

# ADMM-type Methods for Generalized Nash Equilibrium Problems in Hilbert Spaces\*

Eike Börgens<sup>†</sup>

Christian Kanzow<sup>†</sup>

August 28, 2019

## Abstract

We consider the generalized Nash equilibrium problem (GNEP) with  $N$  players in a Hilbert space setting. The joint constraints are eliminated by an augmented Lagrangian-type approach, leading to an ADMM (alternating direction method of multipliers) algorithm. In contrast to standard optimization problems, however, the direct extension of ADMM to GNEPs is not necessarily convergent even for  $N = 2$  players. We therefore use a regularized version of ADMM and present a global convergence result for  $N \geq 2$  players under a partial strong monotonicity and a partial Lipschitz condition. Furthermore, also different from the optimization context, it turns out that the corresponding regularization parameters have to be sufficiently large in order to guarantee global convergence. We therefore also discuss a second ADMM-type method with an adaptive choice of the regularization parameters, with the aim of keeping the regularization parameters smaller and, hence, getting faster convergence. Numerical results are presented for some examples arising in infinite-dimensional Hilbert spaces.

## 1 Introduction

We consider the generalized Nash equilibrium problem (GNEP) with  $N$  players  $\nu$ , where player  $\nu$ -th optimization problem is given by

$$\min_{x_\nu \in \mathcal{X}_\nu} \theta_\nu(x_\nu, x_{-\nu}) + \varphi_\nu(x_\nu) \quad \text{s.t.} \quad \sum_{\nu=1}^N A_\nu x_\nu = b \quad (1)$$

for all  $\nu = 1, \dots, N$ . Here,  $\mathcal{H}_\nu$  and  $\mathcal{K}$  are given Hilbert spaces,  $\varphi_\nu : \mathcal{H}_\nu \rightarrow \mathbb{R}$  are proper, convex, and lower semi-continuous functions,  $\theta_\nu : \mathcal{H}_1 \times \dots \times \mathcal{H}_N \rightarrow \mathbb{R}$  are continuously

---

\*This research was supported by the German Research Foundation (DFG) within the priority program “Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization” (SPP 1962) under grant number KA 1296/24-1.

<sup>†</sup>University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; {eike.boergens,kanzow}@mathematik.uni-wuerzburg.de.

Fréchet-differentiable with  $\theta_\nu(\cdot, x_{-\nu})$  being convex for any fixed  $x_{-\nu}$ ,  $\mathcal{X}_\nu \subset \mathcal{H}_\nu$  are nonempty, closed, and convex sets,  $A_\nu \in \mathcal{L}(\mathcal{H}_\nu, \mathcal{K})$  and  $b \in \mathcal{K}$ . Following standard notation in Nash games, we write  $x = (x_\nu, x_{-\nu})$ , where  $x_{-\nu}$  subsumes all the remaining block components  $x_i$  with  $i \neq \nu$ . This notation is used to emphasize the particular role played by the block component  $x_\nu$  within the entire vector  $x$  and does not mean that the components of  $x$  are re-ordered. In particular, we therefore have  $x = (x_\nu, x_{-\nu}) = (x_1, \dots, x_N)$  and, similarly,  $(y_\nu, x_{-\nu}) = (x_1, \dots, x_{\nu-1}, y_\nu, x_{\nu+1}, \dots, x_N)$ .

We assume that the generalized Nash equilibrium problem (1) has a nonempty feasible set. Since we have explicit constraints  $\mathcal{X}_\nu$ , there is no loss of generality to have  $\theta_\nu$  and  $\varphi_\nu$  real-valued for all  $\nu = 1, \dots, N$ . Moreover, we do not assume the operators  $A_\nu$  to be injective or surjective, a condition that is often used for splitting-type methods in the finite-dimensional context, where the matrices  $A_\nu$  are assumed to have full rank.

For the sake of notational simplicity, we use the abbreviations

$$\begin{aligned} \mathcal{H} &:= \mathcal{H}_1 \times \dots \times \mathcal{H}_N, & \mathcal{X} &:= \mathcal{X}_1 \times \dots \times \mathcal{X}_N \subseteq \mathcal{H}, & x &:= (x_1, \dots, x_N) \in \mathcal{H}, \\ Ax &:= \sum_{\nu=1}^N A_\nu x_\nu, & \varphi(x) &:= \sum_{\nu=1}^N \varphi_\nu(x_\nu), & \mathcal{F} &:= \{x \in \mathcal{X} \mid Ax = b\}. \end{aligned}$$

Canonically,  $\mathcal{H}$  becomes a Hilbert space with the scalar product  $\langle x \mid y \rangle := \langle x_1 \mid y_1 \rangle + \dots + \langle x_N \mid y_N \rangle$ ; the scalar product in the space  $\mathcal{H} \times \mathcal{K}$  is defined analogously. The symbol  $\|\cdot\|$  always denotes the norm induced by the corresponding scalar product (in  $\mathcal{H}_\nu, \mathcal{H}, \mathcal{K}$ , or  $\mathcal{H} \times \mathcal{K}$ ); it should be clear from the context which norm is used. We will omit indexing the norms in order to keep the notation simple.

There exist many approaches for the numerical solution of GNEPs, and the interested reader is referred to the survey papers [13, 15] for more details. But these survey papers consider the finite-dimensional case only. Solution methods in an infinite-dimensional Hilbert space (or Banach space) are still in its infancy. The ideas from the finite-dimensional setting can sometimes be generalized to the Hilbert space setting, like the usage of the Nikaido-Isoda function and the application of Moreau-Yosida-type methods [22, 23]. The augmented Lagrangian methods from [24, 25] may be viewed as extensions of this approach, but they have to solve (standard) Nash equilibrium problems (NEP) in each iteration. Some other methods operating in an infinite-dimensional context are [5, 10, 11, 26], but none of them is a splitting-type method and many of them are written in an optimal control context.

Splitting-type methods can, in principle, be relatively easily extended to standard NEPs. For example, the Gauss-Seidel-type best response algorithm for NEPs uses the update

$$x_\nu^{k+1} = \arg \min_{x_\nu \in \mathcal{X}_\nu} \theta_\nu(x_1^{k+1}, \dots, x_{\nu-1}^{k+1}, x_\nu, x_{\nu+1}^k, \dots, x_N^k)$$

for each player  $\nu = 1, \dots, N$ . Its extension to GNEPs is more delicate since then the feasible set of each player depends on the other players' variables. A complete convergence theory for such a method is given in [14] for the class of (finite-dimensional) potential games, but requires an inner semicontinuity assumption of the set-valued feasibility map which is often violated even for linear constraints. This indicates that Gauss-Seidel-, Jacobi- and other

splitting-type methods may work quite well for standard NEPs, but cause severe trouble for GNEPs.

A simple way to overcome this drawback is to apply a penalty or multiplier-penalty idea in order to get rid of the joint constraints. This avoids the inner semicontinuity assumption and naturally leads to an approach that is called the alternating direction method of multipliers (ADMM). The ADMM method is a very popular approach for large-scale optimization problems with a certain block structure. It has a very satisfactory convergence theory for the case of two blocks, whereas it requires suitable modifications (like a regularization technique) for more than two blocks, see [4, 6, 12, 18, 19, 20, 21] and references therein.

The aim of this paper is therefore to investigate the convergence properties of a direct extension of ADMM for the solution of GNEPs. However, the application of ADMM to GNEPs turns out to be significantly more difficult than for standard optimization problems. In particular, we will see later that the direct translation of ADMM to GNEPs may not converge even for  $N = 2$  players, in contrast to optimization problems with two blocks. We therefore introduce a regularization which is also popular in the optimization context with multiple blocks. The overall method is then well-defined with uniquely determined iterates for arbitrary positive regularization parameters. In order to prove global convergence results, however, the regularization parameters need to be sufficiently large, see also [1, 4, 9, 12, 20, 30] for related literature on regularization-type approaches.

The paper [28] also presents an approach for the solution of (finite-dimensional) GNEPs by a method which is named an ADMM technique by the authors. But their method is quite different from a direct extension of the classical ADMM algorithm to GNEPs, and convergence is shown under relatively strong assumptions. Other splitting-type methods include [3, 7]. Regarding the Jacobi-type method from [3], we refer to the discussion in our final remarks, whereas [7] is mainly a paper on standard NEPs with an extension to GNEPs, but this extension requires a projection in the full space and is therefore not a distributed approach.

This paper is organized as follows. Section 2 contains some basic definitions and preliminary results. Several properties of the GNEP are discussed in Section 3, which also describes our assumptions under which global convergence results are shown in the subsequent sections. Our first ADMM-type method with fixed regularization parameters is introduced in Section 4, which also contains the main convergence result as well as a class of examples showing that regularization is necessary already for  $N = 2$  players, and that the regularization parameters have to be sufficiently large (for any  $N \geq 2$ ). A modified ADMM-type method using an updating technique for the regularization parameter together with the corresponding convergence analysis is presented in Section 5. Some applications and numerical results are given in Section 6. We close with some final remarks in Section 7.

## 2 Preliminaries

We first recall some basic definitions from set-valued and variational analysis in Section 2.1, cf. the excellent monograph [2] for more details. We then state some preliminary results in Section 2.2

## 2.1 Basic Definitions

Given a set-valued operator  $T : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ , the *domain* of  $T$  is defined by  $\text{dom } T := \{x \in \mathcal{H} \mid T(x) \neq \emptyset\}$ , the *graph* of  $T$  is given by  $\text{graph}(T) := \{(x, u) \in \mathcal{H} \times \mathcal{H} \mid u \in T(x)\}$ . The set of all points  $x$  satisfying  $0 \in T(x)$  is denoted by  $\text{zer}(T)$ . The set-valued operator  $T$  is called *monotone* if

$$\langle u - v \mid x - y \rangle \geq 0 \quad \forall (x, u), (y, v) \in \text{graph } T,$$

and *maximally monotone* if it is monotone and its graph is not properly contained in the graph of another monotone operator. It is a well known fact that the graph of a maximal monotone operator is strong-weak-sequentially closed, cf. [2, Prop. 20.38]. A set-valued operator  $T$  is called *strongly monotone* if there is a constant  $\rho > 0$  such that

$$\langle u - v \mid x - y \rangle \geq \rho \|x - y\|^2 \quad \forall (x, u), (y, v) \in \text{graph } T.$$

Furthermore, we call  $T$  *strongly monotone with respect to  $x_{-N}$*  if there is a constant  $\rho > 0$  such that

$$\langle u - v \mid x - y \rangle \geq \rho \|x_{-N} - y_{-N}\|^2 \quad \forall (x, u), (y, v) \in \text{graph } T.$$

Note that this is a weaker property than strong monotonicity. The *subdifferential* of a convex function  $f : \mathcal{H} \rightarrow (-\infty, +\infty]$  at a point  $\bar{x}$  is defined by

$$\partial f(\bar{x}) := \{s \in \mathcal{H} \mid f(x) - f(\bar{x}) \geq \langle s \mid x - \bar{x} \rangle \quad \forall x \in \mathcal{H}\}.$$

The set-valued subdifferential mapping  $x \mapsto \partial f(x)$  is known to be maximally monotone.

