_i$, $s_i^0 = \mathbf{0}_{\bar{n}}$, $z_i^0 = \mathbf{0}_m$, $\lambda_i^0 \in \mathbb{R}^m_{\geq 0}$.

For all $k > 0$: · Communication: The agents exchange the variables $\{\sigma_i^k = x_i^k + s_i^k, \lambda_i^k\}$ with their neighbors.

· Local variables update: each agent $i \in \mathcal{I}$ computes

$$
\begin{aligned}
s_i^{k+1} &= s_i^k - \beta \textstyle\sum_{j=1}^N w_{i,j}(\sigma_i^k - \sigma_j^k) \\
x_i^{k+1} &\leftarrow y \ \text{ s.t. } \ \mathbf{0}_{\bar{n}} \in \alpha \tilde{\boldsymbol{F}}_i(y, y + s_i^{k+1}) + \tfrac{1}{\tau_i}(y - x_i^k) + A_i^\top \lambda_i^k + \textstyle\sum_{j=1}^N w_{i,j}(\sigma_i^k - \sigma_j^k) + \mathrm{N}_{\Omega_i}(y) \\
z_i^{k+1} &= z_i^k + \textstyle\sum_{j=1}^N \nu_{(i,j)} w_{i,j}(\lambda_i^k - \lambda_j^k) \\
\lambda_i^{k+1} &= \mathrm{proj}_{\mathbb{R}^m_{\geq 0}}\big(\lambda_i^k + \delta_i\big(A_i(2x_i^{k+1} - x_i^k) - b_i - (2z_i^{k+1} - z_i^k)\big)\big).
\end{aligned} \tag{24}
$$

Because of the partial-decision information assumption, no agent has access to the actual value of the average strategy. Instead, we equip each agent with an auxiliary error variable $s_i \in \mathbb{R}^{\bar{n}}$, which is an estimate of the quantity $\mathrm{avg}(x) - x_i$. Each agent aims at reconstructing the true aggregate value, based on the information received from its neighbors. In particular, it should hold that $s^k \to \mathbf{1}_N \otimes \mathrm{avg}(x^k) - x^k$ asymptotically, where $s := \mathrm{col}((s_i)_{i\in\mathcal{I}})$. For brevity of notation, we also denote

$$
\sigma_i := x_i + s_i, \quad \sigma := \mathrm{col}((\sigma_i)_{i\in\mathcal{I}}). \tag{27}
$$

Remark 9. By the updates in Algorithm 4, we can infer an important invariance property, namely that $\mathrm{avg}(s^k) = \mathbf{0}_{\bar{n}}$, or equivalently $\mathrm{avg}(x^k) = \mathrm{avg}(\sigma^k)$, for any $k \in \mathbb{N}$, provided that the algorithm is initialized appropriately, i.e., $s_i^0 = \mathbf{0}_{\bar{n}}$, for all $i \in \mathcal{I}$. In fact, the update of σ, as it follows from Algorithm 4, is

$$
\sigma^{k+1} = \sigma^k - \beta \boldsymbol{L}_{\bar{n}} \sigma^k + (x^{k+1} - x^k), \tag{28}
$$

where $\boldsymbol{L}_{\bar{n}} := L \otimes I_{\bar{n}}$. This update is a dynamic tracking for the time-varying quantity $\mathrm{avg}(x)$, similar to those considered for aggregative games in Belgioioso et al. (2021), Koshal et al. (2016) and Gadjov and Pavel (2021). Differently from Gadjov and Pavel (2021), here we introduce the error variables s_i, which allow us to directly recast the iteration in (28) in an operator-theoretic framework. □

Similarly to Section 4, we study the convergence of Algorithm 4 by relating it to the iteration in (13). First, let us define the *extended pseudo-gradient* mapping

$$
\tilde{\boldsymbol{F}}(x, \xi) := \mathrm{col}\left(\big(\tilde{\boldsymbol{F}}_i(x_i, \xi_i)\big)_{i\in\mathcal{I}}\right), \tag{29}
$$

with $\xi := \mathrm{col}((\xi_i)_{i\in\mathcal{I}}) \in \mathbb{R}^n$, and the operators $\tilde{\boldsymbol{F}}_{\mathrm{a}}(x, s) := \mathrm{col}(\alpha \tilde{\boldsymbol{F}}(x, \sigma) + \boldsymbol{L}_{\bar{n}}\sigma, \boldsymbol{L}_{\bar{n}}\sigma)$,

$$
\tilde{\mathcal{A}}(\boldsymbol{\omega}) := \begin{bmatrix} \alpha \tilde{\boldsymbol{F}}(x, \sigma) + \boldsymbol{L}_{\bar{n}}\sigma \\ \boldsymbol{L}_{\bar{n}}\sigma \\ \mathbf{0}_{Em} \\ \boldsymbol{b} \end{bmatrix} + \begin{bmatrix} \boldsymbol{A}^\top \boldsymbol{\lambda} \\ \mathbf{0}_n \\ -\boldsymbol{V}_m \boldsymbol{\lambda} \\ \boldsymbol{V}_m^\top \boldsymbol{v} - \boldsymbol{A}x \end{bmatrix} + \begin{bmatrix} \mathrm{N}_\Omega(x) \\ \mathbf{0}_n \\ \mathbf{0}_{Em} \\ \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\boldsymbol{\lambda}) \end{bmatrix}, \tag{30}
$$

where $\boldsymbol{\omega} := \mathrm{col}(x, s, \boldsymbol{v}, \boldsymbol{\lambda}) \in \mathbb{R}^{2n+Em+Nm}$, and we recall that $\sigma = x + s$ is just a shorthand notation.

Lemma 8. *The mapping $\tilde{\boldsymbol{F}}$ in (29) is $\tilde{\theta}$-Lipschitz continuous, for some $\tilde{\theta} > 0$.* □

Proof. It follows from Lemma 2, by noticing that $\tilde{\boldsymbol{F}}(x, \xi) = \boldsymbol{F}((x, (I_N \otimes \mathbf{1}_{N-1} \otimes I_{\bar{n}})(\frac{N}{N-1}\xi - \frac{1}{N-1}x)))$. ■

Finally, we will assume that the step sizes $\bar{\tau} := \mathrm{diag}((\tau_i I_{\bar{n}})_{i\in\mathcal{I}})$, $\bar{\beta} := \beta I_{N\bar{n}}$, $\bar{\nu} := \mathrm{diag}((\nu_{(i,j)} I_m)_{(i,j)\in\mathcal{E}})$, $\bar{\delta} := \mathrm{diag}((\delta_i I_m)_{i\in\mathcal{I}})$ are chosen such that $\tilde{\Phi} \succ 0$, where

$$
\tilde{\Phi} := \begin{bmatrix} \bar{\tau}^{-1} - \boldsymbol{L}_{\bar{n}} & -\boldsymbol{L}_{\bar{n}} & \mathbf{0} & -\boldsymbol{A}^\top \\ -\boldsymbol{L}_{\bar{n}} & \bar{\beta}^{-1} - \boldsymbol{L}_{\bar{n}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\nu}^{-1} & \boldsymbol{V}_m \\ -\boldsymbol{A} & \mathbf{0} & \boldsymbol{V}_m^\top & \bar{\delta}^{-1} \end{bmatrix}. \tag{31}
$$

