

Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking

Original

Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking / Franci, Barbara; Staudigl, Mathias; Grammatico, Sergio. - (2020), pp. 1274-1279. (2020 European Control Conference (ECC) San Pietroburgo (Russia) 12-15 May 2020) [10.23919/ecc51009.2020.9143676].

Availability:

This version is available at: 11583/3003589 since: 2025-10-10T07:47:18Z

Publisher:

IEEE

Published

DOI:10.23919/ecc51009.2020.9143676

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

IEEE postprint/Author's Accepted Manuscript

©2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

Distributed forward-backward (half) forward algorithms for generalized Nash equilibrium seeking

Barbara Franci¹, Mathias Staudigl² and Sergio Grammatico¹

Abstract—We present two distributed algorithms for the computation of a generalized Nash equilibrium in monotone games. The first algorithm follows from a forward-backward-forward operator splitting, while the second, which requires the pseudo-gradient mapping of the game to be cocoercive, follows from the forward-backward-half-forward operator splitting. Finally, we compare them with the distributed, preconditioned, forward-backward algorithm via numerical experiments.

I. INTRODUCTION

Generalized Nash equilibrium problems (GNEPs) have been widely studied in the literature [1], [2], [3] and such a strong interest is motivated by numerous applications ranging from economics to engineering [4], [5]. In a GNEP, each agent seeks to minimize his own cost function, under local and coupled feasibility constraints. In fact, both the cost function and the constraints depend on the strategies chosen by the other agents. Due to the presence of these shared constraints, the search for generalized Nash equilibria is usually a quite challenging task.

For the computation of a GNE, various algorithms have been proposed, both distributed [6], [7], and semi-decentralized [3], [8]. When dealing with coupling constraints, a common principle is the focus on a special class of equilibria, which reflect some notion of fairness among the agents. This class is known as *variational equilibria* (v-GNE) [9], [3]. Besides fairness considerations, v-GNE is computationally attractive since it can be formulated in terms of variational inequality, which makes it possible to solve them via operator splitting techniques [10], [2]. A recent breakthrough along these lines is the distributed, preconditioned, forward-backward (FB) algorithm conceived in [7] for strongly-monotone games. The key lesson from [7] is that the FB method cannot be directly applied to GNEPs, thus a suitable preconditioning is necessary. From a technical perspective, the FB operator splitting requires that the pseudo-gradient mapping of the game is strongly monotone, an assumption which is not always satisfied.

In this paper we investigate two distributed algorithmic schemes for computing a v-GNE. Motivated by the need

to relax the strong monotonicity assumption on the pseudo-gradient of the game, we first investigate a distributed *forward-backward-forward* (FBF) algorithm [11]. We show that a suitably constructed FBF operator splitting guarantees not only fully distributed computation, but also convergence to a v-GNE under the mere assumption of monotonicity of the involved operators. This enables us to drop the strong monotonicity assumption, which is the main advantage with respect to the FB-based splitting methods [7]. As a second condition, in order to exploit the structure of the monotone inclusion defining the v-GNE problem, we also investigate the *forward-backward-half-forward* (FBHF) algorithm [12]. We would like to point out that both our algorithms are distributed in the sense that each agent needs to know only his local cost function and its local feasible set, and there is no central coordinator that updates and broadcasts the dual variables. The latter is the main difference with semi-decentralized schemes for aggregative games [13], [8].

Compared with the FB and the FBHF algorithms, the FBF requires less restrictive assumptions to guarantee convergence, i.e., plain monotonicity of the pseudo-gradient mapping. On the other hand, the FBF algorithm requires two evaluations of the pseudo-gradient mapping, which means that the agents must communicate at least twice at each iterative step. Confronted with the FBF algorithm, our second proposal, the FBHF algorithm requires only one evaluation of the pseudo-gradient mapping, but needs strong monotonicity to provide theoretical convergence guarantees. Effectively, the FBHF algorithm is guaranteed to converge under the same assumptions as the preconditioned FB [7].

Notation: \mathbb{R} indicates the set of real numbers and $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. $\mathbf{0}_N$ ($\mathbf{1}_N$) is the vector of N zeros (ones). The Euclidean inner product and norm are indicated with $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. Let Φ be a symmetric, positive definite matrix, $\Phi \succ 0$. The induced inner product is $\langle \cdot, \cdot \rangle_\Phi := \langle \Phi \cdot, \cdot \rangle$, and the associated norm is $\|\cdot\|_\Phi := \langle \cdot, \cdot \rangle_\Phi^{1/2}$. We call \mathcal{H}_Φ the Hilbert space with norm $\|\cdot\|_\Phi$. Given a set $\mathcal{X} \subseteq \mathbb{R}^n$, the normal cone mapping is denoted with $N_{\mathcal{X}}(x)$. Id is the identity mapping. Given a set-valued operator A , the graph of A is the set $\text{gph}(A) = \{(x, y) | y \in Ax\}$. The set of zeros is $\text{zer } A = \{x \in \mathbb{R}^n | 0 \in Ax\}$. The resolvent of a maximally monotone operator A is the map $J_A = (\text{Id} + A)^{-1}$. If $g : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is a proper, lower semi-continuous, convex function, its subdifferential is the maximal monotone operator $\partial g(x)$. The proximal operator is defined as $\text{prox}_g^\Phi(v) = J_{\Phi \partial g}(v)$ [10]. Given $x_1, \dots, x_N \in \mathbb{R}^n$, $\mathbf{x} := \text{col}(x_1, \dots, x_N) = [x_1^\top, \dots, x_N^\top]^\top$.

¹Barbara Franci and Sergio Grammatico are with the Delft Center for System and Control, TU Delft, The Netherlands {b.franci-1, s.grammatico}@tudelft.nl.

²Mathias Staudigl is with the Department of Data Science and Knowledge Engineering, Maastricht University, The Netherlands m.staudigl@maastrichtuniversity.nl.

This work was partially supported by NWO under research projects OMEGA (613.001.702) and P2P-TALES (647.003.003), and by the ERC under research project COSMOS (802348). M. Staudigl acknowledges support from the COST Action CA16228 "European Network for Game Theory".

II. MATHEMATICAL SETUP: THE MONOTONE GAME AND VARIATIONAL GENERALIZED NASH EQUILIBRIA

We consider a game with N agents where each agent chooses an action $x_i \in \mathbb{R}^{n_i}$, $i \in \mathcal{I} = \{1, \dots, N\}$.

Each agent i has an extended-valued local cost function $J_i : \mathbb{R}^n \rightarrow (-\infty, \infty]$ of the form

$$J_i(x_i, \mathbf{x}_{-i}) := f_i(x_i, \mathbf{x}_{-i}) + g_i(x_i). \quad (1)$$

where $\mathbf{x}_{-i} = \text{col}(\{x_j\}_{j \neq i})$ is the vector of all decision variables except for x_i , and $g_i : \mathbb{R}^{n_i} \rightarrow (-\infty, \infty]$ is a local idiosyncratic costs function which is possibly non-smooth. Thus, the function J_i in (1) has the typical splitting into smooth and non-smooth parts.