We also recall that the *normal cone* is defined by

$$N_{\mathcal{X}}(x) := \{s \in \mathcal{H} \mid \langle s \mid y - x \rangle \leq 0 \quad \forall y \in \mathcal{X}\}, \quad \text{if } x \in \mathcal{X}; \quad N_{\mathcal{X}}(x) := \emptyset, \quad \text{if } x \notin \mathcal{X}.$$

The Hilbert space adjoint of a linear operator  $A : \mathcal{X} \rightarrow \mathcal{K}$  will be denoted by  $A^*$ . We call  $A \in \mathcal{L}(\mathcal{H})$  self-adjoint if  $A = A^*$ .

An important notion for sequences is introduced in the following definition.

**Definition 2.1** (Fejér monotonicity). Suppose that  $S \subset \mathcal{H}$  is a nonempty subset of a Hilbert space  $\mathcal{H}$ . Then a sequence  $(x^k)_{k \in \mathbb{N}} \subset \mathcal{H}$  is called *Fejér monotone with respect to  $S$*  if the inequality  $\|x^{k+1} - x\| \leq \|x^k - x\|$  holds for all  $k \in \mathbb{N}$  and all  $x \in S$ .

The next result states an interesting property of Fejér monotone sequences, see [2, Thm. 5.33(iv)] for a proof.

**Proposition 2.2.** *Let that  $(x^k)_{k \in \mathbb{N}}$  be a sequence in  $\mathcal{H}$  and let  $S$  be a nonempty subset of  $\mathcal{H}$  such that  $(x^k)_{k \in \mathbb{N}}$  is Fejér monotone with respect to  $S$  and such that every weak accumulation point of  $(x^k)_{k \in \mathbb{N}}$  belongs to  $S$ . Then the whole sequence converges weakly to a point  $\bar{x} \in S$ .*

In our convergence analysis we will also work with an adjusted scalar product and its induced norm. The next result shows that the notions of strong and weak convergence are the same. Its simple proof can be found, e.g., in [4, Lem. 4.5].

**Lemma 2.3.** *Consider the Hilbert space  $\mathcal{H} \times \mathcal{K}$  with the usual scalar product  $\langle \cdot \mid \cdot \rangle$ , and let  $\langle \cdot \mid \cdot \rangle_Q := \langle Q \cdot \mid \cdot \rangle$  for  $Q$  being self-adjoint and strongly monotone. Then the corresponding induced norms  $\|\cdot\|$  and  $\|\cdot\|_Q$  are equivalent. Moreover, weak convergence with respect to  $\langle \cdot \mid \cdot \rangle$  is equivalent to weak convergence with respect to  $\langle \cdot \mid \cdot \rangle_Q$ .*

## 2.2 Basic Inequalities

Here we recall and derive some basic inequalities that will be used in our subsequent analysis. To this end, we first restate the well-known Young's inequality with  $\varepsilon$ .

**Lemma 2.4.** *[Young's Inequality] Suppose that  $a, b \in \mathbb{R}$ . Then*

$$|a \cdot b| \leq \frac{\varepsilon}{2} a^2 + \frac{1}{2\varepsilon} b^2$$

*holds for every  $\varepsilon > 0$ .*

As a consequence of the previous result, we obtain the following estimate.

**Lemma 2.5.** *Let  $\alpha_1, \dots, \alpha_N \in \mathbb{R}$ . Then*

$$\left( \sum_{i=1}^N \alpha_i \right)^2 \leq N \sum_{i=1}^N \alpha_i^2.$$

*Proof.* The claim holds for  $N = 1$ . Now suppose the claim is true for  $N - 1$ . By Young's inequality with  $\varepsilon = 1$ , we obtain

$$\begin{aligned} \left( \sum_{i=1}^N \alpha_i \right)^2 &= \alpha_N^2 + 2 \cdot \left( \sum_{i=1}^{N-1} \alpha_N \cdot \alpha_i \right) + \left( \sum_{i=1}^{N-1} \alpha_i \right)^2 \\ &\stackrel{Young}{\leq} \alpha_N^2 + \sum_{i=1}^{N-1} (\alpha_N^2 + \alpha_i^2) + \left( \sum_{i=1}^{N-1} \alpha_i \right)^2 \\ &\stackrel{Hypoth.}{\leq} \alpha_N^2 + \sum_{i=1}^{N-1} (\alpha_N^2 + \alpha_i^2) + (N-1) \sum_{i=1}^{N-1} \alpha_i^2 \\ &= N \sum_{i=1}^N \alpha_i^2. \end{aligned}$$

This completes the proof. □

Lemma 2.5 immediately yields the following inequality.

**Lemma 2.6.** *For arbitrary  $a_1, \dots, a_N \in \mathcal{H}$ , it holds that*

$$\left\| \sum_{i=1}^N a_i \right\|^2 \leq N \sum_{i=1}^N \|a_i\|^2$$

*Proof.* We obtain

$$\left\| \sum_{i=1}^N a_i \right\|^2 \leq \left( \sum_{i=1}^N \|a_i\| \right)^2 \leq N \sum_{i=1}^N \|a_i\|^2.$$

from the triangle inequality and Lemma 2.5. □

Finally, we restate an identity which follows directly by expanding  $\|w - v\|^2$ .

**Lemma 2.7** (Polarization Identity). *For arbitrary  $w, v \in \mathcal{H}$ , we have*

$$2\langle w | v \rangle = \|w\|^2 + \|v\|^2 - \|w - v\|^2.$$

### 3 Generalized Nash Equilibrium Problems

We first formulate our class of GNEPs as an inclusion in Section 3.1. This formulation plays a central role in our analysis. We then introduce and discuss our assumptions in Section 3.2.

#### 3.1 Reformulation of the Nash Equilibrium Problem

In this work, we consider a special kind of solution of problem (1). To this end, let

$$\Psi(x, y) := \sum_{\nu=1}^N (\theta_{\nu}(x_{\nu}, x_{-\nu}) + \varphi_{\nu}(x_{\nu}) - \theta_{\nu}(y_{\nu}, x_{-\nu}) - \varphi_{\nu}(y_{\nu}))$$

be the *Nikaido-Isoda-function* of the GNEP from (1). Then  $x^* \in \mathcal{H}$  is called a *normalized equilibrium* or a *variational equilibrium* of (1) if  $\sup_{y \in \mathcal{F}} \Psi(x^*, y) = 0$ . Following [13], for example, it is not difficult to see that every variational equilibrium is a generalized Nash equilibrium of the GNEP.

We also introduce the *pseudo-gradient*  $\hat{P}_{\theta} : \mathcal{H} \rightarrow \mathcal{H}$  of the functions  $\theta_{\nu}$  from (1) by

$$\hat{P}_{\theta}(x) := \begin{pmatrix} \nabla_{x_1} \theta_1(x_1, x_{-1}) \\ \vdots \\ \nabla_{x_N} \theta_N(x_N, x_{-N}) \end{pmatrix}. \quad (2)$$

Further note that the definition of  $\varphi$  yields

$$\partial\varphi(x) = \begin{pmatrix} \partial_{x_1} \varphi_1(x_1) \\ \vdots \\ \partial_{x_N} \varphi_N(x_N) \end{pmatrix}.$$

If all  $\varphi_{\nu}$  are differentiable, we notice that  $\partial\varphi = \nabla\varphi = \hat{P}_{\varphi}$ .

This notation allows us to extend a known result from finite-dimensional GNEPs (see, e.g. [13]) to our Hilbert space setting.

The proof of the next Theorem can be found in [3, Thm. 3.1].

**Theorem 3.1** ([3]). *Under the given convexity and smoothness assumptions on  $\theta_{\nu}$  and  $\varphi_{\nu}$ , it holds that  $x^*$  is a variational equilibrium of (1) if and only if  $0 \in \partial\varphi(x^*) + \hat{P}_{\theta}(x^*) + N_{\mathcal{F}}(x^*)$ .*

Note that the previous result remains true for more general convex sets  $\mathcal{X}$ , not necessarily given as in our framework from Section 1.

Under certain regularity conditions we can characterize the normal cone  $N_{\mathcal{F}}(x^*)$ . This leads to a particular notion of a KKT point for which we use the following terminology.

**Definition 3.2.** A pair  $(x^*, \mu^*) \in \mathcal{X} \times \mathcal{K}$  is called a *variational KKT point* of (1) if it satisfies the following *KKT-type conditions*:  $0 \in \partial\varphi(x) + \hat{P}_{\theta}(x) + A^* \mu + N_{\mathcal{X}}(x)$  and  $0 = b - Ax$ .

Note that a variational KKT point has to be feasible with respect to the abstract constraints  $\mathcal{X}$ , whereas it exploits the existence of a multiplier for the equality constraints. This setting is useful for our ADMM-type method where only the linear constraints are penalized, whereas the abstract constraints remain unchanged.

The following result clarifies the relation between variational equilibria and variational KKT points of problem (1). Its proof can be found in [3, Thm. 3.3].