Lemma 9. *The matrix $\tilde{\Phi}$ in (31) is positive definite if $\beta^{-1} > 4\max((d_i)_{i\in\mathcal{I}})$, $\nu_{(i,j)}{}^{-1} > 2\sqrt{(w_{i,j})}$ for all $(i, j) \in \mathcal{E}$, and $\tau_i^{-1} > 4d_i + \|A_i^\top\|_\infty$, $\delta_i^{-1} > \|A_i\|_\infty + \sum_{j=1}^N \sqrt{w_{i,j}}$ for all $i \in \mathcal{I}$.* □

Theorem 4. *Let $d_{\min} := \min((d_i)_{i\in\mathcal{I}})$ and*

$$
\tilde{\alpha}_{\max} := \min\left(\tfrac{4\mu\lambda_2(L)}{\tilde{\theta}^2}, \tfrac{2\sqrt{2}(d_{\min})}{\tilde{\theta}}\right). \tag{32}
$$

Let $\alpha \in (0, \tilde{\alpha}_{\max}]$ and let the step sizes $\bar{\tau}, \bar{\beta}, \bar{\nu}, \bar{\delta}$ be as in Lemma 9. Then, for all $k \in \mathbb{N}$, the inclusion in (24) has a unique solution. Moreover, the sequence $(x^k, s^k, \boldsymbol{z}^k, \boldsymbol{\lambda}^k)_{k\in\mathbb{N}}$ generated by Algorithm 4 converges to an equilibrium $(x^*, \mathbf{1} \otimes \mathrm{avg}(x^*) - x^*, \boldsymbol{z}^*, \boldsymbol{\lambda}^*)$, where x^* is the v-GNE of the game in (1). □

Proof. Similarly to Lemma 6, we first show that Algorithm 4 can be recast as a PPPA, applied to find a zero of the operator $\tilde{\Phi}^{-1}\tilde{\mathcal{A}}$. Then, we restrict our analysis to the invariant subspace

$$
\Sigma := \{(x, s, \boldsymbol{v}, \boldsymbol{\lambda}) \in \mathbb{R}^{2n+Em+Nm} \mid \mathrm{avg}(s) = \mathbf{0}_{\bar{n}}\}. \tag{33}
$$

A detailed proof is in Appendix B. ■

Remark 10. The update in (24) is implicitly defined by a strongly monotone inclusion, or, equivalently, variational inequality (see Appendix B). We emphasize that there are several iterative methods to find the unique solution (with linear rate) (Bauschke & Combettes, 2017, §26) and that, as in Remark 8, convergence is guaranteed even if the solution is approximated at each step (with summable errors). □

Remark 11. If, for some $i \in \mathcal{I}$, there exists a function φ_i such that $\nabla_y \varphi_i(y, s_i^{k+1}) = \tilde{\boldsymbol{F}}_i(y, y + s_i^{k+1})$, then the update of x_i^k in Algorithm 4 can be simplified as

$$
\begin{aligned}
x_i^{k+1} = \underset{y\in\Omega_i}{\mathrm{argmin}}\Big(&\varphi_i(y, s_i^{k+1}) + \tfrac{1}{2\alpha\tau_i}\|y - x_i^k\|^2 \\
&+ \tfrac{1}{\alpha}(A_i^\top \lambda_i^k)^\top y + \tfrac{1}{\alpha}\big(\textstyle\sum_{j=1}^N w_{ij}(\sigma_i^k - \sigma_j^k)\big)^\top y\Big),
\end{aligned}
$$

as in Lemma 6. For scalar games (i.e., $\bar{n} = 1$) this condition holds for all $i \in \mathcal{I}$. Another noteworthy example is that of a cost $f_i(x_i, \mathrm{avg}(x)) = \tilde{f}_i(x_i) + (Q_i \mathrm{avg}(x))^\top x_i$, for some function $\tilde{f}_i$ and symmetric matrix Q_i, which models applications as the Nash–Cournot game described in Koshal et al. (2016) and the resource allocation problem considered in Belgioioso and Grammatico (2017). In this case, $\varphi_i(x_i, s_i) = \tilde{f}_i(x_i) + (Q_i(s_i + x_i))^\top x_i - \frac{N-1}{2N} x_i^\top Q_i x_i$. □

Remark 12. Inertial/relaxed versions of Algorithm 4 can be studied as in Section 5; further, linear convergence can be established for aggregative games without coupling constraints, based on the restricted strong monotonicity of $\boldsymbol{F}_{\mathrm{a}}$ (see the proof of Lemma 11 in Appendix B), as in Theorem 3. □

7. Numerical simulations

7.1. Nash–Cournot game

We consider a Nash–Cournot game (Pavel, 2020, §6), where N firms produce a commodity that is sold to m markets. Each firm $i \in \mathcal{I} = \{1, \dots, N\}$ participates in $n_i \leq m$ of the markets, and decides on the quantities $x_i \in \mathbb{R}^{n_i}$ of commodity to be delivered to these n_i markets. The quantity of product that each firm can deliver is bounded by the local constraints $\mathbf{0}_{n_i} \leq x_i \leq X_i$. Moreover, each market $k = 1, \dots, m$ has a maximal capacity r_k. This results in the shared affine constraint $Ax \leq r$, with $r = \mathrm{col}((r_k)_{k=1,\dots,m})$ and $A = [A_1 \dots A_N]$, where $A_i \in \mathbb{R}^{m \times n_i}$ is the matrix that expresses which markets firm i participates in. Specifically, $[A_i]_{k,j} = 1$ if $[x_i]_j$ is the amount of product sent to the kth market by agent i, $[A_i]_{k,j} = 0$ otherwise, for all $j = 1, \dots, n_i$, $k = 1, \dots, m$. Hence, $Ax = \sum_{i=1}^{N} A_i x_i \in \mathbb{R}^m$ is the vector of the quantities of total product delivered to the markets. Each firm i aims at maximizing its profit, i.e., minimizing the cost function $J_i(x_i, x_{-i}) = 10^{-3} * (c_i(x_i) - p(Ax)^\top A_i x_i)$. Here, $c_i(x_i) = x_i^\top Q_i x_i + q_i^\top x_i$ is firm i's production cost, with $Q_i \in \mathbb{R}^{n_i \times n_i}$, $Q_i \succ 0$, $q_i \in \mathbb{R}^{n_i}$. Instead, $p : \mathbb{R}^m \to \mathbb{R}^m$ associate to each market a price that depends on the amount of product delivered to that market. Specifically, the price for the market k, for $k = 1, \dots, m$, is $[p(x)]_k = \bar{P}_k - \chi_k [Ax]_k$, where $\bar{P}_k$, $\chi_k > 0$.