Standing Assumption 1 (Local cost): For each $i \in \mathcal{I}$, the function g_i in (1) is lower semicontinuous and convex. For each $i \in \mathcal{I}$, $\text{dom}(g_i) = \Omega_i$ is a closed convex set. ■

Examples for the local cost function are indicator functions to enforce set constraints, or penalty functions that promote sparsity, or other desirable structure.

For the function f_i in (1), we assume convexity and differentiability, as usual in the GNEP literature [3].

Standing Assumption 2 (Local convexity): For each $i \in \mathcal{I}$ and for all $\mathbf{x}_{-i} \in \mathbb{R}^{n-n_i}$, the function $f_i(\cdot, \mathbf{x}_{-i})$ in (1) is convex and continuously differentiable. ■

We assume that the game displays joint convexity with affine coupling constraints defining the collective feasible set

$$\mathcal{X} := \{\mathbf{x} \in \Omega \mid A\mathbf{x} - b \leq \mathbf{0}_m\} \quad (2)$$

where $A := [A_1, \dots, A_N] \in \mathbb{R}^{m \times n}$ and $b := \sum_{i=1}^N b_i \in \mathbb{R}^m$. Effectively, each matrix $A_i \in \mathbb{R}^{m \times n_i}$ defines how agent i is involved in the coupling constraints, thus we consider it to be private information of agent i . Then, for each i , given the strategies of all other agents \mathbf{x}_{-i} , the feasible decision set is

$$\mathcal{X}_i(\mathbf{x}_{-i}) := \left\{ y_i \in \Omega_i \mid A_i y_i \leq b - \sum_{j \neq i} A_j x_j \right\}. \quad (3)$$

Next, we assume a constraint qualification condition.

Standing Assumption 3: (Constraint qualification) The set \mathcal{X} in (2) satisfies Slater's constraint qualification. ■

The aim of each agent is to solve its local optimization problem

$$\forall i \in \mathcal{I} : \begin{cases} \min_{x_i \in \Omega_i} & J_i(x_i, \mathbf{x}_{-i}) \\ \text{s.t.} & A_i x_i \leq b - \sum_{j \neq i} A_j x_j. \end{cases} \quad (4)$$

Thus, the solution concept for such a competitive scenario is the generalized Nash equilibrium [9], [3].

Definition 1: (Generalized Nash equilibrium) A collective strategy $\mathbf{x}^* = \text{col}(x_1^*, \dots, x_N^*) \in \mathcal{X}$ is a generalized Nash equilibrium of the game in (4) if, for all $i \in \mathcal{I}$,

$$J_i(x_i^*, \mathbf{x}_{-i}^*) \leq \inf \{ J_i(y, \mathbf{x}_{-i}^*) \mid y \in \mathcal{X}_i(\mathbf{x}_{-i}^*) \}. \quad \blacksquare$$

To derive optimality conditions characterizing GNE, we define agent i 's Lagrangian function as $\mathcal{L}_i(x_i, \lambda_i, \mathbf{x}_{-i}) := J_i(x_i, \mathbf{x}_{-i}) + \lambda_i^\top (A\mathbf{x} - b)$ where $\lambda_i \in \mathbb{R}_{\geq 0}^m$ is the Lagrange multiplier associated with the coupling constraint $A\mathbf{x} \leq b$. Thanks to the sum rule of the subgradient for Lipschitz continuous functions [14, §1.8], we can write the subgradient of agent i as $\partial_{x_i} J_i(x_i, \mathbf{x}_{-i}) = \nabla_{x_i} f_i(x_i, \mathbf{x}_{-i}) + \partial g_i(x_i)$. Therefore, Under Assumption 3, the Karush–Kuhn–Tucker

(KKT) theorem ensures the existence of a pair $(x_i^*, \lambda_i^*) \in \Omega_i \times \mathbb{R}_{\geq 0}^m$, such that

$$\forall i \in \mathcal{I} : \begin{cases} \mathbf{0}_{n_i} \in \nabla_{x_i} f_i(x_i^*; \mathbf{x}_{-i}^*) + \partial g_i(x_i^*) + A_i^\top \lambda_i^* \\ \mathbf{0}_m \in \mathbb{N}_{\mathbb{R}_{\geq 0}^m}(\lambda_i^*) - (A\mathbf{x}^* - b). \end{cases} \quad (5)$$

We conclude the section by postulating a standard assumption for GNEP's [3], and inclusion problems in general [10], concerning the monotonicity and Lipschitz continuity of the mapping that collects the partial gradients $\nabla_i f_i$.

Standing Assumption 4 (Monotonicity): The mapping

$$F(\mathbf{x}) := \text{col}(\nabla_{x_1} f_1(\mathbf{x}), \dots, \nabla_{x_N} f_N(\mathbf{x})) \quad (6)$$

is monotone on Ω , i.e., for all $\mathbf{x}, \mathbf{y} \in \Omega$, $\langle F(\mathbf{x}) - F(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0$. and $\frac{1}{\beta}$ -Lipschitz continuous, $\beta > 0$, i.e., for all $\mathbf{x}, \mathbf{y} \in \Omega$, $\|F(\mathbf{x}) - F(\mathbf{y})\| \leq \frac{1}{\beta} \|\mathbf{x} - \mathbf{y}\|$. ■

Among all possible GNEs of the game, this work focuses on the computation of a *variational GNE* (v-GNE) [3, Def. 3.10], i.e. a GNE in which all players share consensus on the dual variables:

$$\forall i \in \mathcal{I} : \begin{cases} \mathbf{0}_{n_i} \in \nabla_{x_i} f_i(x_i^*; \mathbf{x}_{-i}^*) + \partial g_i(x_i^*) + A_i^\top \lambda^* \\ \mathbf{0}_m \in \mathbb{N}_{\mathbb{R}_{\geq 0}^m}(\lambda^*) - (A\mathbf{x}^* - b). \end{cases} \quad (7)$$

III. DISTRIBUTED GENERALIZED NASH EQUILIBRIUM SEEKING VIA OPERATOR SPLITTING

In this section, we present the proposed distributed algorithms. We allow each agent to have information on his own local problem data only, i.e., J_i, Ω_i, A_i and b_i . We let each agent i control its local decision x_i , and a local copy $\lambda_i \in \mathbb{R}_{\geq 0}^m$ of dual variables, as well as a local auxiliary variable $z_i \in \mathbb{R}^m$ used to enforce consensus of the dual variables. To actually reach consensus on the dual variables, we let the agents exchange information via an undirected weighted *communication graph*, represented by its weighted adjacency matrix $\mathbf{W} = [w_{i,j}]_{i,j} \in \mathbb{R}^{N \times N}$. We assume $w_{i,j} > 0$ iff (i, j) is an edge in the communication graph. The set of neighbours of agent i in the graph is $\mathcal{N}_i^\lambda = \{j \mid w_{i,j} > 0\}$.