**Theorem 3.3.** *The following statements hold:*

- (a) *If  $(x^*, \mu^*) \in \mathcal{H} \times \mathcal{K}$  is a variational KKT pair of (1), then  $x^*$  is a variational equilibrium.*
- (b) *Conversely, assume that  $A \in \mathcal{L}(\mathcal{H}, \mathcal{K})$  has closed range, and that  $\text{int } \mathcal{X} \cap \{x \in \mathcal{H} \mid Ax = b\} \neq \emptyset$ . If  $x^* \in \mathcal{F}$  is a variational equilibrium of (1), then there exists a multiplier  $\mu^*$  such that  $(x^*, \mu^*) \in \mathcal{H} \times \mathcal{K}$  is a variational KKT pair of (1).*

In our subsequent algorithm for the solution of the GNEP from (1), we will compute a variational KKT point. Theorem 3.3 shows that this always yields a variational equilibrium, and that this approach is actually equivalent to finding a variational equilibrium under a certain regularity condition.

In order to rewrite the variational KKT conditions in a more compact form, let us further introduce the notation

$$\mathcal{W} := \mathcal{X}_1 \times \dots \times \mathcal{X}_N \times \mathcal{K}, \quad w := (x_1, \dots, x_N, \mu), \quad \psi(w) := \varphi(x), \quad (3)$$

where the last expression is just a formal re-definition of the mapping  $\varphi$  with the only difference being that  $\psi$  is viewed as a function of all variables  $w$ , whereas  $\varphi$  depends only on  $x$ . Hence

$$\partial\psi(w) = \begin{pmatrix} \partial\varphi(x) \\ \{0\} \end{pmatrix}, \quad (4)$$

where the corresponding subdifferentials are taken with respect to  $w$  and  $x$ , respectively. Moreover, we define the pseudo-gradient as a mapping of the whole vector  $w = (x, \mu)$  by

$$P_\theta(w) := \begin{pmatrix} \widehat{P}_\theta(x) \\ \{0\} \end{pmatrix}. \quad (5)$$

Finally, let us define  $G : \mathcal{H} \times \mathcal{K} \rightarrow \mathcal{H} \times \mathcal{K}$  by

$$G(w) := \begin{pmatrix} A_1^* \mu \\ \vdots \\ A_N^* \mu \\ b - \sum_{\nu=1}^N A_\nu x_\nu \end{pmatrix}. \quad (6)$$

The particular structure of  $G$  immediately yields the following result.

**Lemma 3.4.** *The mapping  $G$  as defined in (6) satisfies  $\langle G(w) - G(\bar{w}) \mid w - \bar{w} \rangle = 0$  for all  $w, \bar{w} \in \mathcal{W}$ ; in particular,  $G$  is a continuous monotone operator.*

The above notation gives a compact representation of the variational KKT conditions.

**Lemma 3.5.** *The vector pair  $w^* = (x^*, \mu^*) \in \mathcal{X} \times \mathcal{K}$  is a variational KKT point of (1) if and only if  $w^* \in \mathcal{W}^*$ , where  $\mathcal{W}^* := \{w \in \mathcal{W} \mid 0 \in \partial\psi(w) + P_\theta(w) + G(w) + N_{\mathcal{W}}(w)\}$ .*

*Proof.* The proof follows from the previous definitions, taking into account that, due to the Cartesian structure of  $\mathcal{W}$ , we have  $N_{\mathcal{W}}(w) = N_{\mathcal{X}_1}(x_1) \times \dots \times N_{\mathcal{X}_N}(x_N) \times N_{\mathcal{K}}(\mu)$  and  $N_{\mathcal{K}}(\mu) = \{0\}$  since  $\mathcal{K}$  is the entire space.  $\square$

Applying the definitions of the normal cone and the convex subdifferential to the statement of Lemma 3.5, we obtain the following characterization of the variational KKT points.

**Lemma 3.6.** *The vector pair  $w^* = (x^*, \mu^*) \in \mathcal{X} \times \mathcal{K}$  is a variational KKT point of (1) if and only if there is a  $g^* \in \partial\psi(w^*)$  such that*

$$0 \leq \langle g^* + P_\theta(w^*) + G(w^*) \mid w - w^* \rangle,$$

for all  $w \in \mathcal{W}$ .

Let us define the multifunction

$$T(w) := \partial\psi(w) + P_\theta(w) + G(w) + N_{\mathcal{W}}(w). \quad (7)$$

The domain of  $T$  is obviously given by the nonempty set  $\mathcal{W}$ . Then the set  $\mathcal{W}^*$  from Lemma 3.5 can be expressed as  $\mathcal{W}^* = \{w \in \mathcal{W} \mid 0 \in T(w)\}$ , i.e.

$$0 \in T(w^*) \iff w^* = (x^*, \mu^*) \text{ is a KKT point of (1)}. \quad (8)$$

. This indicates that the set-valued mapping  $T$  plays a central role in our analysis. Its most important property is formulated in the following result, whose proof can be found in [3, Prop. 3.6].

**Proposition 3.7.** *Suppose that the pseudo-gradient  $P_\theta : \mathcal{H} \times \mathcal{K} \rightarrow \mathcal{H} \times \mathcal{K}$  defined in (5) is a monotone mapping. Then the set-valued function  $T$ , defined in (7), is maximally monotone.*

## 3.2 Assumptions

In this section we state the standing assumptions of this work, and discuss and compare these assumptions with those usually made in splitting-type methods.

Our aim is to generalize the well-known alternating direction method of multipliers for optimization problems to generalized Nash equilibrium problems with shared constraints. We also want to deal with more than two players. Hence we cannot expect to deal with weaker conditions as known for the multi-block ADMM, i.e. the ADMM with more than two functions, for finite-dimensional optimization problems. In [8], it is shown that the multi-block ADMM is not necessarily convergent when there are more than two (only convex) functions involved. On the other hand, the recent paper [29] shows that the multi-block ADMM is convergent when all except of two functions are strongly convex.



In our generalized Nash equilibrium setting this tells us that it will not be enough to require  $\widehat{P}_\theta$  to be monotone and that we will need  $\widehat{P}_\theta$  to be strongly monotone in a certain part of the variable  $x = (x_1, \dots, x_N)$ . In fact, we will require  $\widehat{P}_\theta$  to be strongly monotone in all except of one variable, i.e. in one variable more than in [29]. Taking into account that our class of problems is far more general and difficult to deal with, we believe that this condition is sufficiently weak.

Different from the multi-block ADMM, the objective functions of our players also depend on the strategies of the other players. In order to have more control of the consecutive iterates, we also need a certain Lipschitz continuity condition on the gradient of the player's objective functions. The Lipschitz condition applied here is weaker than the usual Lipschitz assumption typically used in the convergence analysis of first-order methods since we assume the Lipschitz continuity to hold only with respect to certain input variables. Hence both our strong monotonicity condition and Lipschitz condition are weaker than the standard assumptions used in the context of splitting methods.

A precise statement of our assumptions follows.

**Assumption 3.8.** (U) For all  $\nu = 1, \dots, N$ , let  $A_\nu \in \mathcal{L}(\mathcal{H}_\nu, \mathcal{K})$ ,  $b \in \mathcal{K}$ ,  $\varphi_\nu : \mathcal{H}_\nu \rightarrow \mathbb{R}$  be proper, lower semi-continuous, convex,  $\text{dom } \varphi_\nu = \mathcal{H}_\nu$ ,  $\theta_\nu(\cdot, x_{-\nu})$  convex and continuously Fréchet-differentiable with the derivative being continuous with respect to all variables  $x = (x_1, \dots, x_N)$ . and let  $\mathcal{X}_\nu$  be nonempty, closed, and convex.

(S) Suppose that  $\partial\varphi + \widehat{P}_\theta$  is monotone and strongly monotone in  $x_{-N} = (x_1, \dots, x_{N-1})$ , i.e. there is a  $\rho > 0$  such that for all  $x, y \in \mathcal{H}$  and all  $g_x \in \partial\varphi(x)$ ,  $g_y \in \partial\varphi(y)$  we have

$$\langle (g_x + P_\theta(x)) - (g_y + P_\theta(y)) \mid x - y \rangle \geq \rho \|x_{-N} - y_{-N}\|^2 = \rho \sum_{\nu=1}^{N-1} \|x_\nu - y_\nu\|^2.$$

(L) Assume that  $\nabla_\nu \theta_\nu$  is Lipschitz continuous in the last  $N - \nu$  components, i.e. there is an  $L_\nu > 0$  such that

$$\begin{aligned} \|\nabla_\nu \theta_\nu(x_1, \dots, x_\nu, y_{\nu+1}, \dots, y_N) - \nabla_\nu \theta_\nu(x_1, \dots, x_\nu, x_{\nu+1}, \dots, x_N)\|^2 \\ \leq L_\nu^2 \sum_{i=\nu+1}^N \|x_i - y_i\|^2 \end{aligned} \quad (9)$$

for all  $\nu = 1, \dots, N$ .

Assumption 3.8 (S) gives us the opportunity to treat problems which are strongly monotone in  $N - 1$  components only. This allows us to deal with conic and inequality constraints through the approach described in [3].

For notational convenience, let us define

$$C_\nu := \max_{i=\nu+1, \dots, N} \|A_\nu^* \mathcal{A}_i\|^2. \quad (10)$$

## 4 ADMM Method with Fixed Regularization

This section is devoted to a regularized ADMM method for the solution of the GNEP from (1). The method presented here uses a fixed regularization parameter. A modification which includes an update of the regularization parameter will be discussed in the subsequent section. The precise statement of the algorithm is given in Section 4.1, whereas its global convergence is analyzed in Section 4.2. Finally, Section 4.3 contains a brief discussion regarding the necessity of the regularization. In particular, it turns out that the regularization is (in general) necessary even in the case of just two players.

### 4.1 Statement of Algorithm

The following regularized alternating direction method of multipliers for the solution of (1) will be investigated in this section. Its basic idea is the following: The joint constraints get augmented in order to obtain a separable structure in the remaining constraints. We then use the direct extension of the ADMM and view the resulting optimization problems of each player as minimization problems of the variables  $x^\nu$  alone. A proximal term is added to improve the convergence properties.