We set $N = 20$, $m = 7$. The market structure (i.e., which firms are allowed to participate in which of the m markets) is defined as in Pavel (2020, Fig. 1); thus $x = \mathrm{col}((x_i)_{i \in \mathcal{I}}) \in \mathbb{R}^n$ and $n = 32$. The firms cannot access the production of all the competitors, but they are allowed to communicate with their neighbors on a randomly generated connected graph. We select randomly with uniform distribution r_k in $[1, 2]$, Q_i diagonal with diagonal elements in $[1, 8]$, q_i in $[1, 2]$, $\bar{P}_k$ in $[10, 20]$, χ_k in $[1, 3]$, X_i in $[5, 10]$, for all $i \in \mathcal{I}$, $k = 1, \dots, m$.

The resulting setup satisfies all our theoretical assumptions (Pavel, 2020, Section VI). We set $\alpha = \alpha_{\max} \approx 0.7$ as in Lemma 3 and we choose the step sizes as in Lemma 5 to satisfy all the conditions of Theorem 2.

We compare the performance of Algorithm 1 versus that of the pseudo-gradient method in Pavel (2020, Alg. 1), which is to the best of our knowledge the only other available single-layer fixed-step scheme to solve GNE problems under partial-decision information. In Pavel (2020, Alg. 1), we choose the parameter c that maximize the step sizes τ, ν, σ, provided that the conditions in Pavel (2020, Th. 2) are satisfied. This results in very small step sizes, e.g., $\tau^* \approx 10^{-5}$.

The results are illustrated in Fig. 1, where the two Algorithms are initialized with the same random initial conditions. Pavel (2020, Alg. 1) is extremely slow, due to the small step sizes; and our PPPA method shows a much faster convergence. According to our numerical experience, the bounds on the parameters are conservative, and in effect we observe faster convergence for larger step sizes. For Pavel (2020, Alg. 1), the fastest convergence is attained by setting the step sizes 10^4 times bigger than the theoretical bounds; for larger steps, convergence is lost.

We repeat the simulation for different numbers of agents (and random market structures). Differently from Algorithm 1, the upper bounds for the step sizes in Pavel (2020, Alg. 1) decrease when N grows (see Section 5.1), resulting in a greater performance degradation, as shown in Fig. 2 (with theoretical parameters for our PPPA, and steps 10^3 times larger than their upper bounds for Pavel, 2020, Alg. 1).

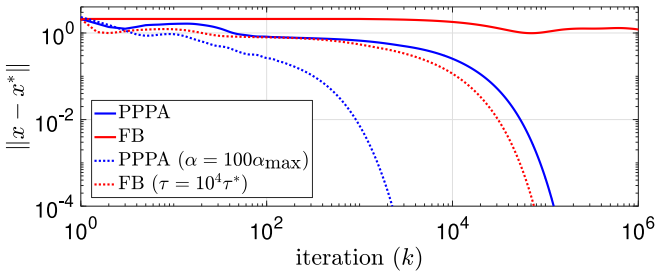


Fig. 1. Distance from the v-GNE for our PPPA (Algorithm 1) and the FB algorithm in Pavel (2020, Alg. 1), for different parameters (the solid line for the theoretical step sizes).

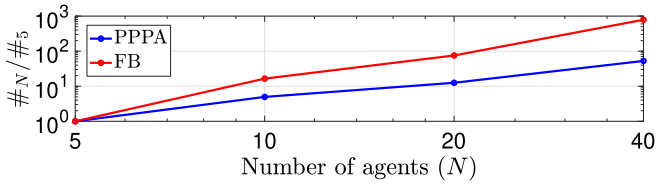


Fig. 2. Variation of the number of iterations $\#_N$ needed to reach a precision of $\|x^k - x^*\| \leq 10^{-2}$ for different values of the number of agents N (in logarithmic scale): our PPPA (Algorithm 1) versus the FB method in Pavel (2020, Alg. 1).

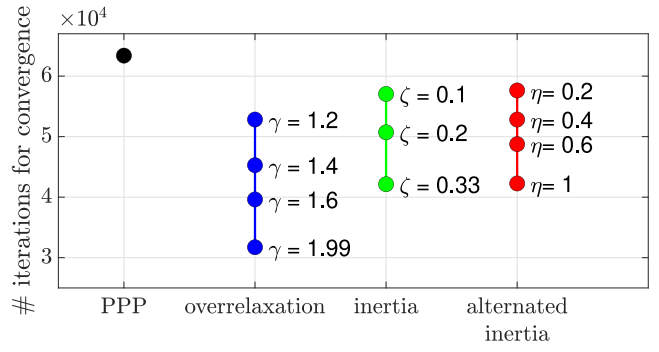


Fig. 3. Number of iterations needed to reach a precision of $\|x^k - x^*\| \leq 10^{-2}$, with different acceleration schemes and parameters.

Finally, we apply the acceleration schemes discussed in Section 5 to Algorithm 1, with parameters that theoretically ensure convergence. The impact is remarkable, up to halving the number of iterations needed for convergence, as shown in Fig. 3.

7.2. Charging of plug-in electric vehicles

We consider the charging scheduling problem for a group of plug-in electric vehicles, modeled by an aggregative game (Grammatico, 2017). Each user $i \in \mathcal{I} = \{1, \dots, N\}$ plans the charging of its vehicle for an horizon of 24 hours, discretized into $\bar{n}$ intervals; the goal is to choose the energy injections $x_i \in \mathbb{R}^{\bar{n}}$ of each time interval to minimize its cost $J_i(x_i, \mathrm{avg}(x)) = g_i(x_i) + p(\mathrm{avg}(x))^\top x_i$, where $g_i(x_i) = x_i^\top Q_i x_i + c_i^\top x^i$ is the battery degradation cost, and $p(\xi) = a(\xi + d) + b\mathbf{1}_{\bar{n}}$ is the cost of energy, with b a baseline price, a the inverse of the price elasticity and $d \in \mathbb{R}^{\bar{n}}$ the inelastic demand (not related to vehicle charging) along the horizon. We assume a maximum injection per interval and a desired final charge level for each user, resulting in the local constraints $\Omega_i = \{y \in [\mathbf{0}_{\bar{n}}, \bar{x}_i] \mid \mathbf{1}_{\bar{n}}^\top y = \gamma_i\}$. Moreover, we consider the transmission line constraints $\mathbf{0}_{\bar{n}} \leq \sum_{i \in \mathcal{I}} x_i \leq \bar{c}N$.