Standing Assumption 5 (Graph connectivity): The matrix \mathbf{W} is symmetric and irreducible. ■

Define the weighted Laplacian as $\mathbf{L} := \text{diag}\{(\mathbf{W}\mathbf{1}_N)_1, \dots, (\mathbf{W}\mathbf{1}_N)_N\} - \mathbf{W}$. It holds that $\mathbf{L}^\top = \mathbf{L}$, $\ker(\mathbf{L}) = \text{span}(\mathbf{1}_N)$ and that, given Standing Assumption 5, \mathbf{L} is positive semi-definite with real and distinct eigenvalues $0 = s_1 < s_2 \leq \dots \leq s_N$. Moreover, given the maximum degree of the graph \mathcal{G}^λ , $\Delta := \max_{i \in \mathcal{I}} (\mathbf{W}\mathbf{1}_N)_i$, it holds that $\Delta \leq s_N \leq 2\Delta$. Denoting by $\kappa := |\mathbf{L}|$, it holds that $\kappa \leq 2\Delta$. We define the tensorized Laplacian as the matrix $\bar{\mathbf{L}} = \mathbf{L} \otimes I_m$. We set $\bar{b} = (b_1, \dots, b_N)^\top$, $\mathbf{x} = \text{col}(x_1, \dots, x_N)$ and similarly \mathbf{z} and $\boldsymbol{\lambda}$.

Let $\mathbf{A} = \text{diag}(A_1, \dots, A_N)$ and define

$$\begin{aligned} \mathcal{A}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) &:= \text{col}(F(\mathbf{x}), \mathbf{0}_{mN}, \bar{\mathbf{L}}\boldsymbol{\lambda} + \bar{b}), \\ \mathcal{B}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) &:= \text{col}(\mathbf{A}^\top \boldsymbol{\lambda}, \bar{\mathbf{L}}\boldsymbol{\lambda}, -\mathbf{A}\mathbf{x} - \bar{\mathbf{L}}\mathbf{z}) \end{aligned} \quad (8)$$

Let us also define the operator $\mathcal{D} := \mathcal{A} + \mathcal{B}$, and the set-valued operator

$$\mathcal{C}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = G(\mathbf{x}) \times \{\mathbf{0}_{mN}\} \times \mathbb{N}_{\mathbb{R}_{\geq 0}^m}(\boldsymbol{\lambda}) \quad (9)$$

where $G(\mathbf{x}) = \partial g_1(x_1) \times \dots \times \partial g_N(x_N)$.

Let us summarize the properties of the operators above.

Lemma 1: The following statements hold:

- (i) \mathcal{A} is maximally monotone and $L_{\mathcal{A}} = (\frac{1}{\beta} + \kappa)$ -Lipschitz continuous.
- (ii) \mathcal{B} is maximally monotone and $L_{\mathcal{B}} = (2|\mathbf{A}| + 2\kappa)$ -Lipschitz continuous.
- (iii) \mathcal{D} is maximally monotone and $L_{\mathcal{D}} = L_{\mathcal{A}} + L_{\mathcal{B}}$ -Lipschitz continuous.
- (iv) $\mathcal{C} + \mathcal{D}$ is maximally monotone.

Proof: (i) The operator \mathcal{A} is maximally monotone by [10, Prop. 20.23]. Furthermore, given $\kappa = |\mathbf{L}|$ and $\mathbf{u} := \text{col}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda})$, it holds that

$$\begin{aligned} \|\mathcal{A}\mathbf{u} - \mathcal{A}\mathbf{u}'\| &\leq \|F(\mathbf{x}) - F(\mathbf{x}')\| + \|\bar{\mathbf{L}}(\mathbf{z} - \mathbf{z}')\| \\ &\leq \left(\frac{1}{\beta} + \kappa\right) (\|\mathbf{x} - \mathbf{x}'\| + \|\mathbf{z} - \mathbf{z}'\|), \end{aligned} \quad (10)$$

showing that \mathcal{A} is $L_{\mathcal{A}} := (\frac{1}{\beta} + \kappa)$ -Lipschitz continuous.

(ii) The operator \mathcal{B} is maximally monotone by [10, Cor. 20.28]. By a computation similar to (10), it can be shown that \mathcal{B} is $L_{\mathcal{B}} = (2|\mathbf{A}| + 2\kappa)$ -Lipschitz continuous.

(iii) The operator \mathcal{D} is maximally monotone since $\text{dom}(\mathcal{B}) = \Omega$ [10, Prop. 20.23]. It is $L_{\mathcal{D}} = L_{\mathcal{A}} + L_{\mathcal{B}}$ -Lipschitz continuous because sum of Lipschitz operators.

(iv) The operator \mathcal{C} is maximally monotone by [7, Lem. 5] and $\mathcal{C} + \mathcal{D}$ is maximally monotone by [10, Prop. 20.23]. ■

The following result, which immediately follows from the definition of the inner product $\langle \cdot, \cdot \rangle_{\Phi}$, holds for monotone operators and it will be recalled later on.

Lemma 2: Let $\Phi \succ 0$ and \mathcal{T} be a monotone operator, then $\Phi^{-1}\mathcal{T}$ is monotone in the Hilbert space \mathcal{H}_{Φ} . ■

Now, given the operators \mathcal{A} , \mathcal{B} and \mathcal{C} as in (8) and (9), a simple proof shows that the zeros of the sum $\mathcal{A} + \mathcal{B} + \mathcal{C}$ are v-GNE of the game in (4). In fact, this follows verbatim from [7, Thm. 2], so we omit the detailed proof here.

Lemma 3: The set $\text{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C})$ is the set of v-GNE of the game in (4) and it is non-empty. ■

A. Forward-backward operator splitting

The aim of this section is to revisit a distributed forward-backward (FB) splitting algorithm for the distributed computation of a v-GNE, see Algorithm 1 [7]. From now on,

Algorithm 1 Preconditioned Forward Backward

Initialization: $x_i^0 \in \Omega_i$, $\lambda_i^0 \in \mathbb{R}_{\geq 0}^m$, and $z_i^0 \in \mathbb{R}^m$.

Iteration k : Agent i

(1) Receives x_j^k for $j \in \mathcal{N}_i^J$, λ_j^k for $j \in \mathcal{N}_i^{\lambda}$, then updates

$$\begin{aligned} x_i^{k+1} &= \text{prox}_{g_i}^{\rho_i}[x_i^k - \rho_i(\nabla_{x_i} f_i(x_i^k, \mathbf{x}_{-i}^k) - A_i^T \lambda_i^k)] \\ z_i^{k+1} &= z_i^k + \sigma_i \sum_{j \in \mathcal{N}_i^{\lambda}} w_{i,j} (\lambda_i^k - \lambda_j^k) \end{aligned}$$