**Algorithm 4.1.** (*ADMM Method with Fixed Regularization*)

(S.0) Choose a starting point  $(x^0, \mu^0) \in \mathcal{X} \times \mathcal{K}$ , parameters  $\beta > 0$ ,  $\gamma_\nu > 0$  for all  $\nu = 1, \dots, N$ , and set  $k := 0$ .

(S.1) If a suitable termination criterion is satisfied: STOP.

(S.2) For  $\nu = 1, \dots, N$ , compute

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \theta_\nu(x_1^{k+1}, \dots, x_{\nu-1}^{k+1}, x_\nu, x_{\nu+1}^k, \dots, x_N^k) \right. \\ \left. + \langle \mu^k | A_\nu x_\nu \rangle_{\mathcal{K}} + \frac{\gamma_\nu}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 \right. \\ \left. + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{i=1}^{\nu-1} A_i x_i^{k+1} + \sum_{i=\nu+1}^N A_i x_i^k - b\|_{\mathcal{K}}^2 \right\}. \quad (11a)$$

(S.3) Define

$$\mu^{k+1} := \mu^k + \beta \left( \sum_{\nu=1}^N A_\nu x_\nu^{k+1} - b \right). \quad (11b)$$

(S.4) Set  $k \leftarrow k + 1$ , and go to (S.1).

We note that all subproblems (11a) are strongly convex for all  $\nu$  and all iterations  $k$ . Hence all iterates  $x^{k+1} := (x_1^{k+1}, \dots, x_N^{k+1})$  are well-defined and uniquely determined. Throughout our convergence analysis, we assume implicitly that Algorithm 4.1 generates an infinite number of iterates.

The main computational overhead in Algorithm 4.1 comes from the solution of the optimization subproblems in (S.2). However, in contrast to augmented Lagrangian-type methods [24, 25], for example, these subproblems are only optimization problems and not Nash equilibrium problems themselves. Moreover, the subproblems occurring in (S.2) can typically be solved in an efficient way, sometimes even analytically.

## 4.2 Convergence

In this section, we analyze the convergence of Algorithm 4.1 under Assumption 3.8. To keep the notation reasonably simple, we make extensive use of the abbreviations introduced in (3)-(6) (recall that  $w = (x, \mu)$ ). Moreover, it will be convenient to define the auxiliary vectors

$$\hat{x}^{\nu,k} := (x_1^{k+1}, \dots, x_\nu^{k+1}, x_{\nu+1}^k, \dots, x_N^k) \quad \text{for } \nu = 1, \dots, N, \quad (12)$$

in particular, it holds that  $\hat{x}^{N,k} = x^{k+1}$ .

We begin our analysis with two simple results.

**Lemma 4.2.** *Suppose that Assumption 3.8 (S) holds, and let  $w^* = (x^*, \mu^*)$  be a variational KKT-point of (1). Then*

$$\langle g + P_\theta(w) + G(w) + s \mid w - w^* \rangle \geq \langle g + P_\theta(w) + G(w) \mid w - w^* \rangle \geq \rho \|x_{-N} - x_{-N}^*\|^2$$

holds for all  $x \in \mathcal{X}$ ,  $\mu \in \mathcal{K}$ ,  $g \in \partial\psi(w)$  and  $s \in N_{\mathcal{W}}(w)$ , where, again,  $w = (x, \mu)$ .

*Proof.* The first inequality follows from  $s \in N_{\mathcal{W}}(w)$ , which implies  $\langle s \mid w - w^* \rangle \geq 0$ . In order to verify the second inequality, first note that the assumed strong monotonicity of  $\partial\varphi + \hat{P}_\theta$  with respect to  $x_{-N}$  together with Lemma 3.4 yields

$$\langle g + P_\theta(w) + G(w) - (g^* + P_\theta(w^*) + G(w^*)) \mid w - w^* \rangle \geq \rho \|x_{-N} - x_{-N}^*\|^2$$

for all  $g^* \in \partial\psi(w^*)$  and  $g \in \partial\psi(w)$ . Now, in view of Lemma 3.6, there exists a particular element  $g^* \in \partial\psi(w^*)$  corresponding to the variational KKT pair  $w^*$  such that  $\langle g^* + P_\theta(w^*) + G(w^*) \mid w - w^* \rangle \geq 0$  holds for all  $w \in \mathcal{W}$ . Combining this with the previous inequality for this particular  $g^*$ , the statement follows.  $\square$

The second preliminary result states a partial uniqueness of a variational KKT pair.

**Lemma 4.3.** *Suppose that Assumption 3.8 (S) holds, and let  $w^* = (x^*, \mu^*)$  and  $\bar{w} = (\bar{x}, \bar{\mu})$  be two variational KKT points of (1). Then  $x_{-N}^* = \bar{x}_{-N}$ .*

*Proof.* By Definition 3.2 of a variational KKT point, we see that there exist  $\bar{g} \in \partial\psi(\bar{w})$  and  $\bar{s} \in N_{\mathcal{W}}(\bar{w})$  such that  $0 = \bar{g} + P_\theta(\bar{w}) + G(\bar{w}) + \bar{s}$ . Using Lemma 4.2 and the fact that  $w^*$  is a variational KKT pair, we therefore obtain

$$0 = \langle 0 \mid \bar{w} - w^* \rangle = \langle \bar{g} + P_\theta(\bar{w}) + G(\bar{w}) + \bar{s} \mid \bar{w} - w^* \rangle \geq \rho \|\bar{x}_{-N} - x_{-N}^*\|^2,$$

hence  $\bar{x}_{-N} = x_{-N}^*$ .  $\square$

In order to write the optimality conditions of (11) as a suitable inclusion, we introduce two linear operators  $Q_N, R_N \in \mathcal{L}(\mathcal{H} \times \mathcal{K})$ . The first operator

$$Q_N(w) := Q_N(x, \mu) := \begin{pmatrix} \gamma_1 x_1 \\ \vdots \\ \gamma_N x_N \\ \frac{1}{\beta} \mu \end{pmatrix} \quad (13)$$

represents the regularization. It is easy to see that  $Q_N$  is self-adjoint and strongly monotone for all  $\gamma_\nu > 0$  and  $\beta > 0$ . The second operator is given by

$$R_N(w) := \begin{pmatrix} 0 & -\beta A_1^* A_2 & \dots & \dots & -\beta A_1^* A_N & 0 \\ 0 & 0 & -\beta A_2^* A_3 & \dots & -\beta A_2^* A_N & 0 \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ 0 & & & & -\beta A_{N-1}^* A_N & 0 \\ 0 & 0 & \dots & & 0 & 0 \\ 0 & 0 & \dots & & 0 & 0 \end{pmatrix} w. \quad (14)$$

These two operators allow the following formulation of the subproblems from (11) as a generalized equation.

**Lemma 4.4.** *The iteration (11) is equivalent to finding  $w^{k+1} \in \mathcal{H} \times \mathcal{K}$  such that*

$$0 \in T(w^{k+1}) + Q_N(w^{k+1} - w^k) + R_N(w^{k+1} - w^k) + \begin{pmatrix} \nabla_{x_1} \theta_1(\hat{x}^{1,k}) - \nabla_{x_1} \theta_1(x^{k+1}) \\ \vdots \\ \nabla_{x_{N-1}} \theta_{N-1}(\hat{x}^{N-1,k}) - \nabla_{x_{N-1}} \theta_{N-1}(x^{k+1}) \\ 0 \\ 0 \end{pmatrix}, \quad (15)$$

where  $T$  is the set-valued operator from (7).

*Proof.* First note that, due to the Cartesian structure of  $\mathcal{W}$ , we have  $N_{\mathcal{W}}(w^{k+1}) = N_{\mathcal{X}_1}(x_1^{k+1}) \times \dots \times N_{\mathcal{X}_N}(x_N^{k+1}) \times N_{\mathcal{K}}(\mu^{k+1})$ . Furthermore, using the sum rule of the convex subdifferential, cf. [2, Cor. 16.48] (which can be applied because of Assumption 3.8 (U)) as well as the definitions of  $\hat{x}^{\nu,k}$  and  $\mu^{k+1}$ , the optimality conditions of the convex optimization problem (11a) can be written as

$$\begin{aligned} 0 &\in \partial_{x_\nu} \varphi_\nu(x_\nu^{k+1}) + \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) + A_\nu^* \mu^k + \beta A_\nu^* (A \hat{x}^{\nu,k} - b) + \gamma_\nu (x_\nu^{k+1} - x_\nu^k) + N_{\mathcal{X}_\nu}(x_\nu^{k+1}) \\ &= \partial_{x_\nu} \varphi_\nu(x_\nu^{k+1}) + \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) + A_\nu^* \mu^{k+1} - \beta \sum_{i=\nu+1}^N A_\nu^* A_i (x_i^{k+1} - x_i^k) \\ &\quad + \gamma_\nu (x_\nu^{k+1} - x_\nu^k) + N_{\mathcal{X}_\nu}(x_\nu^{k+1}) \\ &= \left[ \partial_{x_\nu} \varphi_\nu(x_\nu^{k+1}) + \nabla_{x_\nu} \theta_\nu(x^{k+1}) + A_\nu^* \mu^{k+1} + N_{\mathcal{X}_\nu}(x_\nu^{k+1}) \right] + \left[ \gamma_\nu (x_\nu^{k+1} - x_\nu^k) \right] \\ &\quad + \left[ -\beta \sum_{i=\nu+1}^N A_\nu^* A_i (x_i^{k+1} - x_i^k) \right] + \left[ \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1}) \right] \end{aligned}$$

for all  $\nu = 1, \dots, N$ . It is easy to see that the terms in brackets correspond to the  $\nu$ th row of the operator on the right-hand side of (15) (for  $\nu = N$  we exploited the fact that  $\hat{x}^{N,k} = x^{k+1}$ ).