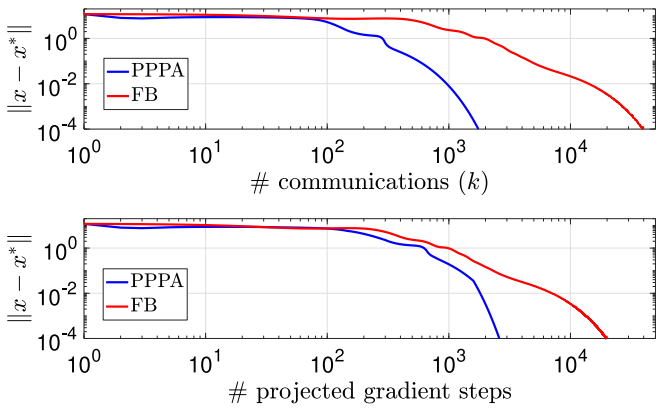


Fig. 4. Distance of the primal variable from the v-GNE. Our PPPA (Algorithm 4) outperforms the FB algorithm in Gadjov and Pavel (2021, Alg. 1), in terms of both communication rounds and performed projected gradient steps.

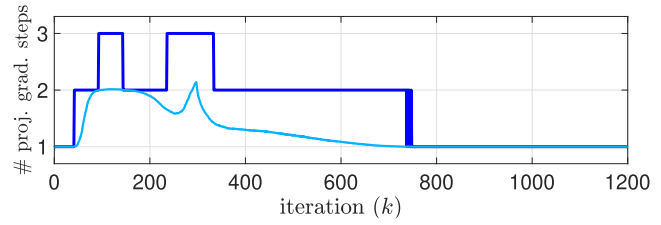


Fig. 5. Maximum (blue) and average (light blue) number of projected gradient steps performed by the agents at each iteration in Algorithm 4, with guaranteed accuracy of $\varepsilon^k = 1/k^2$.

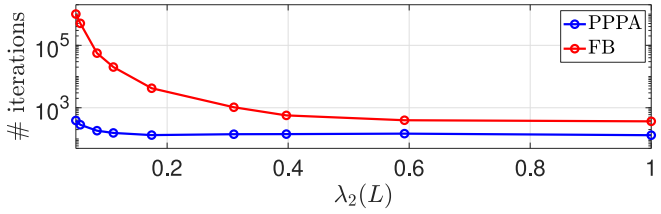


Fig. 6. Number of iterations to reach a precision of $\|x - x^*\| \leq 10^{-2}$ for different values of the algebraic connectivity, where $\lambda_2(L) = 1$ indicates a complete graph (all the graphs are doubly stochastic): our PPPA (Algorithm 4) versus the FB in Gadjov and Pavel (2021, Alg. 1).

We set $N = 1000$, $\bar{n} = 12$. For all $i \in \mathcal{I}$, we select with uniform distribution c_i in $[0.55, 0.95]$, $Q_i \succ 0$ with diagonal and off-diagonal elements in $[0.2, 0.8]$ and $[0, 0.05]$, respectively, γ_i in $[0.6, 1]$; $[\bar{x}_i]_j = 0.25$ with probability 20%, $[\bar{x}_i]_j = 0$ otherwise. We set $[\bar{c}]_j$ as 0.04 if $j \in \{1, 2, 3, 11, 12\}$, as 0.01 otherwise (corresponding to more restrictive limitations in the daytime); $a = 0.38$, $b = 0.6$ and d as in Grammatico (2017). We check numerically that Standing Assumptions 1, 2 hold, and let the agents communicate over a randomly generated connected graph. We implement Algorithm 4, by performing only a finite number of gradient steps per iteration; each agent i uses the stopping criterion in (19) to ensure an accuracy of $\varepsilon_i^k = 1/k^2$. Fig. 4 compares the performance of Algorithm 4 and Gadjov and Pavel (2021, Alg. 1) (which requires two rounds of communication per iteration), with step sizes set to their theoretical upper bounds. Notably, our PPPA significantly outperforms (Gadjov & Pavel, 2021, Alg. 1), even in terms of total projected gradient steps required (for Algorithm 4, we consider the maximum among the agents at each iteration). Interestingly, Fig. 5 shows that the maximum number of performed gradient steps at each iteration is 3 and decreases as the iteration converges, despite the increasing accuracy required in the local optimizations (see also Remark 8).

Differently from our PPPA, the upper bounds for the step sizes in Gadjov and Pavel (2021) are proportional to the quantity $\mu_{\bar{A}}$ in Gadjov and Pavel (2021, Lem. 4), hence they depend on $\lambda_2(L)$, θ_0, μ, θ (but not on N, cf. Sections 7.1, 5.1); in turn, we expect these parameters to affect to a larger extent the convergence speed for the FB method. In Fig. 6 we compare the two algorithms, with $N = 10$, for different values of the communication graph connectivity: in the considered range, the number of iterations to converge varies by a factor 2 for Algorithm 4, by a factor 10^3 for Gadjov and Pavel (2021, Alg. 1).

8. Conclusion

Inexact preconditioned proximal-point methods are extremely efficient to design fully-distributed single-layer generalized Nash equilibrium seeking algorithms. The advantage is that convergence can be guaranteed for much larger step sizes compared to pseudo-gradient-based algorithms. In fact, in our numerical experience, our algorithms proved much faster than the existing methods, resulting in a considerable reduction of communication and computation requirements. Besides, the operator-theoretic approach facilitates the design of acceleration schemes, also in the partial-decision information setup. As future work, it would be highly valuable to relax our monotonicity and connectivity assumptions, namely to allow for merely monotone game mappings and jointly connected networks, and to address the case of nonlinear coupling constraints.

Appendix A. Proof of Theorem 1

For all $k \in \mathbb{N}$, let $z^k := \omega^k + \gamma^k(u^k - \omega^k)$, so that $\omega^{k+1} = z^k + \gamma^k e^k$. Consider any $\omega^* \in C$. We have, for all $k \in \mathbb{N}$,

$$\begin{aligned} &\|z^k - \omega^*\|_P^2 \\ &= \|\omega^k - \omega^*\|_P^2 - 2\gamma^k\langle \omega^k - u^k \mid \omega^k - \omega^*\rangle_P + (\gamma^k)^2\|u^k - \omega^k\|_P^2 \\ &\leq \|\omega^k - \omega^*\|_P^2 - \gamma^k(2 - \gamma^k)\|u^k - \omega^k\|_P^2, \end{aligned} \tag{A.1}$$

where the inequality follows by Lemma 4.

(i) By (A.1), $\|z^k - \omega^*\|_P \leq \|\omega^k - \omega^*\|_P$, and the conclusion follows by the Cauchy–Schwarz inequality.