(2) Receives z_j^{k+1} for $j \in \mathcal{N}_i^{\lambda}$, then updates

$$\begin{aligned} \lambda_i^{k+1} &= \text{proj}_{\mathbb{R}_{\geq 0}^m} \{ \lambda_i^k - \tau_i [A_i(2x_i^{k+1} - x_i^k) - b_i \\ &\quad + \sum_{j \in \mathcal{N}_i^{\lambda}} w_{i,j} [2(z_i^{k+1} - z_{j,k+1}) - (z_i^k - z_j^k)] \\ &\quad + \sum_{j \in \mathcal{N}_i^{\lambda}} w_{i,j} (\lambda_i^k - \lambda_j^k)] \} \end{aligned}$$

the triplet $\mathbf{u} := \text{col}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda})$ defines the state variable of a distributed algorithm. Given the state at iteration k , $\mathbf{u}^k = (\mathbf{x}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$, the FB algorithm can be written as fixed-point iteration of the form $\mathbf{u}^{k+1} = T_{\text{FB}} \mathbf{u}^k$, where

$$T_{\text{FB}} := J_{\Phi_{\text{FB}}^{-1}(\mathcal{C} + \mathcal{B})} (\text{Id} - \Phi_{\text{FB}}^{-1} \mathcal{A}). \quad (11)$$

and Φ_{FB} is the preconditioning matrix defined as

$$\Phi_{\text{FB}} := \begin{bmatrix} \rho^{-1} & 0 & -\mathbf{A}^{\top} \\ 0 & \sigma^{-1} & -\bar{\mathbf{L}} \\ -\mathbf{A} & -\bar{\mathbf{L}} & \tau^{-1} \end{bmatrix}. \quad (12)$$

The matrices $\rho = \text{diag}\{\rho_1 I_{n_1}, \dots, \rho_N I_{n_N}\}$, σ and τ (defined analogously) collect the step sizes of the primal, the auxiliary and the dual updates, respectively. By choosing the step sizes appropriately, the preconditioning matrix Φ can be made positive definite [6]. The FB algorithm is known to converge to a zero of a monotone inclusion $0 \in \mathcal{A} + \mathcal{B} + \mathcal{C}$ when the operators are maximally monotone and the single-valued operator $\Phi_{\text{FB}}^{-1} \mathcal{A}$ is cocoercive [10, Thm. 26.14]. Thus, the pseudo-gradient mapping F in (6) should satisfy the following assumption.

Assumption 1 (Strong monotonicity): For all $\mathbf{x}, \mathbf{x}' \in \Omega$, $\langle F(\mathbf{x}) - F(\mathbf{x}'), \mathbf{x} - \mathbf{x}' \rangle \geq \eta \|\mathbf{x} - \mathbf{x}'\|^2$, for some $\eta > 0$. ■

To ensure the cocoercivity condition, we refer to the following result.

Lemma 4: [7, Lem. 5 and Lem. 7] Let $\Phi \succ 0$ and F as in (6) satisfy Assumption 1. Then, the following hold:

- (i) \mathcal{A} is θ -cocoercive with $\theta \leq \min\{1/2\Delta, \eta\beta^2\}$.
- (ii) $\Phi^{-1} \mathcal{A}$ is $\alpha\theta$ -cocoercive with $\alpha = 1/|\Phi^{-1}|$. ■

We recall that convergence to a v-GNE has been demonstrated in [7, Th. 3], if the step sizes in (12) are chosen small enough [7, Lem. 6].

B. Forward-backward-forward splitting

In this section, we propose our distributed forward-backward-forward (FBF) scheme, Algorithm 2.

In compact form, the FBF algorithm generates two sequences $(\mathbf{u}^k, \mathbf{v}^k)_{k \geq 0}$ as follows:

$$\begin{aligned} \mathbf{u}^k &= J_{\Psi^{-1}\mathcal{C}} (\mathbf{v}^k - \Psi^{-1} \mathcal{D} \mathbf{v}^k) \\ \mathbf{v}^{k+1} &= \mathbf{u}^k + \Psi^{-1} (\mathcal{D} \mathbf{v}^k - \mathcal{D} \mathbf{u}^k). \end{aligned} \quad (13)$$

In (13), Ψ is the block-diagonal matrix of the step sizes:

$$\Psi = \text{diag}(\rho^{-1}, \sigma^{-1}, \tau^{-1}), \quad (14)$$

We recall that $\mathcal{D} = \mathcal{A} + \mathcal{B}$ is single-valued, maximally monotone and Lipschitz continuous by Lemma 1. Each iteration differs from the scheme in (11) by one additional forward step and the fact that the resolvent is now defined in terms of the operator \mathcal{C} only. Writing the coordinates as $\mathbf{u}^k = (\tilde{\mathbf{x}}^k, \tilde{\mathbf{z}}^k, \tilde{\boldsymbol{\lambda}}^k)$ and $\mathbf{v}^k = (\mathbf{x}^k, \mathbf{z}^k, \boldsymbol{\lambda}^k)$, the updates are explicitly given in Algorithm 2.

FBF operates on the splitting $\mathcal{C} + \mathcal{D}$ and it can be compactly written as the fixed-point iteration $\mathbf{v}^{k+1} = T_{\text{FBF}} \mathbf{v}^k$, where the mapping T_{FBF} is defined as

$$T_{\text{FBF}} := \Psi^{-1} \mathcal{D} + (\text{Id} - \Psi^{-1} \mathcal{D}) \circ J_{\Psi^{-1} \mathcal{C}} \circ (\text{Id} - \Psi^{-1} \mathcal{D}). \quad (15)$$

To ensure convergence of Algorithm 2 to a v-GNE of the game in (4), we need the next assumption.

Assumption 2: $|\Psi^{-1}| < 1/L_{\mathcal{D}}$, with Ψ as in (14) and $L_{\mathcal{D}}$ being the Lipschitz constant of \mathcal{D} as in Lemma 1. ■

Algorithm 2 Distributed Forward Backward Forward

Initialization: $x_i^0 \in \Omega_i, \lambda_i^0 \in \mathbb{R}_{\geq 0}^m$, and $z_i^0 \in \mathbb{R}^m$.Iteration k : Agent i (1) Receives x_j^k for $j \in \mathcal{N}_i^J$, λ_j^k and $z_{j,k}$ for $j \in \mathcal{N}_i^\lambda$ then updates

$$\begin{aligned}\tilde{x}_i^k &= \text{prox}_{g_i}^{\rho_i}[x_i^k - \rho_i(\nabla_{x_i} f_i(x_i^k, \mathbf{x}_{-i}^k) - A_i^T \lambda_i^k)] \\ \tilde{z}_i^k &= z_i^k + \sigma_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} (\lambda_i^k - \lambda_j^k) \\ \tilde{\lambda}_i^k &= \text{proj}_{\mathbb{R}_{\geq 0}^m} \{ \lambda_i^k - \tau_i (A_i x_i^k - b_i) \\ &\quad + \tau \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(z_i^k - z_j^k) - (\lambda_i^k - \lambda_j^k)] \}\end{aligned}$$

(2) Receives \tilde{x}_j^k for $j \in \mathcal{N}_i^J$, $\tilde{\lambda}_j^k$ and \tilde{z}_j^k for $j \in \mathcal{N}_i^\lambda$ then updates

$$\begin{aligned}x_i^{k+1} &= \tilde{x}_i^k - \rho_i(\nabla_{x_i} f_i(x_i^k, \mathbf{x}_{-i}^k) - \nabla_{\tilde{x}_i} f_i(\tilde{x}_i^k, \tilde{\mathbf{x}}_{-i}^k) \\ &\quad - \rho_i A_i^T (\lambda_i^k - \tilde{\lambda}_i^k)) \\ z_i^{k+1} &= \tilde{z}_i^k + \sigma_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(\lambda_i^k - \lambda_j^k) - (\tilde{\lambda}_i^k - \tilde{\lambda}_j^k)] \\ \lambda_i^{k+1} &= \tilde{\lambda}_i^k + \tau_i A_i (\tilde{x}_i^k - x_i^k) \\ &\quad - \tau_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(z_i^k - z_j^k) - (\tilde{z}_i^k - \tilde{z}_j^k)] \\ &\quad + \tau_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(\lambda_{i,k} - \lambda_j^k) - (\tilde{\lambda}_i^k - \tilde{\lambda}_j^k)]\end{aligned}$$

Theorem 1: Let Assumption 2 hold. The sequence $(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ generated by Algorithm 2 converges to $\text{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C})$, thus the primal variable converges to a v-GNE of the game in (4).