The last row of (15) results from the updating rule (11b), which can be rewritten as

$$0 = \frac{1}{\beta}(\mu^{k+1} - \mu^k) - \left( \sum_{\nu=1}^N A_{\nu} x_{\nu}^{k+1} - b \right),$$

which, in turn, is equivalent to

$$0 \in \left[ -Ax^{k+1} + b + N_{\mathcal{K}}(\mu^{k+1}) \right] + \left[ \frac{1}{\beta}(\mu^{k+1} - \mu^k) \right] + [0] + [0].$$

since  $N_{\mathcal{K}}(\mu^{k+1}) = \{0\}$ . This completes the proof.  $\square$

Using the definitions of the operator  $T$  and the normal cone, Lemma 4.4 immediately yields the following result.

**Lemma 4.5.** *The iteration (11) is equivalent to finding  $w^{k+1} := (x_1^{k+1}, \dots, x_N^{k+1}, \mu^{k+1})$  and  $g^{k+1} \in \partial\psi(w^{k+1})$  such that*

$$\begin{aligned} & \langle g^{k+1} + P_{\theta}(w^{k+1}) + G(w^{k+1}) + Q_N(w^{k+1} - w^k) \mid w - w^{k+1} \rangle \\ & + \sum_{\nu=1}^{N-1} \langle \nabla_{x_{\nu}} \theta_{\nu}(\hat{x}^{\nu,k}) - \nabla_{x_{\nu}} \theta_{\nu}(x_{\nu}^{k+1}) \mid x_{\nu} - x_{\nu}^{k+1} \rangle \\ & - \beta \sum_{\nu=1}^{N-1} \sum_{i=\nu+1}^N \langle A_{\nu}^* A_i (x_i^{k+1} - x_i^k) \mid x_{\nu} - x_{\nu}^{k+1} \rangle \geq 0 \quad \forall w \in \mathcal{W}. \end{aligned} \quad (16)$$

The previous results allow to prove our main global convergence theorem.

**Theorem 4.6.** *Assume that problem (1) admits a variational KKT-point, and that Assumption 3.8 holds. Further suppose that  $\gamma_1 > 0$  and*

$$\gamma_{\nu} > \frac{1}{\rho} \sum_{i=1}^{\nu-1} L_i^2 + \frac{1}{\rho} N \beta^2 \sum_{i=1}^{\nu-1} C_i \quad \text{for all } \nu = 2, \dots, N. \quad (17)$$

*Then the iterates  $w^{k+1}$  generated by Algorithm 4.1 converge weakly to a variational KKT pair  $(x_1^*, \dots, x_N^*, \mu^*)$  of (1). Furthermore,  $x_{-N}^k$  converges strongly to  $x_{-N}^*$ .*

Before proving this theorem, let us add some comments. The first regularization parameter  $\gamma_1$  can be chosen as an arbitrary positive constant. On the other hand, the remaining regularization parameters have to satisfy condition (17) which implies, in particular, that  $\gamma_2 < \gamma_3 < \dots < \gamma_N$ , i.e. the lower bounds are getting more and more restrictive. Large values of these regularization parameters typically slow down the convergence rate. We therefore present a modification of Algorithm 4.1 with an adaptive updating rule of these regularization parameters to avoid unnecessarily large values for these parameters in the next section. Since the convergence theory of this adapted version depends on the convergence result for Algorithm 4.1, we have to investigate the properties of Algorithm 4.1 first.

Let us shortly discuss the special case of  $N = 2$  players in Theorem 4.6. Assumption 3.8 then requires  $\partial\varphi + \widehat{P}_\theta$  to be strongly monotone in  $x_1$  only, and  $\nabla_{x_1}\theta_1$  to be Lipschitz continuous in the second component  $x_2$ . Both conditions are significantly weaker than the usual assumption of  $\widehat{P}_\theta$  being strongly monotone and Lipschitz continuous. The conditions regarding the regularization parameters are  $\gamma_1 > 0$  and  $\gamma_2 > \frac{1}{\rho}(L_1^2 + 2\beta^2\|A_1^*A_2\|)$ . Then we obtain that  $(x_1^{k+1}, x_2^{k+1}, \mu^{k+1})$  converges weakly to a solution and that  $x_1^{k+1}$  even converges strongly.

*Proof of Theorem 4.6.* First recall that the linear operator  $Q_N$  from (13) is self-adjoint and strongly monotone. Hence it induces a scalar product  $\langle x, y \rangle_{Q_N} := \langle x, Q_N y \rangle$  in the Hilbert space  $\mathcal{H} \times \mathcal{K}$ . The norm induced by this scalar product will be denoted by  $\|\cdot\|_{Q_N}$  throughout this proof.

Using Lemma 4.2, we obtain

$$\langle g^{k+1} + P_\theta(w^{k+1}) + G(w^{k+1}) \mid w^{k+1} - w^* \rangle \geq \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \quad (18)$$

for an arbitrary  $g^{k+1} \in \partial\psi(w^{k+1})$ . Setting  $w = w^*$  in (16), applying the Cauchy-Schwarz inequality (CSI) and using Young's inequality twice, the first time with some  $\varepsilon > 0$  and the second time with some  $\delta > 0$  (at this stage of the proof,  $\delta, \varepsilon > 0$  are arbitrary, however, later these values will be specified in a suitable way), we obtain

$$\begin{aligned} 0 &\stackrel{\text{Lem. 4.5}}{\leq} \langle g^{k+1} + P_\theta(w^{k+1}) + G(w^{k+1}) + Q_N(w^{k+1} - w^k) \mid w^* - w^{k+1} \rangle \\ &\quad + \sum_{\nu=1}^{N-1} \langle \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1}) \mid x_\nu^* - x_\nu^{k+1} \rangle \\ &\quad - \beta \sum_{\nu=1}^{N-1} \sum_{i=\nu+1}^N \langle A_\nu^* A_i(x_i^{k+1} - x_i^k) \mid x_\nu^* - x_\nu^{k+1} \rangle \\ &\stackrel{(18)}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\ &\quad + \sum_{\nu=1}^{N-1} \langle \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1}) \mid x_\nu^* - x_\nu^{k+1} \rangle \\ &\quad - \beta \sum_{\nu=1}^{N-1} \left\langle \sum_{i=\nu+1}^N A_\nu^* A_i(x_i^{k+1} - x_i^k) \mid x_\nu^* - x_\nu^{k+1} \right\rangle \\ &\stackrel{\text{CSI}}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\ &\quad + \sum_{\nu=1}^{N-1} \left\| \nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1}) \right\| \cdot \|x_\nu^{k+1} - x_\nu^*\| \\ &\quad + \beta \sum_{\nu=1}^{N-1} \left\| \sum_{i=\nu+1}^N A_\nu^* A_i(x_i^{k+1} - x_i^k) \right\| \cdot \|x_\nu^{k+1} - x_\nu^*\| \\ &\stackrel{\text{Young}}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \end{aligned}$$

$$\begin{aligned}
& + \sum_{\nu=1}^{N-1} \left( \frac{\varepsilon}{2} \|\nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1})\|^2 + \frac{1}{2\varepsilon} \|x_\nu^{k+1} - x_\nu^*\|^2 \right) \\
& + \sum_{\nu=1}^{N-1} \left( \frac{\beta^2 \delta}{2} \left\| \sum_{i=\nu+1}^N A_\nu^* A_i (x_i^{k+1} - x_i^k) \right\|^2 + \frac{1}{2\delta} \|x_\nu^{k+1} - x_\nu^*\|^2 \right) \\
& \stackrel{(9)}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \sum_{\nu=1}^{N-1} \frac{\varepsilon}{2} L_\nu^2 \sum_{i=\nu+1}^N \|x_i^{k+1} - x_i^k\|^2 + \frac{1}{2\varepsilon} \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \sum_{\nu=1}^{N-1} \frac{\beta^2 \delta}{2} \left\| \sum_{i=\nu+1}^N A_\nu^* A_i (x_i^{k+1} - x_i^k) \right\|^2 + \frac{1}{2\delta} \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& \stackrel{\text{Lem. 2.6}}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \rho \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \sum_{\nu=1}^{N-1} \frac{\varepsilon}{2} L_\nu^2 \sum_{i=\nu+1}^N \|x_i^{k+1} - x_i^k\|^2 + \frac{1}{2\varepsilon} \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \sum_{\nu=1}^{N-1} \frac{\beta^2 \delta N}{2} \sum_{i=\nu+1}^N \|A_\nu^* A_i\|^2 \|x_i^{k+1} - x_i^k\|^2 + \frac{1}{2\delta} \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& = \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \left( \rho - \frac{1}{2\varepsilon} - \frac{1}{2\delta} \right) \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \sum_{\nu=1}^{N-1} \sum_{i=\nu+1}^N \left( \frac{\varepsilon}{2} L_\nu^2 + \frac{\beta^2 \delta N}{2} \|A_\nu^* A_i\|^2 \right) \|x_i^{k+1} - x_i^k\|^2 \\
& \stackrel{(10)}{\leq} \langle w^{k+1} - w^k \mid w^* - w^{k+1} \rangle_{Q_N} - \left( \rho - \frac{1}{2\varepsilon} - \frac{1}{2\delta} \right) \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \frac{1}{2} \sum_{\nu=1}^{N-1} \sum_{i=\nu+1}^N \left( \varepsilon L_\nu^2 + \beta^2 \delta N C_\nu \right) \|x_i^{k+1} - x_i^k\|^2 \\
& = - \langle w^{k+1} - w^k \mid w^{k+1} - w^* \rangle_{Q_N} - \left( \rho - \frac{1}{2\varepsilon} - \frac{1}{2\delta} \right) \|x_{-N}^{k+1} - x_{-N}^*\|^2 \\
& + \frac{1}{2} \sum_{\nu=2}^N \|x_\nu^{k+1} - x_\nu^k\|^2 \cdot \sum_{i=1}^{\nu-1} \left( \varepsilon L_i^2 + \beta^2 \delta N C_i \right) \\
& \stackrel{\text{Lem. 2.7}}{=} \frac{1}{2} \|w^k - w^*\|_{Q_N}^2 - \frac{1}{2} \|w^{k+1} - w^k\|_{Q_N}^2 - \frac{1}{2} \|w^{k+1} - w^*\|_{Q_N}^2 \\
& - \left( \rho - \frac{1}{2\varepsilon} - \frac{1}{2\delta} \right) \|x_{-N}^{k+1} - x_{-N}^*\|^2 + \frac{1}{2} \sum_{\nu=2}^N \|x_\nu^{k+1} - x_\nu^k\|^2 \cdot \sum_{i=1}^{\nu-1} \left( \varepsilon L_i^2 + \beta^2 \delta N C_i \right),
\end{aligned}$$

where the penultimate equality follows from an elementary reordering of the corresponding terms. Multiplying this inequality by 2, using the definition of  $Q_N$ , and exploiting the block

structure of  $w = (x, \mu) = (x_1, x_{-1}, \mu)$ , we therefore obtain

$$\begin{aligned} & \|w^{k+1} - w^*\|_{Q_N}^2 + \left(2\rho - \frac{1}{\varepsilon} - \frac{1}{\delta}\right) \|x_{-N}^{k+1} - x_{-N}^*\|^2 + \frac{1}{\beta} \|\mu^{k+1} - \mu^k\|^2 + \gamma_1 \|x_1^{k+1} - x_1^k\|^2 \\ & + \sum_{\nu=2}^N \|x_\nu^{k+1} - x_\nu^k\|^2 \cdot \left( \gamma_\nu - \varepsilon \sum_{i=1}^{\nu-1} L_i^2 - \delta \beta^2 N \sum_{i=1}^{\nu-1} C_i \right) \leq \|w^k - w^*\|_{Q_N}^2. \end{aligned} \quad (19)$$