(ii) By $(\gamma^k\|e^k\|_P)_{k\in\mathbb{N}} \in \ell^1$ and point *(i)*, $(\omega^k)_{k\in\mathbb{N}}$ is bounded. Let $c := \sup_{k\in\mathbb{N}}\|\omega^k - \omega^*\|_P < \infty$ and $\epsilon^k := 2c(\gamma^k\|e^k\|_P) + (\gamma^k\|e^k\|_P)^2$, for all $k \in \mathbb{N}$. Clearly, $(\epsilon^k)_{k\in\mathbb{N}} \in \ell^1$. Moreover, for all $k \in \mathbb{N}$ we have

$$\begin{aligned} &\|w^{k+1} - w^*\|_P^2 \\ &\leq (\|z^k - w^*\|_P + \gamma^k\|e^k\|_P)^2 \\ &\leq \|w^k - w^*\|_P^2 - \gamma^k(2 - \gamma^k)\|u^k - w^k\|_P^2 + \epsilon^k, \end{aligned} \tag{A.2}$$

and the thesis follows by recursion.

(iii) It follows by (A.2) and the results in Combettes (2001, Prop. 3.2(i)) and Combettes (2001, Th. 3.8).

(iv) By definition of resolvent, $\omega^k - u^k \in \mathcal{B}(u^k)$; hence

$$\langle u^k - \omega^* \mid \omega^k - u^k\rangle_P \geq \mu_{\mathcal{B}}\|u^k - \omega^*\|_P^2. \tag{A.3}$$

By the Cauchy–Schwarz inequality, $\|\omega^k - u^k\|_P \geq \mu_{\mathcal{B}}\|u^k - \omega^*\|_P$. Thus, (A.3) yields

$$\begin{aligned} &\|\omega^k - \omega^*\|_P^2 \\ &= \|u^k - \omega^*\|_P^2 + 2\langle u^k - \omega^* \mid \omega^k - u^k\rangle_P + \|\omega^k - u^k\|_P^2 \\ &\geq (1 + \mu_{\mathcal{B}})^2\|u^k - \omega^*\|_P^2. \end{aligned} \tag{A.4}$$

If $\gamma^k \leq 1$, by the Cauchy–Schwarz inequality and (A.4), we have $\|z^k - \omega^*\|_P \leq (1-\gamma^k)\|\omega^k - \omega^*\|_P + \gamma^k\|u^k - \omega^*\|_P \leq (1 - \frac{\gamma^k\mu_{\mathcal{B}}}{1+\mu_{\mathcal{B}}})\|\omega^k - \omega^*\|_P$. For $\gamma^k > 1$, we can write

$$\|z^k - \omega^*\|_P^2$$

$$= (1-\gamma^k)^2\|\omega^k-\omega^*\|_P^2 + \gamma^k(2-\gamma^k)\|u^k-\omega^*\|_P^2 + 2\gamma^k(1-\gamma^k)\langle u^k-\omega^* \mid \omega^k-u^k\rangle_P$$
$$\leq (1-\gamma^k)^2\|\omega^k-\omega^*\|_P^2 + \gamma^k(2(1+\mu_{\mathcal{B}})-\gamma^k(1+2\mu_{\mathcal{B}}))\|u^k-\omega^*\|_P^2 \quad \text{(A.5)}$$
$$\leq (\max(1-\tfrac{\gamma^k\mu_{\mathcal{B}}}{1+\mu_{\mathcal{B}}}, \gamma^k-1))^2\|\omega^k-\omega^*\|_P^2, \quad \text{(A.6)}$$

where the first equality follows by rearranging the terms in (A.1); in the first inequality we used (A.3); the last inequality follows by taking into account that the second term in (A.5) is nonpositive if $\gamma^k \in (1, 1+\frac{1}{1+2\mu_{\mathcal{B}}}]$ and can be upper bounded via (A.4) if $\gamma^k \in [1+\frac{1}{1+2\mu_{\mathcal{B}}}, 2)$. Finally, assume that $\omega^k \in C$, and choose $u^k = \omega^k$. Then (A.6) implies $\omega^k = \omega^*$, hence C must be a singleton. ■

Appendix B. Proof of Theorem 4

Analogously to Lemma 6, it can be shown that Algorithm 4 is equivalent to the iteration

$$\boldsymbol{\omega}^{k+1} \in \mathrm{J}_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}(\boldsymbol{\omega}^k), \quad \boldsymbol{\omega}^0 = \bar{\boldsymbol{\omega}}^0, \quad \text{(B.1)}$$

where $\bar{\boldsymbol{\omega}}^0 = (\boldsymbol{x}^0, \mathbf{0}_n, \mathbf{0}_{Em}, \boldsymbol{\lambda}^0)$, for some $\boldsymbol{x}^0 \in \Omega$, $\boldsymbol{\lambda}^0 \in \mathbb{R}^{Nm}_{\geq 0}$, modulo the transformation $\boldsymbol{z}^k = \boldsymbol{V}_m^\top \boldsymbol{v}^k$.

First, we show that the iteration in (B.1) is uniquely defined. For all $i \in \mathcal{I}$, let $\mathcal{F}_i(y, \vartheta^k) := \alpha\tilde{\boldsymbol{F}}_i(y, y+s_i^{k+1}) + \frac{1}{\tau_i}(y-x_i^k) + A_i^\top\lambda_i^k + \sum_{j=1}^N w_{i,j}(\sigma_i^k - \sigma_j^k) + \mathrm{N}_{\Omega_i}(y)$, where $\vartheta^k = (x^k, s^{k+1}, s^k, \boldsymbol{\lambda}^k)$. We note that $\tilde{\boldsymbol{F}}_i$ is $\tilde{\theta}$-Lipschitz, because $\tilde{\boldsymbol{F}}$ is $\tilde{\theta}$-Lipschitz by Lemma 8. Then, by monotonicity of the normal cone, we have $\langle y-y' \mid \mathcal{F}_i(y,\vartheta^k) - \mathcal{F}_i(y',\vartheta^k)\rangle \geq (\tau_i^{-1} - \alpha\sqrt{2}\tilde{\theta})\|y-y'\|^2$, for any $y, y' \in \mathbb{R}^n$, for any ϑ^k. By the assumption on α, $\mathcal{F}_i$ is strongly monotone in y for any ϑ^k, hence the inclusion in (24) has a unique solution, for any ϑ^k (Bauschke & Combettes, 2017, Cor. 23.37). Therefore, it also holds that $\mathrm{dom}(\mathrm{J}_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}) = \mathbb{R}^{2n+Em+Nn}$ and that $\mathrm{J}_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}$ is single-valued.