Proof: The fixed-point iteration with T_{FBF} as in (15) can be derived from (13) by substituting \mathbf{u}_k . Therefore, the sequence $(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ generated by Algorithm 2 converges to a v-GNE by [10, Th.26.17] and [11, Th.3.4] since $\Psi^{-1}\mathcal{A}$ is monotone by Lemma 2 and $\mathcal{A} + \mathcal{B} + \mathcal{C}$ is maximally monotone by Lemma 1. See Appendix III-B for details. ■

We emphasize that Algorithm 2 does not require strong monotonicity (Assumption 1) of the pseudo-gradient mapping F in (6). Moreover, we note that the FBF algorithm requires two evaluations of the individual gradients, which requires computing the operator \mathcal{D} twice per iteration. At the level of the individual agents, this means that we need two communication rounds per iteration in order to exchange the necessary information. Compared with the FB algorithm, the non-strong monotonicity assumption comes at the price of increased communications at each iteration.

C. Forward-backward-half forward splitting

Should the strong monotonicity condition (Assumption 1) be satisfied, an alternative to the FB is the *forward-backward-half-forward* (FBHF) operator splitting, developed in [12]. Thus, our second GNE seeking algorithm is a distributed FBHF, described in Algorithm 3.

In compact form, the FBHF algorithm reads as

$$\begin{aligned}\mathbf{u}^k &= \text{J}_{\Psi^{-1}\mathcal{C}}(\mathbf{v}^k - \Psi^{-1}(\mathcal{A} + \mathcal{B})\mathbf{v}^k) \\ \mathbf{v}^{k+1} &= \mathbf{u}^k + \Psi^{-1}(\mathcal{B}\mathbf{v}^k - \mathcal{B}\mathbf{u}^k).\end{aligned}\quad (16)$$

Algorithm 3 Distributed Forward Backward Half Forward

Initialization: $x_i^0 \in \Omega_i, \lambda_i^0 \in \mathbb{R}_{\geq 0}^m$, and $z_i^0 \in \mathbb{R}^m$.Iteration k : Agent i (1) Receives x_j^k for $j \in \mathcal{N}_i^J$, λ_j^k and $z_{j,k}$ for $j \in \mathcal{N}_i^\lambda$ then updates

$$\begin{aligned}\tilde{x}_i^k &= \text{prox}_{g_i}^{\rho_i}[x_i^k - \rho_i(\nabla_{x_i} f_i(x_i^k, \mathbf{x}_{-i}^k) - A_i^T \lambda_i^k)] \\ \tilde{z}_i^k &= z_i^k + \sigma_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} (\lambda_i^k - \lambda_j^k) \\ \tilde{\lambda}_i^k &= \text{proj}_{\mathbb{R}_{\geq 0}^m} \{ \lambda_i^k - \tau_i (A_i x_i^k - b_i) \\ &\quad + \tau \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(z_i^k - z_j^k) - (\lambda_i^k - \lambda_j^k)] \}\end{aligned}$$

(2) Receives $\tilde{\lambda}_j^k$ and $\tilde{z}_{j,k}$ for $j \in \mathcal{N}_i^\lambda$ then updates

$$\begin{aligned}x_i^{k+1} &= \tilde{x}_i^k + \rho_i A_i^T (\lambda_i^k - \tilde{\lambda}_{i,k}) \\ z_i^{k+1} &= \tilde{z}_i^k + \sigma_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(\lambda_i^k - \lambda_j^k) - (\tilde{\lambda}_i^k - \tilde{\lambda}_j^k)] \\ \lambda_i^{k+1} &= \tilde{\lambda}_i^k + \tau_i A_i (\tilde{x}_i^k - x_i^k) \\ &\quad - \tau_i \sum_{j \in \mathcal{N}_i^\lambda} w_{i,j} [(z_i^k - z_j^k) - (\tilde{z}_i^k - \tilde{z}_j^k)]\end{aligned}$$

We note that the iterates of FBHF are similar to those of the FBF, but the second forward step requires the operator \mathcal{B} only. More simply, we can write the FBHF as the fixed-point iteration $\mathbf{v}^{k+1} = T_{\text{FBHF}}\mathbf{v}^k$, where

$$T_{\text{FBHF}} = (\text{Id} - \Psi^{-1}\mathcal{B}) \circ \text{J}_{\Psi^{-1}\mathcal{C}} \circ (\text{Id} - \Psi^{-1}\mathcal{D}) + \Psi^{-1}\mathcal{B}. \quad (17)$$

Also in this case, we have a bound on the step sizes.

Assumption 3: $|\Psi^{-1}| \leq \min\{2\theta_{\mathcal{A}}, 1/L_{\mathcal{B}}\}$, with $\theta_{\mathcal{A}}$ as in Lemma 4 and $L_{\mathcal{B}}$ as in Lemma 1. ■

We note that in Assumption 3, the step sizes in Ψ can be chosen larger compared to those in Assumption 2, since the upper bound is related to the Lipschitz constant of the operator \mathcal{B} , not of $L_{\mathcal{D}} = L_{\mathcal{A}} + L_{\mathcal{B}}$ as for the FBF (Assumption 2). A similar comparison can be done with respect to the FB algorithm. Intuitively, larger step sizes should be beneficial in term of convergence speed.

We can now establish our convergence result for the FBHF algorithm.

Theorem 2: Let Assumptions 1 and 3 hold. The sequence $(\mathbf{x}^k, \boldsymbol{\lambda}^k)$ generated by Algorithm 3 converges to $\text{zer}(\mathcal{A} + \mathcal{B} + \mathcal{C})$, thus the primal variable converges to a v-GNE of the game in (4). ■

Proof: Algorithm 3 is the fixed point iteration in (17) whose convergence is guaranteed by [12, Th. 2.3] under Assumption 3 because $\Psi^{-1}\mathcal{A}$ is cocoercive by Lemma 4. See Appendix VI-B for details. ■