Assumption (17) now guarantees that we can find suitable  $\varepsilon, \delta \in (\frac{1}{\rho}, \infty)$  such that

$$\hat{\gamma}_\nu := \left( \gamma_\nu - \varepsilon \sum_{i=1}^{\nu-1} L_i^2 - \delta \beta^2 N \sum_{i=1}^{\nu-1} C_i \right) > 0 \quad \text{for all } \nu = 2, \dots, N.$$

For the sake of completeness, let us also define  $\hat{\gamma}_1 := \gamma_1 > 0$ . Note that the choice of  $\delta, \varepsilon$  implies

$$\hat{\rho} := 2\rho - \frac{1}{\varepsilon} - \frac{1}{\delta} > 0.$$

With these positive scalars, we can rewrite (19) as

$$\begin{aligned} & \|w^{k+1} - w^*\|_{Q_N}^2 + \hat{\rho} \|x_{-N}^{k+1} - x_{-N}^*\|^2 + \frac{1}{\beta} \|\mu^{k+1} - \mu^k\|^2 + \sum_{\nu=1}^N \hat{\gamma}_\nu \|x_\nu^{k+1} - x_\nu^k\|^2 \\ & \leq \|w^k - w^*\|_{Q_N}^2. \end{aligned} \quad (20)$$

Summation of (20) for  $k = 0, 1, \dots, l$  yields

$$\sum_{k=0}^l \left( \hat{\rho} \|x_{-N}^{k+1} - x_{-N}^*\|^2 + \frac{1}{\beta} \|\mu^{k+1} - \mu^k\|^2 + \sum_{\nu=1}^N \hat{\gamma}_\nu \|x_\nu^{k+1} - x_\nu^k\|^2 \right) \leq \|w^0 - w^*\|_{Q_N}^2.$$

Taking the limit  $l \rightarrow \infty$ , we therefore obtain  $\|\mu^{k+1} - \mu^k\| \rightarrow 0$ ,  $\|x_\nu^{k+1} - x_\nu^k\| \rightarrow 0$  for all  $\nu = 1, \dots, N$ , and  $\|x_{-N}^{k+1} - x_{-N}^*\| \rightarrow 0$ . Hence  $w^{k+1} - w^k \rightarrow 0$  and  $x_{-N}^{k+1} \rightarrow x_{-N}^*$ , this convergence being the second part of our claim. It therefore remains to be shown that  $\{x_N^k\}$  and  $\{\mu^k\}$  converge weakly, and that the weak limit is a solution.

In view of (20), the sequence  $\{w^{k+1}\}$  is bounded, thus it has a weakly convergent subsequence  $w^{k+1} \rightharpoonup_{\mathcal{I}} \bar{w}$ . The fact that  $w^{k+1} - w^k \rightarrow 0$  then also implies  $w^k \rightharpoonup_{\mathcal{I}} \bar{w}$ . By the assumed Lipschitz continuity of  $\nabla_{x_\nu} \theta_\nu$  in the last  $N - \nu$  components, we obtain

$$\|\nabla_{x_\nu} \theta_\nu(\hat{x}^{\nu,k}) - \nabla_{x_\nu} \theta_\nu(x^{k+1})\|^2 \leq L_\nu^2 \sum_{i=\nu+1}^N \|x_i^{k+1} - x_i^k\|^2 \rightarrow 0.$$

Recall from Lemma 4.4 that the optimality conditions of the subproblems from (11) can be rewritten as

$$-Q_N(w^{k+1} - w^k) - R_N(w^{k+1} - w^k) - \begin{pmatrix} \nabla_{x_1} \theta_1(\hat{x}^{1,k}) - \nabla_{x_1} \theta_1(x^{k+1}) \\ \vdots \\ \nabla_{x_{N-1}} \theta_{N-1}(\hat{x}^{N-1,k}) - \nabla_{x_{N-1}} \theta_{N-1}(x^{k+1}) \\ 0 \\ 0 \end{pmatrix} \in T(w^{k+1}),$$



where  $T$  denotes the operator from (7). Now, our previous discussion shows that the left-hand side converges strongly to zero, whereas  $w^{k+1} \rightharpoonup_{\mathcal{I}} \bar{w}$ . Thus, by the strong-weak-sequential closedness of the graph of a maximally monotone operator and the maximal monotonicity of  $T$ , cf. Proposition 3.7, we obtain  $0 \in T(\bar{w})$ . Lemma 3.5 and (8) therefore imply that  $\bar{w}$  is a variational KKT point of (1). Since (20) eventually implies that the sequence  $\{w^k\}$  is Fejér monotone with respect to the solution set and that every cluster point is a variational KKT point, it follows from Proposition 2.2 that  $w^{k+1} \rightharpoonup \bar{w}$ .  $\square$

### 4.3 Necessity of Regularization

Taking into account the situation of the standard ADMM method for optimization problems, one might expect that (a) no regularization is necessary for GNEPs with  $N = 2$  players, (b) arbitrary (possibly small) regularization parameters  $\gamma_\nu > 0$  are sufficient for global convergence of GNEPs with  $N \geq 3$  players. The subsequent discussion shows that none of these statements hold, hence a regularization is necessary also for two players, and the corresponding regularization parameters have to be sufficiently large. This also shows that the GNEP is a significantly more difficult class of problems. Altogether, this justifies that one cannot expect to prove a much stronger convergence result as the one given in Theorem 4.6.

In order to verify the above statements, let us consider the finite-dimensional GNEP

$$\begin{aligned} \min_{x_1 \in \mathbb{R}^{n_1}} \quad & \frac{1}{2} x_1^T U_{11} x_1 + x_1^T U_{12} x_2 \\ \min_{x_2 \in \mathbb{R}^{n_2}} \quad & \frac{1}{2} x_2^T U_{22} x_2 + x_2^T U_{21} x_1 \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 = 0 \end{aligned}$$

with  $N = 2$  players. Since both objective functions are quadratic, the corresponding subproblems (11) are simple quadratic programs, hence their optimality conditions result in a linear system of equations. In fact, an elementary calculation shows that the corresponding updating scheme from Lemma 4.4 boils down to the matrix iteration

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \mu^{k+1} \end{pmatrix} = \underbrace{\begin{pmatrix} U_{11} + \gamma_1 \text{id} & -\beta A_1^* A_2 & A_1^* \\ U_{21} & U_{22} + \gamma_2 \text{id} & A_2^* \\ -A_1 & -A_2 & \frac{1}{\beta} \text{id} \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \text{id} & -U_{12} - \beta A_1^* A_2 & 0 \\ 0 & \gamma_2 \text{id} & 0 \\ 0 & 0 & \frac{1}{\beta} \text{id} \end{pmatrix}}_{=:M} \begin{pmatrix} x_1^k \\ x_2^k \\ \mu^k \end{pmatrix}.$$

Standard results on splitting methods guarantee that this iteration converges, for all starting points, to a solution if and only if  $\rho(M) < 1$ , where  $\rho(M)$  denotes the spectral radius of  $M$ . In other words, for  $\rho(M) \geq 1$ , there exist starting points such that the above matrix iteration does not converge to a solution. Hence, in the following, we only have to compute the spectral radius for a particular instance of the above GNEP (note that a similar reasoning was also used in [8] in order to verify that the standard ADMM method is not necessarily convergent for finite-dimensional optimization problems with more than two blocks).

To be more specific, let us take  $U_{11} = 1$ ,  $U_{12} = -10$ ,  $U_{21} = 10$  and  $U_{22} = 1$ . Then it is easy to see that  $\hat{P}_\theta$  is strongly monotone with modulus  $\rho = 1$ , and Lipschitz continuous with

constant  $L = 11$ . In particular, this implies that Assumption 3.8 holds. Let us further choose  $A_1 = (1, 0)^T$ ,  $A_2 = (0, 1)^T$ ,  $\beta = 1$ ,  $\gamma_1 = 0.01$ . Then the spectral radius of  $M$  depends on the parameter  $\gamma_2$ . Using a simple program, we can compute this spectral radius for different values of  $\gamma_2$ . The corresponding result is shown in Fig. 1 which shows that  $\rho(M) < 1$  holds for values of  $\gamma_2$  larger than (approximately) 32.1, whereas for all values of  $\gamma_2$  less than 32.1, the spectral radius is larger than one. This example clearly shows that both a regularization is also necessary for two-player games, and that convergence cannot be expected, in general, for arbitrary  $\gamma_\nu > 0$ .

We close this section by noting that our theoretical lower bound for  $\gamma_2$  from (17) yields  $\gamma_2 \geq 100$  since  $L_1 = 10$  and  $A_1^* A_2 = 0$  in our case. Hence our theoretical lower bound is not sharp, though also not a too bad overestimate.