We turn our attention to the set Σ in (33). As in Remark 9, for any $\varsigma \in \Sigma$, $\mathrm{J}_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}(\varsigma) \in \Sigma$; hence Σ is invariant for (B.1). Moreover, $\boldsymbol{\omega}^0 \in \Sigma$. Hence, in (B.1), it is enough to consider the operator $J_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}|_\Sigma$, where $\mathcal{B}|_\Sigma$ is the restriction of the operator $\mathcal{B}$ to Σ, i.e., $\mathcal{B}|_\Sigma(\boldsymbol{\omega}) = \mathcal{B}(\boldsymbol{\omega})$ if $\boldsymbol{\omega} \in \Sigma$, $\mathcal{B}|_\Sigma(\boldsymbol{\omega}) = \varnothing$ otherwise. By invariance and (16), it also follows that $J_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}}\,|_\Sigma = J_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}|_\Sigma}\,|_\Sigma$. Thus, the iteration in (B.1) is rewritten as

$$\boldsymbol{\omega}^{k+1} = \mathrm{J}_{\tilde{\Phi}^{-1}\tilde{\mathcal{A}}|_\Sigma}(\boldsymbol{\omega}^k), \quad \boldsymbol{\omega}^0 = \bar{\boldsymbol{\omega}}^0 \in \Sigma. \quad \text{(B.2)}$$

We show the convergence of (B.2) by studying the properties of $\tilde{\mathcal{A}}|_\Sigma$. We start by characterizing the zero set.

Lemma 10. *The following statements hold:*

(i) If $\mathrm{col}(x^*, s^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\tilde{\mathcal{A}}|_\Sigma)$, *then* $s^* = \mathbf{1}_N \otimes \mathrm{avg}(x^*) - x^*$ *and* x^* *is the v-GNE of the game in* (1).
(ii) $\mathrm{zer}(\tilde{\mathcal{A}}|_\Sigma) \neq \varnothing$. □

Proof. Let $\boldsymbol{V}_q := V \otimes I_q$, $\boldsymbol{L}_q := L \otimes I_q = \boldsymbol{V}_q^\top \boldsymbol{V}_q$, for any $q > 0$; hence, under Standing Assumption 3, we have

$$\mathrm{null}\left(\boldsymbol{L}_q\right) = \mathrm{null}\left(\boldsymbol{V}_q\right) = \mathrm{range}(\mathbf{1}_N \otimes I_q) \quad \text{(B.3)}$$
$$\mathrm{range}(\boldsymbol{V}_q^\top) \supseteq \mathrm{range}\left(\boldsymbol{L}_q\right) = \mathrm{null}(\mathbf{1}_N^\top \otimes I_q). \quad \text{(B.4)}$$

(i) Let us consider any $\boldsymbol{\omega}^* = \mathrm{col}(x^*, s^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\tilde{\mathcal{A}}|_\Sigma)$, and let $\sigma^* = x^* + s^*$; then we have

$$\mathbf{0}_{\bar{n}} \in \alpha\tilde{\boldsymbol{F}}(x^*, \sigma^*) + \boldsymbol{L}_{\bar{n}}\sigma^* + \mathrm{N}_\Omega(x^*) + \boldsymbol{A}^\top\boldsymbol{\lambda}^* \quad \text{(B.5a)}$$
$$\mathbf{0}_{\bar{n}} = \boldsymbol{L}_{\bar{n}}\sigma^* \quad \text{(B.5b)}$$
$$\mathbf{0}_{Em} = -\boldsymbol{V}_m\boldsymbol{\lambda}^* \quad \text{(B.5c)}$$
$$\mathbf{0}_{Nm} \in \boldsymbol{b} + \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\boldsymbol{\lambda}^*) - \boldsymbol{A}x^* + \boldsymbol{V}_m^\top\boldsymbol{v}^* \quad \text{(B.5d)}$$

By (B.5c) and by (B.3), we have $\boldsymbol{\lambda}^* = \mathbf{1}_N \otimes \lambda^*$, for some $\lambda^* \in \mathbb{R}^m$; by (B.5b) and since $\boldsymbol{\omega}^* \in \Sigma$, it must hold $\sigma^* = x^* + s^* = 1_N \otimes \mathrm{avg}(x^*)$. It is then enough to prove that the pair (x^*, λ^*) satisfies the KKT conditions in (3). By (B.5a) and by recalling that $\boldsymbol{A}^\top(\mathbf{1}_N \otimes \lambda^*) = A^\top\lambda^*$ and $\tilde{\boldsymbol{F}}(x^*, \mathbf{1}_N \otimes x^*) = F(x^*)$, we retrieve the first KKT condition in (3). We obtain the second KKT condition by left-multiplying both sides of (B.5d) with $(\mathbf{1}_N^\top \otimes I_m)$ and using that $(\mathbf{1}_N^\top \otimes I_m)\boldsymbol{b} = b$, $(\mathbf{1}_N^\top \otimes I_m)\boldsymbol{L}_m = 0$ by (B.3) and symmetry of L, $(\mathbf{1}_N^\top \otimes I_m)\boldsymbol{A} = A$ and $(\mathbf{1}_N^\top \otimes I_m)\mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\mathbf{1}_N \otimes \lambda^*) = N\mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*) = \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*)$.

(ii) Let us consider any pair (x^*, λ^*) satisfying the KKT conditions in (3) (one such pair exists by Assumption 2). We next show that there exists $\boldsymbol{v}^* \in \mathbb{R}^{Em}$ such that $\boldsymbol{\omega}^* = \mathrm{col}(x^*, \mathbf{1}_N \otimes \mathrm{avg}(x^*) - x^*, \boldsymbol{z}^*, \mathbf{1}_N \otimes \lambda^*) \in \mathrm{zer}(\tilde{\mathcal{A}}|_\Sigma)$. Clearly, $\boldsymbol{\omega}^* \in \Sigma$. Besides, $\boldsymbol{\omega}^*$ satisfies the conditions (B.5a)–(B.5c), as in point *(i)*. By (3), there exists $u^* \in \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*)$ such that $Ax^* - b - u^* = \mathbf{0}_n$. Also, $\mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(1_N \otimes \lambda^*) = \prod_{i\in\mathcal{I}} \mathrm{N}_{\mathbb{R}^m_{\geq 0}}(\lambda^*)$, and it follows by properties of cones that $\mathrm{col}\left(u_1^*, \ldots, u_N^*\right) \in \mathrm{N}_{\mathbb{R}^{Nm}_{\geq 0}}(\mathbf{1}_N \otimes \lambda^*)$, with $u_1^* = \cdots = u_N^* = \frac{1}{N}u^*$. Hence $(\mathbf{1}_N^\top \otimes I_m)\left(-\boldsymbol{A}x^* + \boldsymbol{b} + \mathrm{col}\left(u_1^*, \ldots, u_N^*\right)\right) = b - Ax^* + u^* = \mathbf{0}_m$, or $-\boldsymbol{A}x^* + \boldsymbol{b} + \mathrm{col}\left(u_1^*, \ldots, u_N^*\right) \in \mathrm{null}(\mathbf{1}_N^\top \otimes I_m) \subseteq \mathrm{range}(\boldsymbol{V}_m^\top)$, by (B.4). Therefore there exists $\boldsymbol{v}^*$ such that also the condition (B.5d) is satisfied, for which $\boldsymbol{\omega}^* \in \mathrm{zer}(\tilde{\mathcal{A}})$. ■

Next, similar to Lemma 3, we show restricted monotonicity of the operator $\tilde{\mathcal{A}}|_\Sigma$.