IV. CASE STUDY AND NUMERICAL SIMULATIONS

We consider a networked Cournot game with market capacity constraints [7]. As a numerical setting, we use a set of 20 companies and 7 markets, similarly to [7]. Each company i has a local constraint $x_i \in (0, \delta_i)$ where each component of δ_i is randomly drawn from $[1, 1.5]$. The maximal capacity of each market j is b_j , randomly drawn from $[0.5, 1]$. The local cost function of company i is $c_i(x_i) = \pi_i \sum_{j=1}^{n_i} ([x_i]_j)^2 + r^\top x_i$, where $[x_i]_j$ indicates

the j component of x_i . For all $i \in \mathcal{I}$, π_i is randomly drawn from [1, 8], and the components of r_i are randomly drawn from [0.1, 0.6]. Notice that $c_i(x_i)$ is strongly convex with Lipschitz continuous gradient. The price is taken as a linear function $P = \bar{P} - DAx$ where each component of $\bar{P} = \text{col}(\bar{P}_1, \dots, \bar{P}_7)$ is randomly drawn from [2, 4] while the entries of $D = \text{diag}(d_1, \dots, d_7)$ are randomly drawn from [0.5, 1]. Recall that the cost function of company i is influenced by the variables of the agents selling in the same market. Such informations can be retrieved from [7, Fig. 1]. Since $c_i(x_i)$ is strongly convex with Lipschitz continuous gradient and the prices are linear, the pseudo gradient of f_i is strongly monotone. The communication graph \mathcal{G}^λ for the dual variables is a cycle graph with the addition of the edges (2, 15) and (6, 13). As local cost functions g_i we use the indicator functions. In this way, the proximal step is a projection on the local constraints sets.

The aim of these simulations is to compare the proposed schemes. The step sizes are taken differently for every algorithm. In particular, we take ρ_{FB} , σ_{FB} and τ_{FB} as in [7, Lem. 6], ρ_{FBF} , σ_{FBF} and τ_{FBF} such that Assumption 2 is satisfied and ρ_{FBHF} , σ_{FBHF} and τ_{FBHF} such that Assumption 3 holds. We select them to be the maximum possible.

The initial points λ_i^0 and z_i^0 are set to 0 while the local decision variable x_i^0 is randomly taken in the feasible sets.

The plots in Fig. 1 show the performance parameter $\frac{\|x_{k+1} - x^*\|}{\|x^*\|}$, that is, the convergence to a solution x^* , and the CPU time (in seconds) used by each algorithm. We run 10 simulations, changing the parameters of the cost function to show that the result are replicable. The darker line represent the average path towards the solution.

The plot in Fig 1 shows that with suitable parameters convergence to a solution is faster with the FBF algorithm which, however, is computationally more expensive than the FB and FBHF algorithms.

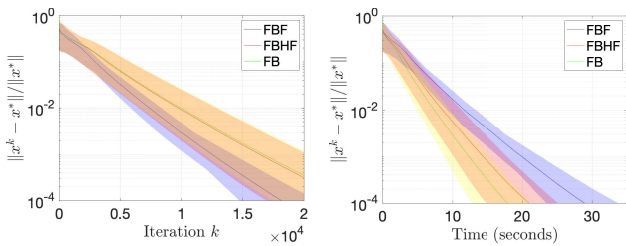


Fig. 1. Relative distance from v-GNE (left) and cumulative CPU time (right).

V. CONCLUSION

The FBF and the FBHF splitting methods generate distributed equilibrium seeking algorithms for solving generalized Nash equilibrium problems. Compared to the FB, the FBF has the advantage to converge under the non-strong monotonicity assumption. This comes at the price of increased communications between the agents. If strong monotonicity holds, an alternative to the FBF is the FBHF that, in our numerical experience is less computationally expensive than the FBF.

VI. APPENDIX

A. Convergence of the forward-backward-forward

We show the convergence proof for the FBF. From now on, $\mathbf{H} = \mathbb{R}^n \times \mathbb{R}^{mN} \times \mathbb{R}^{mN}$ and $\text{fix}(T) = \{x \in \mathbf{H} : Tx = x\}$.

Proposition 1: If Assumption 2 holds, $\text{fix}(T_{\text{FBF}}) = \mathcal{Z}$.

Proof: We first show that $\mathcal{Z} \subseteq \text{fix}(T_{\text{FBF}})$. Let $u^* \in \mathcal{Z}$:

$$\begin{aligned} 0 \in \mathcal{C}u^* + \mathcal{D}u^* &\Leftrightarrow -\mathcal{D}u^* \in \mathcal{C}u^* \\ &\Leftrightarrow u^* = J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*) \\ &\Leftrightarrow \Psi^{-1}\mathcal{D}u^* = \Psi^{-1}\mathcal{D}J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*) \\ &\Leftrightarrow u^* = T_{\text{FBF}}u^*. \end{aligned}$$

Conversely, let $u^* \in \text{fix}(T_{\text{FBF}})$. Then $u^* - J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*) = \Psi^{-1}\mathcal{D}u^* - \Psi^{-1}\mathcal{D}J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*)$ and

$$\begin{aligned} \|u^* - J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*)\| &\leq \\ &\leq \alpha^{-1}\|\mathcal{D}u^* - \mathcal{D}J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*)\| \\ &\leq \frac{L}{\alpha}\|u^* - J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*)\|. \end{aligned}$$

Hence, $u^* = J_{\Psi^{-1}\mathcal{C}}(u^* - \Psi^{-1}\mathcal{D}u^*)$. \blacksquare

Proposition 2: For all $u^* \in \text{fix}(T_{\text{FBF}})$ and $v \in \mathbf{H}$, there exists $\varepsilon \geq 0$ such that

$$\|T_{\text{FBF}}v - u^*\|_{\Psi}^2 = \|v - u^*\|_{\Psi}^2 - (1 - (L/\alpha)^2) \|v - u^*\|_{\Psi}^2 - 2\varepsilon. \quad (18)$$

Proof: Let $u^* \in \text{fix}(T_{\text{FBF}})$ and $u = J_{\Psi^{-1}\mathcal{C}}(v - \Psi^{-1}\mathcal{D}v)$, $v^+ = T_{\text{FBF}}v$, for $v \in \mathbf{H}$ arbitrary. Then,

$$\begin{aligned} \|v - u^*\|_{\Psi}^2 &= \|v - u + u - v^+ + v^+ - u^*\|_{\Psi}^2 \\ &= \|v - u\|_{\Psi}^2 + \|u - v^+\|_{\Psi}^2 + \|v^+ - u^*\|_{\Psi}^2 \\ &\quad + 2\langle v - u, u - u^* \rangle_{\Psi} + 2\langle u - v^+, v^+ - u^* \rangle_{\Psi}. \end{aligned}$$

Since, $2\langle u - v^+, v^+ - u^* \rangle_{\Psi} = 2\langle u - v^+, v^+ - u \rangle_{\Psi} + 2\langle u - v^+, u - u^* \rangle_{\Psi} = -2\|u - v^+\|_{\Psi}^2 + 2\langle u - v^+, u - u^* \rangle_{\Psi}$. This gives $\|v - u^*\|_{\Psi}^2 = \|v - u\|_{\Psi}^2 - \|u - v^+\|_{\Psi}^2 + \|v^+ - u^*\|_{\Psi}^2 + 2\langle u - v^+, u - u^* \rangle_{\Psi}$. By definition of the updates, we have for $\bar{v} \equiv Bv$, $\bar{u} \equiv Bu$, $\hat{v} \in \mathcal{C}u$, the identities $u + \Psi^{-1}\hat{v} = v - \Psi^{-1}\bar{v}$ and $v^+ = u + \Psi^{-1}(\bar{v} - \bar{u})$. Furthermore, since $0 \in \mathcal{D}u^* + \mathcal{C}u^*$, there exists $\hat{v}^* \in \mathcal{C}u^*$ and $\bar{u}^* \equiv \mathcal{D}u^*$ such that $0 = \bar{u}^* + \hat{v}^*$. It follows that $v - v^+ = v - u - \Psi^{-1}(\bar{v} - \bar{u}) = \Psi^{-1}(\hat{v} + \bar{u})$.