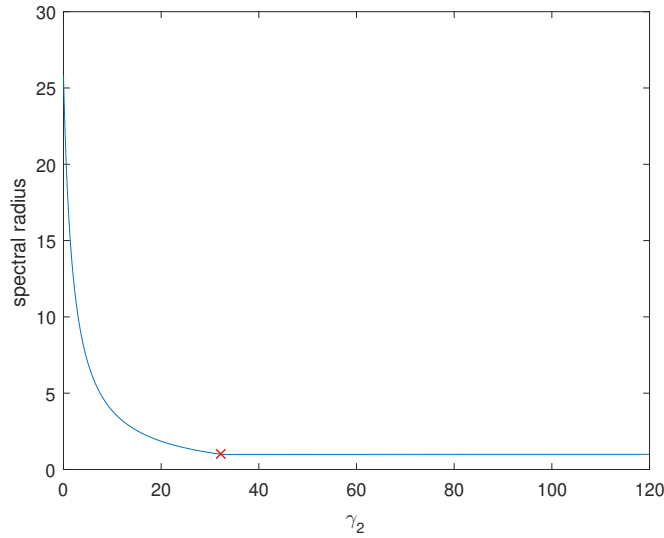


Figure 1: The spectral radius of the iteration matrix plotted against  $\gamma_2$ . The red cross indicates where the value drops below one.

## 5 ADMM Method with Adaptive Regularization

Theorem 4.6 proves convergence of Algorithm 4.1 for all sufficiently large regularization parameters  $\gamma_\nu$ . As the counterexample from Section 4.3 shows, one cannot expect convergence without taking these regularization parameters large enough. On the other hand, the same counterexample also indicates that the theoretical bounds from (17) are not sharp. Moreover, numerical tests indicate that Algorithm 4.1 often converges for significantly smaller regularization parameters. Since smaller regularization parameters typically lead to faster convergence, this topic plays a crucial role for the practical solution of GNEPs by ADMM-type methods.

We therefore present a modification of Algorithm 4.1 in Section 5.1 which uses an adaptive updating of the regularization parameter. Its convergence is discussed in 5.2 and heavily based on the existing results from the previous section.

## 5.1 Statement of Algorithm

The following method is a modification of Algorithm 4.1 which replaces the fixed regularization parameter by an adaptive updating procedure. In addition, it includes a second acceptance criterion for the new iterate which might be satisfied also for some small values of the regularization parameter. For a concise formulation of this modified algorithm, let us introduce the abbreviations

$$r^k := \sum_{\nu=1}^{N-1} \left\| g_\nu^k + \nabla_{x_\nu} \theta_\nu(x^k) + A_\nu^* \mu^k + s_\nu^k \right\|^2 + \|\gamma^{k-1}(x_N^k - x_N^{k-1})\|^2 + \left\| \sum_{\nu=1}^N A_\nu x_\nu^k - b \right\|^2, \quad (21)$$

for all  $k$ , with arbitrary elements  $g_\nu^k \in \partial_{x_\nu} \varphi_\nu(x^k)$  and a  $s_\nu^k \in N_{\mathcal{X}_\nu}(x^k)$ , see below for a discussion. To keep the notation simple, we also assume that all players take the same regularization parameter at each step, i.e.  $\gamma_1^k = \dots = \gamma_N^k =: \gamma^k$ .

### Algorithm 5.1. (ADMM Method with Adaptive Regularization)

(S.0) Choose a starting point  $(x^0, \mu^0) \in \mathcal{X} \times \mathcal{K}$ , parameters  $\alpha \in (0, 1)$ ,  $\beta, \gamma^0 > 0$ ,  $\tau > 1$ , a  $\Upsilon > 0$  satisfying (17), and set  $k := 0$ .

(S.1) If a suitable termination criterion is satisfied: STOP.

(S.2a) For  $\nu = 1, \dots, N$ , compute

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \theta_\nu(x_1^{k+1}, \dots, x_{\nu-1}^{k+1}, x_\nu, x_{\nu+1}^k, \dots, x_N^k) \right. \\ \left. + \langle \mu^k | A_\nu x_\nu \rangle_{\mathcal{K}} + \frac{\gamma^k}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 \right. \\ \left. + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{i=1}^{\nu-1} A_i x_i^{k+1} + \sum_{i=\nu+1}^N A_i x_i^k - b\|_{\mathcal{K}}^2 \right\}. \quad (22a)$$

(S.2b) For all  $\nu = 1, \dots, N-1$ , compute  $g_\nu^{k+1} \in \partial_{x_\nu} \varphi_\nu(x^{k+1})$  and a  $s_\nu^{k+1} \in N_{\mathcal{X}_\nu}(x^{k+1})$ .

(S.3) Define

$$\mu^{k+1} := \mu^k + \beta \left( \sum_{\nu=1}^N A_\nu x_\nu^{k+1} - b \right). \quad (22b)$$

(S.4) If  $\gamma^k \geq \Upsilon$  or

$$r^{k+1} \leq \alpha r^k \quad (\text{with } r^k \text{ defined in (21)}) \quad (23)$$

set  $\gamma^{k+1} := \gamma^k$

else

set  $\gamma^{k+1} := \gamma^k + \tau$ .

(S.5) Set  $k \leftarrow k + 1$ , and go to (S.1).

Notice that the main steps (22) of this algorithm are the same as (11) from Algorithm 4.1. Thus Lemma 4.2, Lemma 4.4, and Lemma 4.5 still hold. The idea of the method is to start the iteration with a small  $\gamma^0$  and to increase this regularization parameter only if this is really necessary. Note that an estimate  $\Upsilon$  for the lower bounds from (17) is still required, but not in such a crucial way as in Algorithm 4.1 where the corresponding lower bound directly influences the size of the regularization parameters, whereas here we can also take a bad overestimate for  $\Upsilon$  and still deal with relatively small regularization parameters.

This is due to the second acceptance criterion used in (S.4). This is a sufficient decrease condition which implies that there exists a sequence  $\varepsilon^{k+1} \in \mathcal{H} \times \mathcal{K}$  such that  $\varepsilon^{k+1} \rightarrow 0$  and  $\varepsilon^{k+1} \in T(w^{k+1})$ . Essentially (23) is a linear convergence condition whose satisfaction is reasonable to expect since our problem satisfies a strong convexity condition, and splitting-type algorithms are often linearly convergent under a strong convexity/monotonicity assumption.

In order to understand the definition of  $r^k$  used in the sufficient decrease condition (23), note that  $r^{k+1}$  represents the sum of the optimality conditions of the subproblems from (22a). The only term which needs some explanation is the difference  $\gamma^k(x_N^{k+1} - x_N^k)$  that occurs in  $r^{k+1}$ . We claim that this term represents the first-order optimality condition of (22a) for player  $\nu = N$  (that is why the first term in the definition of  $r^{k+1}$  is a sum from  $\nu = 1$  to  $N - 1$  only). In order to see this, note that this optimality condition is given by

$$0 = g_N^{k+1} + \nabla_{x_N} \theta_N(x^{k+1}) + A_N^* \mu^k + \gamma^k(x_N^{k+1} - x_N^k) + \beta A_N^*(Ax^{k+1} - b) + s_N^{k+1}$$

for suitable elements  $g_N^{k+1} \in \partial_{x_N} \varphi_N(x^{k+1})$  and a  $s_N^{k+1} \in N_{\mathcal{X}_N}(x^{k+1})$ . Using the definition (22b) of  $\mu^{k+1}$ , we therefore obtain

$$\begin{aligned} \gamma^k(x_N^k - x_N^{k+1}) &= g_N^{k+1} + \nabla_{x_N} \theta_N(x^{k+1}) + A_N^* \mu^k + \beta A_N^*(Ax^{k+1} - b) + s_N^{k+1} \\ &= g_N^{k+1} + \nabla_{x_N} \theta_N(x^{k+1}) + A_N^* \mu^{k+1} + s_N^{k+1}, \end{aligned} \quad (24)$$

which explains our definition of  $r^k$ .

Note that, for all other players  $\nu = 1, \dots, N - 1$ , Algorithm 5.1 requires the computation of arbitrary elements  $g_\nu^{k+1} \in \partial_{x_\nu} \varphi_\nu(x^{k+1})$  and a  $s_\nu^{k+1} \in N_{\mathcal{X}_\nu}(x^{k+1})$ . Apart from the fact that  $\varphi_\nu$  might be differentiable or the set  $\mathcal{X}_\nu$  is the entire space, so that these elements are uniquely defined, these elements do not need to be those particular ones that occur in the optimality conditions of the corresponding optimization problems of player  $\nu$ . These subgradients are used in our method only in the sufficient decrease condition, and there we can work with arbitrary elements.

Finally, let us note that we decided to use the update  $\gamma^{k+1} := \gamma^k + \tau$  for our algorithm. Alternatively, we could have taken the update  $\gamma^{k+1} := \tau \gamma^k$  for some  $\tau > 1$ , but our choice keeps  $\gamma^k$  smaller and seems to work quite well in numerical experiments.

## 5.2 Convergence

The global convergence of Algorithm 5.1 is heavily based on the known global convergence of Algorithm 4.1. Basically, we just have to verify that the additional decrease condition (23) does not destroy the global convergence properties from Theorem 4.6. To this end, let us take

a closer look at Algorithm 4.1. The sequence  $\{\gamma^k\}$  is monotonically increasing. Moreover, the updating  $\gamma^{k+1} = \gamma^k + \tau$  can occur only finitely many times due to the test  $\gamma^k \geq \Upsilon$  in (S.4). Hence it is clear that  $\gamma^k$  eventually stays fixed. For the convergence analysis, we can therefore assume that  $\gamma^k$  is a constant sequence, say  $\gamma^k = \gamma$  for all  $k$  sufficiently large.

Taking this into account, there are two possible situations: Either  $\gamma \geq \Upsilon$  or  $\gamma < \Upsilon$ . In the first case, we are precisely in the situation of the previous section, and convergence follows from Theorem 4.6. Hence it remains to consider the second case with  $\gamma < \Upsilon$ , which means that the sufficient decrease condition (23) is satisfied for all iterations  $k$  sufficiently large. This situation will be discussed separately in the following result.