Lemma 11. *Let* $\alpha \in (0, \tilde{\alpha}_{\max}]$, *with* $\tilde{\alpha}_{\max}$ *as in* (32). *Then* $\tilde{\mathcal{A}}|_\Sigma$ *is restricted monotone.* □

Proof. The operator $\tilde{\mathcal{A}}\,|_\Sigma$ is the sum of three components, as in (30). The third is monotone by properties of the normal cones (Bauschke & Combettes, 2017, Th. 20.25), the second because it is a linear skew-symmetric operator (Bauschke & Combettes, 2017, Ex. 20.35) (and restriction does not cause loss of monotonicity, by definition). For the first term, let $(\boldsymbol{\omega}, \boldsymbol{u}) \in \mathrm{gra}(\tilde{\mathcal{A}}|_\Sigma)$, $\boldsymbol{\omega} := \mathrm{col}(x, s, \boldsymbol{v}, \boldsymbol{\lambda})$, $\boldsymbol{\omega}^* = \mathrm{col}(x^*, s^*, \boldsymbol{v}^*, \boldsymbol{\lambda}^*) \in \mathrm{zer}(\tilde{\mathcal{A}}|_\Sigma)$, $\sigma = x+s$, $\sigma^* = s^*+x^*$. By Lemma 10, $s^* = \mathbf{1}_N \otimes \mathrm{avg}(x^*) - x^*$. Then, by Gadjov and Pavel (2021, Lemma 4), there is a $\tilde{\mu} > 0$ such that $\langle \mathrm{col}(x-x^*, s-s^*) \mid \tilde{\boldsymbol{F}}_{\mathrm{a}}(x,s) - \tilde{\boldsymbol{F}}_{\mathrm{a}}(x^*, s^*)\rangle = \langle x-x^* \mid \alpha\tilde{\boldsymbol{F}}(x,\sigma) - \alpha\tilde{\boldsymbol{F}}(x^*,\sigma^*)\rangle + \langle \sigma-\sigma^* \mid \boldsymbol{L}_{\bar{n}}(\sigma-\sigma^*)\rangle \geq \tilde{\mu}\|\mathrm{col}(x-x^*, \sigma-\sigma^*)\|^2 \geq \mu_{\tilde{\boldsymbol{F}}_{\mathrm{a}}}\|\mathrm{col}(x-x^*, s-s^*)\|^2$, where $\mu_{\tilde{\boldsymbol{F}}_{\mathrm{a}}} := (3-\sqrt{5})\tilde{\mu}/2$ and the last inequality follows by definition of σ and bounds on quadratic forms. ■

Finally, the preconditioning matrix $\tilde{\Phi}$ is positive definite by Lemma 9. As in Lemma 7, by Lemma 11, it holds that $\tilde{\Phi}^{-1}\tilde{\mathcal{A}}|_\Sigma$ is restricted monotone in $\mathcal{H}_{\tilde{\Phi}}$. In view of (B.2), the conclusion follows analogously to Theorem 2. ■

References

Bauschke, H. H., & Combettes, P. L. (2017). *Convex analysis and monotone operator theory in Hilbert spaces, Vol. 2011*. Springer.

Belgioioso, G., & Grammatico, S. (2017). Semi-decentralized Nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions. *IEEE Control Systems Letters*, *1*(2), 400–405.

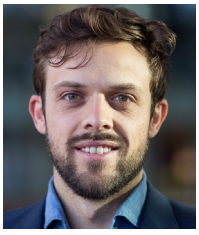
Belgioioso, G., & Grammatico, S. (2018). Projected-gradient algorithms for generalized equilibrium seeking in aggregative games are preconditioned forward-backward methods. In *2018 European Control Conference (ECC)* (pp. 2188–2193).

Belgioioso, G., & Grammatico, S. (2020). Semi-decentralized generalized Nash equilibrium seeking in monotone aggregative games. arXiv preprint, Available online: arxiv.org/pdf/2003.04031.

Belgioioso, G., Nedić, A., & Grammatico, S. (2021). Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks. *IEEE Transactions on Automatic Control*, *66*(5), 2061–2075.