Hence,

$$\begin{aligned} \|v - u^*\|_{\Psi}^2 &= \|v - u\|_{\Psi}^2 - \|u - v^+\|_{\Psi}^2 + \\ &\quad + \|v^+ - u^*\|_{\Psi}^2 + 2\langle u - u^*, \hat{v} + \bar{u} \rangle \\ &= \|v - u\|_{\Psi}^2 - \|u - v^+\|_{\Psi}^2 + \|v^+ - u^*\|_{\Psi}^2 \\ &\quad + 2\langle \hat{v} - \hat{v}^* - \bar{u}^* + \bar{u}, u - u^*, u - u^* \rangle. \end{aligned}$$

Since $(u, \hat{v}), (u^*, \hat{v}^*) \in \text{gph}(C)$, $(u^*, \bar{u}^*), (u, \bar{u}) \in \text{gph}(B)$, it follows from the monotonicity that $\varepsilon := \langle \hat{v} - \hat{v}^* - \bar{u}^* + \bar{u}, u - u^*, u - u^* \rangle \geq 0$. Finally, observe that $u - v^+ = \Psi^{-1}(\mathcal{D}u - \mathcal{D}v)$, and that

$$\begin{aligned} \|\Psi^{-1}(\mathcal{D}u - \mathcal{D}v)\|_{\Psi}^2 &= \langle \Psi^{-1}(\mathcal{D}u - \mathcal{D}v), \mathcal{D}u - \mathcal{D}v \rangle \\ &\leq \lambda_{\max}(\Psi^{-1})\|\mathcal{D}u - \mathcal{D}v\|^2 \leq L^2\lambda_{\max}(\Psi^{-1})\|u - v\|^2 \\ &\leq L^2 \frac{\lambda_{\max}(\Psi^{-1})}{\lambda_{\min}(\Psi)} \|u - v\|_{\Psi}^2. \end{aligned}$$

Since $\alpha = 1/\lambda_{\max}(\Psi^{-1}) = \lambda_{\min}(\Psi)$, it follows from the Lipschitz continuity of the operator B that $\|u - v^+\|_{\Psi}^2 \leq (L/\alpha)^2\|u - v\|_{\Psi}^2$ and the statement is proven. \blacksquare

Corollary 1: If $L/\alpha < 1$, the map $T_{\text{FBF}} : \mathbf{H} \rightarrow \mathbf{H}$ is quasicontractive in the Hilbert space $(\mathbf{H}, \langle \cdot, \cdot \rangle_{\Psi})$, i.e.

$$\forall v \in \mathbf{H} \forall u^* \in \text{fix}(T_{\text{FBF}}) \|T_{\text{FBF}}v - u^*\|_{\Psi} \leq \|v - u^*\|_{\Psi}.$$

Proposition 3: If Assumption 2 holds, the sequence generated by the FBF algorithm, $(v^k)_{k \geq 0}$, is bounded in norm, and all its accumulation points are elements in \mathcal{Z} .

Proof: Form (18) we deduce that $(v^k)_{k \geq 0}$ is Fejér monotone with respect to $\text{fix}(T_{\text{FBF}}) = \mathcal{Z}$. Therefore, it is bounded norm. It remains to show that all accumulation points are in \mathcal{Z} . By an obvious abuse of notation, let $(v^k)_{k \geq 0}$ denote a converging subsequence with limit u^* . From (18) it follows $\|u^k - v^k\|_{\Psi} \rightarrow 0$, and hence $\|u^k - v^k\| \rightarrow 0$ as $k \rightarrow \infty$. By continuity, it therefore follows as well $\|\mathcal{D}u^k - \mathcal{D}v^k\| \rightarrow 0$ as $k \rightarrow \infty$. Since $u^k = J_{\Psi^{-1}\mathcal{C}}(v^k - \Psi^{-1}\mathcal{D}v^k)$, it follows that $w^k := \Psi(v^k - u^k) + \mathcal{D}u^k - \mathcal{D}v^k \in \mathcal{D}u^k + \mathcal{C}u^k$. Since $w^k \rightarrow 0$ and the operator $\mathcal{C} + \mathcal{D}$ is maximally monotone by Lemma 1 and has a closed graph [11, Lem. 3.2], we conclude $0 \in \mathcal{D}u^* + \mathcal{C}u^*$. Hence, $u^* \in \mathcal{Z}$. ■

B. Convergence of the forward-backward-half-forward

We here provide the convergence proof for the FBHF.

Proposition 4: If Assumption 3 holds, the sequence generated by the FBHF algorithm converges to \mathcal{Z} .

Since, $w - u \in \Psi^{-1}\mathcal{C}u$, it follows that $(u, w - u) \in \text{gph}(\Psi^{-1}\mathcal{C})$. Additionally, $0 \in \mathcal{D}u^* + \mathcal{C}u^*$, implying that $(u^*, -\Psi^{-1}\mathcal{D}u^*) \in \text{gph}(\Psi^{-1}\mathcal{C})$. Monotonicity of the involved operators, implies that $\langle u - u^*, w - u - \Psi^{-1}\mathcal{D}u^* \rangle_{\Psi} \leq 0$, and $\langle u - u^*, \Psi^{-1}(\mathcal{B}u^* - \mathcal{B}u) \rangle_{\Psi} \leq 0$. Using these two inequalities, we see

$$\begin{aligned} \langle u - u^*, u - w - \Psi^{-1}\mathcal{B}u \rangle_{\Psi} &= \langle u - u^*, \Psi^{-1}\mathcal{A}u^* \rangle_{\Psi} \\ &+ \langle u - u^*, u - w - \Psi^{-1}\mathcal{D}u^* \rangle_{\Psi} \\ &+ \langle u - u^*, \Psi^{-1}(\mathcal{D}u^* - \mathcal{B}u) \rangle_{\Psi} \leq \langle u - u^*, \Psi^{-1}\mathcal{A}u^* \rangle_{\Psi} \end{aligned}$$

Therefore,

$$\begin{aligned} 2\langle u - u^*, \Psi^{-1}(\mathcal{B}v - \mathcal{B}u) \rangle_{\Psi} &= \\ 2\langle u - u^*, \Psi^{-1}\mathcal{B}v + w - u \rangle_{\Psi} &+ 2\langle u - u^*, u - w - \Psi^{-1}\mathcal{B}u \rangle_{\Psi} \\ &\leq 2\langle u - u^*, \Psi^{-1}\mathcal{B}v + w - u \rangle_{\Psi} + 2\langle u - u^*, \Psi^{-1}\mathcal{A}u^* \rangle_{\Psi} \\ &= 2\langle u - u^*, \Psi^{-1}\mathcal{D}v + w - u \rangle_{\Psi} \\ &+ 2\langle u - u^*, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi} \\ &= 2\langle u - u^*, v - u \rangle_{\Psi} + 2\langle u - u^*, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi}, \end{aligned} \quad (19)$$

where in the last equality we have used the identity $w = v - \Psi^{-1}\mathcal{D}v$. Using the cosine formula, (19) becomes

$$\begin{aligned} 2\langle u - u^*, \Psi^{-1}(\mathcal{B}v - \mathcal{B}u) \rangle_{\Psi} &\leq \|v - u^*\|_{\Psi}^2 - \|u - u^*\|_{\Psi}^2 \\ &- \|v - u\|_{\Psi}^2 + 2\langle u - u^*, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi}. \end{aligned} \quad (20)$$

The cocoercivity of $\Psi^{-1}\mathcal{A}$ in $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\Psi})$ gives for all $\varepsilon > 0$

$$\begin{aligned} 2\langle u - u^*, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi} &= \\ 2\langle v - u^*, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi} &+ 2\langle u - v, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi} \\ &\leq -2\alpha\theta \|\Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v)\|_{\Psi}^2 + 2\langle u - v, \Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v) \rangle_{\Psi} \\ &= -2\alpha\theta \|\Psi^{-1}(\mathcal{A}u^* - \mathcal{A}v)\|_{\Psi}^2 + \frac{1}{\varepsilon} \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &+ \varepsilon \|v - u\|_{\Psi}^2 - \varepsilon \|v - u - \frac{1}{\varepsilon}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &= \varepsilon \|v - u\|_{\Psi}^2 - (2\alpha\theta - \frac{1}{\varepsilon}) \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &- \varepsilon \|v - u - \frac{1}{\varepsilon}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2. \end{aligned}$$

Combining this estimate with (20), we see

$$\begin{aligned} 2\langle u - u^*, \Psi^{-1}(\mathcal{B}v - \mathcal{B}u) \rangle_{\Psi} &\leq \|v - u^*\|_{\Psi}^2 - \|u - u^*\|_{\Psi}^2 \\ &- \|v - u\|_{\Psi}^2 + \varepsilon \|v - u\|_{\Psi}^2 - (2\alpha\theta - \frac{1}{\varepsilon}) \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &- \varepsilon \|v - u - \frac{1}{\varepsilon}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \|v^+ - u^*\|_{\Psi}^2 &= \|u + \Psi^{-1}(\mathcal{B}v - \mathcal{B}u) - u^*\|_{\Psi}^2 \\ &= \|u - u^*\|_{\Psi}^2 + 2\langle u - u^*, \Psi^{-1}(\mathcal{B}v - \mathcal{B}u) \rangle_{\Psi} \\ &+ \|\Psi^{-1}(\mathcal{B}v - \mathcal{B}u)\|_{\Psi}^2 \\ &\leq \|u - u^*\|_{\Psi}^2 + \|\Psi^{-1}(\mathcal{B}v - \mathcal{B}u)\|_{\Psi}^2 - \|u - u^*\|_{\Psi}^2 \\ &- (2\alpha\theta - \frac{1}{\varepsilon}) \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 + \|v - u^*\|_{\Psi}^2 - \|v - u\|_{\Psi}^2 \\ &+ \varepsilon \|v - u\|_{\Psi}^2 - \varepsilon \|v - u - \frac{1}{\varepsilon}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2. \end{aligned}$$

Since, $\|\Psi^{-1}(\mathcal{B}v - \mathcal{B}u)\|_{\Psi}^2 \leq (L/\alpha)^2 \|v - u\|_{\Psi}^2$, the above reads as

$$\begin{aligned} \|T_{\text{FBHF}}v - u^*\|_{\Psi}^2 &\leq \|v - u^*\|_{\Psi}^2 - L^2 \left(\frac{1-\varepsilon}{L^2} - \frac{1}{\alpha^2} \right) \|v - u\|_{\Psi}^2 \\ &- \frac{1}{\alpha\varepsilon} \left(2\theta\varepsilon - \frac{1}{\alpha} \right) \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &- \varepsilon \|v - u - \frac{1}{\varepsilon}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2. \end{aligned}$$

In order to choose the largest interval for $1/\alpha$ ensuring that the second and third terms are negative, we set $\chi \leq \min\{2\theta, 1/L\}$. Then,

$$\begin{aligned} \|T_{\text{FBHF}}v - u^*\|_{\Psi}^2 &\leq \|v - u^*\|_{\Psi}^2 - L^2 \left(\chi^2 - \frac{1}{\alpha^2} \right) \|v - u\|_{\Psi}^2 \\ &- \frac{2\theta}{\alpha\chi} \left(\chi - \frac{1}{\alpha} \right) \|\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2 \\ &- \frac{\chi}{2\theta} \|v - u - \frac{2\theta}{\chi}\Psi^{-1}(\mathcal{A}v - \mathcal{A}u^*)\|_{\Psi}^2. \end{aligned}$$

From here, we obtain convergence of the sequence $(v^k)_{k \geq 0}$ as a consequence of [12, Thm. 2.3] for $1/\alpha \in (0, \chi)$.

REFERENCES

- [1] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica: Journal of the Econometric Society*, pp. 520–534, 1965.
- [2] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2007.
- [3] F. Facchinei and C. Kanzow, "Generalized Nash equilibrium problems," *Annals of Operations Research*, vol. 175, no. 1, pp. 177–211, 2010.
- [4] L. Pavel, "An extension of duality to a game-theoretic framework," *Automatica*, vol. 43, no. 2, pp. 226–237, 2007.
- [5] A. A. Kulkarni and U. V. Shanbhag, "On the variational equilibrium as a refinement of the generalized Nash equilibrium," *Automatica*, vol. 48, no. 1, pp. 45–55, 2012.
- [6] G. Belgioioso and S. Grammatico, "Projected-gradient algorithms for generalized equilibrium seeking in aggregative games are pre-conditioned forward-backward methods," in *2018 European Control Conference (ECC)*. IEEE, 2018, pp. 2188–2193.
- [7] P. Yi and L. Pavel, "An operator splitting approach for distributed generalized Nash equilibria computation," *Automatica*, vol. 102, pp. 111–121, 2019.
- [8] G. Belgioioso and S. Grammatico, "Semi-decentralized nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions," *IEEE control systems letters*, vol. 1, no. 2, pp. 400–405, 2017.
- [9] G. Debreu, "A social equilibrium existence theorem," *Proceedings of the National Academy of Sciences*, vol. 38, no. 10, pp. 886–893, 1952.
- [10] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer - CMS Books in Mathematics, 2016.
- [11] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 431–446, 2000.
- [12] L. Briceño Arias and D. Davis, "Forward-backward-half forward algorithm for solving monotone inclusions," *SIAM Journal on Optimization*, vol. 28, no. 4, pp. 2839–2871, 2018.

- [13] S. Grammatico, "Dynamic control of agents playing aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4537–4548, 2017.
- [14] F. Clarke, Y. S. Ledyaev, R. Stern, and P. Wolenski, *Nonsmooth analysis and Control theory*. Springer Berlin Heidelberg, 1998.