Bianchi, M., Belgioioso, G., & Grammatico, S. (2020). A fully-distributed proximal-point algorithm for Nash equilibrium seeking with linear convergence rate. In *2020 59th IEEE Conference on Decision and Control (CDC)* (pp. 2303–2308).
Bianchi, M., & Grammatico, S. (2020). A continuous-time distributed generalized Nash equilibrium seeking algorithm over networks for double-integrator agents. In *2020 European Control Conference (ECC)* (pp. 1474–1479).
Bianchi, M., & Grammatico, S. (2021). Fully distributed Nash equilibrium seeking over time-varying communication networks with linear convergence rate. *IEEE Control Systems Letters*, *5*(2), 499–504.
Boţ, R. I., Csetnek, E. R., & Hendrich, C. (2015). Inertial Douglas–Rachford splitting for monotone inclusion problems. *Applied Mathematics and Computation*, *256*, 472–487.
Chen, G., Ming, Y., & Hong, Y. (2021). Distributed algorithm for ϵ-generalized Nash equilibria with uncertain coupled constraints. *Automatica*, *123*, 109313.
Combettes, P. L. (2001). Quasi-fejérian analysis of some optimization algorithms. In D. Butnariu, Y. Censor, & S. Reich (Eds.), *Studies in computational mathematics*: *Vol. 8*, *Inherently parallel algorithms in feasibility and optimization and their applications* (pp. 115–152). Elsevier.
El Farouq, N. (2001). Pseudomonotone variational inequalities: Convergence of proximal methods. *Journal of Optimization Theory and Applications*, *109*(2), 311–326.
Facchinei, F., & Kanzow, C. (2010). Generalized Nash equilibrium problems. *Annals of Operations Research*, *175*, 177–211.
Facchinei, F., & Pang, J. (2007). *Finite-dimensional variational inequalities and complementarity problems*. Springer.
Facchinei, F., & Pang, J. (2009). Nash equilibria: the variational approach. In D. P. Palomar, & Y. C. Eldar (Eds.), *Convex optimization in signal processing and communications* (pp. 443–493). Cambridge University Press.
Frihauf, P., Krstic, M., & Basar, T. (2012). Nash equilibrium seeking in noncooperative games. *IEEE Transactions on Automatic Control*, *57*(5), 1192–1207.
Gadjov, D., & Pavel, L. (2019). A passivity-based approach to Nash equilibrium seeking over networks. *IEEE Transactions on Automatic Control*, *64*(3), 1077–1092.
Gadjov, D., & Pavel, L. (2021). Single-timescale distributed GNE seeking for aggregative games over networks via forward–backward operator splitting. *IEEE Transactions on Automatic Control*, *66*(7), 3259–3266.
Godsil, C., & Royle, G. (2013). *Algebraic graph theory. Graduate texts in mathematics, vol.207*. Springer Science & Business Media.
Grammatico, S. (2017). Dynamic control of agents playing aggregative games with coupling constraints. *IEEE Transactions on Automatic Control*, *62*(9), 4537–4548.
Iutzeler, F., & Hendrickx, J. M. (2019). A generic online acceleration scheme for optimization algorithms via relaxation and inertia. *Optimization Methods & Software*, *34*(2), 383–405.
Ivkin, N., Rothchild, D., Ullah, E., Braverman, V., Stoica, I., & Arora, R. (2019). Communication-efficient distributed SGD with sketching. *Advances in Neural Information Processing Systems*, *32*, 13144–13154.
Koshal, J., Nedić, A., & Shanbhag, U. V. (2016). Distributed algorithms for aggregative games on graphs. *Operations Research*, *64*, 680–704.
Kulkarni, A. A., & Shanbhag, U. V. (2012). On the variational equilibrium as a refinement of the generalized Nash equilibrium. *Automatica*, *48*(1), 45–55.
Lei, J., & Shanbhag, U. V. (2018). Linearly convergent variable sample-size schemes for stochastic Nash games: Best-response schemes and distributed gradient-response schemes. In *2018 IEEE Conference on Decision and Control (CDC)* (pp. 3547–3552).
Li, N., Chen, L., & Dahleh, M. A. (2015). Demand response using linear supply function bidding. *IEEE Transactions on Smart Grid*, *6*(4), 1827–1838.
Moudafi, A. (2020). About proximal-type methods for a class of nonmonotone operators. *Bulletin of the Iranian Mathematical Society*, *46*(1), 115–125.
Pang, Y., & Hu, G. (2020). Distributed Nash equilibrium seeking with limited cost function knowledge via a consensus-based gradient-free method. *IEEE Transactions on Automatic Control*.
Parise, F., Gentile, B., & Lygeros, J. (2020). A distributed algorithm for almost-Nash equilibria of average aggregative games with coupling constraints. *IEEE Transactions on Control of Network Systems*, *7*(2), 770–782.
Pavel, L. (2020). Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach. *IEEE Transactions on Automatic Control*, *65*(4), 1584–1597.
Saad, W., Han, Z., Poor, H. V., & Basar, T. (2012). Game-theoretic methods for the smart grid: An overview of microgrid systems, demand-side management, and smart grid communications. *IEEE Signal Processing Magazine*, *29*(5), 86–105.
Salehisadaghiani, F., & Pavel, L. (2016). Distributed Nash equilibrium seeking: A gossip-based algorithm. *Automatica*, *72*, 209–216.
Salehisadaghiani, F., Shi, W., & Pavel, L. (2019). Distributed Nash equilibrium seeking under partial-decision information via the alternating direction method of multipliers. *Automatica*, *103*, 27–35.
Scutari, G., Facchinei, F., Pang, J., & Palomar, D. P. (2014). Real and complex monotone communication games. *IEEE Transactions on Information Theory*, *60*(7), 4197–4231.
Swenson, B., Kar, S., & Xavier, J. (2015). Empirical centroid fictitious play: An approach for distributed learning in multi-agent games. *IEEE Transactions on Signal Processing*, *63*(15), 3888–3901.
Tatarenko, T., Shi, W., & Nedić, A. (2020). Geometric convergence of gradient play algorithms for distributed Nash equilibrium seeking. *IEEE Transactions on Automatic Control*, *66*(11), 5342–5353.
Ye, M., & Hu, G. (2017). Distributed Nash equilibrium seeking by a consensus based approach. *IEEE Transactions on Automatic Control*, *62*(9), 4811–4818.
Yi, P., & Pavel, L. (2019). Distributed generalized Nash equilibria computation of monotone games via double-layer preconditioned proximal-point algorithms. *IEEE Transactions on Control of Network Systems*, *6*(1), 299–311.
Yi, P., & Pavel, L. (2019). An operator splitting approach for distributed generalized Nash equilibria computation. *Automatica*, *102*, 111–121.
Yu, C., van der Schaar, M., & Sayed, A. H. (2017). Distributed learning for stochastic generalized Nash equilibrium problems. *IEEE Transactions on Signal Processing*, *65*(15), 3893–3908.

Mattia Bianchi is a Ph.D. candidate in the Delft Center for Systems and Control, TU Delft, The Netherlands. He received the Bachelor's degree in Information Engineering and the Master's degree in Control Systems Engineering, both from University of L'Aquila, Italy, in July 2016 and July 2018, respectively. From February to June 2018 he visited the Control Systems Technology group, TU Eindhoven, The Netherlands. His research interests include game theory and operator theory to solve decision and control problems for complex systems of systems, in the presence of network coupling and uncertainty.



Giuseppe Belgioioso is a postdoctoral researcher at the Automatic Control Laboratory (ifA), Swiss Federal Institute of Technology (ETH Zürich), Switzerland. He received the bachelor's degree in Information Engineering in 2012 and the master's degree (cum laude) in Automatic Control Engineering in 2015, both at the University of Padova, Italy. In 2020, he obtained the Ph.D. degree in Automatic Control at Eindhoven University of Technology (TU/e), The Netherlands. In 2015, he was also a research scientist (assegnista) at the Control Systems Group (Automatica), University of Padova, Italy. In 2019, he visited the School of Electrical, Computer and Energy Engineering at Arizona State University (ASU), USA.

Sergio Grammatico (S'19) is an Associate Professor at the Delft Center for Systems and Control, TU Delft, The Netherlands. Born in Italy in 1987, he received the Bachelor's degree in Computer Engineering, the Master's degree in Automatic Control Engineering, and the Ph.D. degree in Automatic Control, all from the University of Pisa, Italy, in February 2008, October 2009, and March 2013 respectively. He also received a Master's degree in Engineering Science from the Sant'Anna School of Advanced Studies, the Italian Superior Graduate School (Grande École) for Applied Sciences, in November 2011. In 2013–2015, he was a postdoc researcher in the Automatic Control Laboratory, ETH Zurich, Switzerland. In 2015–2018, he was an Assistant Professor in the Department of Electrical Engineering, Control Systems, TU Eindhoven. He was awarded a 2005 F. Severi B.Sc. Scholarship by the Italian High-Mathematics National Institute, and a 2008 M.Sc. Fellowship by the Sant'Anna School of Advanced Studies. He was awarded 2013 and 2014 "TAC Outstanding Reviewer" by the IEEE Control Systems Society. He was recipient of the Best Paper Award at the 2016 ISDG Int. Conf. on Network Games, Control and Optimization. He is currently an Associate Editor of the IEEE Trans. on Automatic Control (2018–present) and of Automatica (2020–present).