**Proposition 5.2.** *Suppose that (1) admits a variational KKT point, that the sequence  $\{w^k\} = \{(x^k, \mu^k)\}$  is generated by Algorithm 5.1, that Assumption 3.8 holds, and that there is a  $k_0 \in \mathbb{N}$  such that (23) holds for all  $k \geq k_0$ . Then the sequence  $\{w^{k+1}\}$  converges weakly to a variational equilibrium of (1), and the iterates  $\{x_{-N}^{k+1}\}$  converge strongly.*

*Proof.* The proof is divided into four parts: In the first two parts, we will prove that  $\{\mu^{k+1}\}$  and  $\{x_N^{k+1}\}$  are Cauchy sequences and, therefore, (strongly) convergent. This we will use in part 3 to conclude that the sequence  $\{x_{-N}^{k+1}\}$  is strongly convergent. Finally, in part 4 we will show that every limit point of the sequence  $\{w^{k+1}\}$  is a variational equilibrium of (1). Throughout this proof, we denote by  $w^* = (x^*, \mu^*)$  an arbitrary variational KKT pair, whose  $x_{-N}^*$  part is unique due to Lemma 4.3.

*Part 1:* First notice that  $\sqrt{\alpha} \in (0, 1)$ . Taking the square root from (23), we obtain inductively

$$\begin{aligned} & \left( \sum_{\nu=1}^{N-1} \|g_\nu^{k+1} + \nabla_{x_\nu} \theta_\nu(x^{k+1}) + A_\nu^* \mu^{k+1} + s_\nu^{k+1}\|^2 + \|\gamma^k (x_N^{k+1} - x_N^k)\|^2 + \|Ax^{k+1} - b\|^2 \right)^{1/2} \\ & \leq \dots \leq \alpha^{k/2} r \end{aligned} \quad (25)$$

where  $r := r^1$  with  $r^1$  defined in (21); note that  $r$  is essentially a constant (just depending on  $x^1$  and  $x^0$ ). Summation yields

$$\begin{aligned} & \sum_{k=1}^{\infty} \|Ax^{k+1} - b\| \\ & \leq \sum_{k=1}^{\infty} \left( \sum_{\nu=1}^N \|g_\nu^{k+1} + \nabla_{x_\nu} \theta_\nu(x^{k+1}) + A_\nu^* \mu^{k+1} + s_\nu^{k+1}\|^2 + \|Ax^{k+1} - b\|^2 \right)^{1/2} \\ & \leq r \sum_{k=1}^{\infty} \sqrt{\alpha}^k = \frac{r}{1 - \sqrt{\alpha}}. \end{aligned} \quad (26)$$

Hence (22b) implies

$$\sum_{k=1}^{\infty} \|\mu^{k+1} - \mu^k\| = \beta \sum_{k=1}^{\infty} \|Ax^{k+1} - b\| \leq \beta \frac{r}{1 - \sqrt{\alpha}}.$$

Consequently, the triangle inequality shows that  $\{\mu^k\}$  a Cauchy sequence and, therefore, convergent to some element  $\bar{\mu}$ .

*Part 2:* Similar to the derivation of (26), we also obtain from (25) that

$$\sum_{k=1}^{\infty} \gamma^k \|x_N^{k+1} - x_N^k\| \leq \dots \leq \frac{r}{1 - \sqrt{\alpha}}.$$

Since  $\{\gamma^k\}$  is eventually constant, this implies that  $\{x_N^k\}$  is a Cauchy sequence, thus convergent to some element  $\bar{x}_N$ .

*Part 3:* Note that (23) or (25) together with (24) and the previous parts imply

$$g^{k+1} + P_{\theta}(w^{k+1}) + G(w^{k+1}) + s^{k+1} \rightarrow 0 \quad (27)$$

where  $g^{k+1} := (g_1^{k+1}, \dots, g_N^{k+1}, 0) \in \partial\psi(w^{k+1})$  and  $s^{k+1} := (s_1^{k+1}, \dots, s_N^{k+1}, 0) \in N_{\mathcal{W}}(w^{k+1})$  are chosen as in (S2b) or, for  $\nu = N$ , given by (24). Assume now that there is a subsequence  $\mathcal{I}$  and a  $c > 0$  such that  $\|x_{-N}^{k+1} - x_{-N}^*\| \geq c > 0$  for all  $k \in \mathcal{I}$ . Then (27), Lemma 4.2 and the boundedness of  $x_N^{k+1}, \mu^{k+1}$  imply

$$\rho \|x_{-N}^{k+1} - x_{-N}^*\| \leq \frac{\langle g^{k+1} + P_{\theta}(w^{k+1}) + G(w^{k+1}) + s^{k+1} \mid w^{k+1} - w^* \rangle}{\|x_{-N}^{k+1} - x_{-N}^*\|} \rightarrow_{\mathcal{I}} 0.$$

Therefore  $\|x_{-N}^{k+1} - x_{-N}^*\| \rightarrow_{\mathcal{I}} 0$ , a contradiction. Hence  $x_{-N}^{k+1} \rightarrow x_{-N}^*$  (recall that the  $x_{-N}^*$ -part is unique for all variational KKT points).

*Part 4:* We have already shown that  $x_{-N}^{k+1} \rightarrow x_{-N}^*$ ,  $x_N^{k+1} \rightarrow \bar{x}_N$ ,  $\mu^{k+1} \rightarrow \bar{\mu}$ . It remains to verify that  $\bar{w} := ((x_{-N}^*, \bar{x}_N), \bar{\mu})$  is a variational KKT pair. By the strong convergence of the block components, we have  $w^{k+1} - w^k \rightarrow 0$ . Therefore Assumption 3.8 (L) yields

$$\|\nabla_{x_{\nu}} \theta_{\nu}(\hat{x}^{\nu,k}) - \nabla_{x_{\nu}} \theta_{\nu}(x^{k+1})\|^2 \leq L_{\nu}^2 \sum_{i=\nu+1}^N \|x_i^{k+1} - x_i^k\|^2 \rightarrow 0$$

Since  $\gamma^k$  is eventually constant, we can use the linear operators defined in (13), (14) and obtain

$$\begin{aligned} \Delta^k &:= Q_N(w^{k+1} - w^k) + R_N(w^{k+1} - w^k) + \begin{pmatrix} \nabla_{x_1} \theta_1(\hat{x}^{1,k}) - \nabla_{x_1} \theta_1(x^{k+1}) \\ \vdots \\ \nabla_{x_{N-1}} \theta_{N-1}(\hat{x}^{N-1,k}) - \nabla_{x_{N-1}} \theta_{N-1}(x^{k+1}) \\ 0 \\ 0 \end{pmatrix} \\ &\rightarrow 0. \end{aligned}$$

In view of Lemma 4.4, we have  $-\Delta^k \in T(w^{k+1})$ . Using the strong-weak-sequential closedness (thus also strong-strong sequential closedness) of the graph of a maximally monotone operator and exploiting Proposition 3.7, we obtain  $0 \in T(\bar{w})$ . Hence  $\bar{w} = ((x_{-N}^*, \bar{x}), \bar{\mu})$  is a variational KKT pair, see (8).  $\square$

Combining Theorem 4.6 and Proposition 5.2, we obtain from the discussion at the beginning of this section the following convergence result for Algorithm 5.1.

**Theorem 5.3.** *Assume problem (1) admits a variational KKT point and that Assumption 3.8 holds. Further suppose that*

$$\Upsilon > \frac{1}{\rho} \sum_{i=1}^{N-1} L_i^2 + \frac{1}{\rho} \beta^2 N \sum_{i=1}^{N-1} C_i. \quad (28)$$

*Then the iterates  $w^{k+1}$  generated by the Algorithm 5.1 converge weakly to a variational equilibrium  $w^*$  of (1). Furthermore,  $\{x_{-N}^k\}$  converges strongly to  $x_{-N}^*$ .*

## 6 Applications and Numerical Results

This section presents some applications and numerical results. Since Algorithm 5.1 is usually much faster convergent than Algorithm 4.1, we concentrate on this method. The aim is to demonstrate that the ADMM-type approach sometimes terminates after a suprisingly small number of iterations despite the fact that the global convergence rate of ADMM-like schemes is usually not that fast.

Since our examples include (linear) inequality constraints, let us first note that we can also handle problems like

$$\min_{x_\nu \in \mathcal{X}_\nu} \theta_\nu(x_\nu, x_{-\nu}) + \varphi_\nu(x_\nu) \quad \text{s.t.} \quad \sum_{\mu=1}^N B_\mu x_\mu - b \in C \quad (29)$$

with  $C$  being a convex cone. In fact, using a slack variable, it was already noted in [3] that it is possible to reformulate such problems into our framework from (1) with linear equality constraints. This allows the application of our methods also to problems of the form (29).

We stop Algorithm 5.1 as soon as the value of  $r^k$  as defined in (21) is less than  $10^{-8}$ , which corresponds to the variational KKT conditions being satisfied up to an accuracy of  $10^{-4}$  in the  $\mathcal{H} \times \mathcal{K}$ -norm. Notice that, in general, this termination criterion is not guaranteed to work due to the fact that the subgradients  $s^k$  and  $g^k$ , which occur in the definition of  $r^k$ , are taken arbitrarily and might not be the correct ones which imply that  $r^k$  tends to zero. On the other hand, if  $r^k$  converges to zero, this is clearly an indication that we are close to a variational KKT point. In our numerical experiments, we did not observe any difficulties with this termination criterion.

The subproblems (22a) were solved using the MATLAB<sup>®</sup> function `fmincon`, with the gradient and Hessian user supplied. The `OptimalityTolerance` and `ConstraintTolerance` were left on the default values  $10^{-6}$ . All test runs use the following parameters:

$$\alpha := 0.99999, \quad \beta := 1000, \quad \tau := 1.$$

The implementation applies a slightly different update of the parameter  $\gamma^k$ : Whenever  $\gamma^k$  got increased at some iteration  $k$ , we kept  $\gamma^k$  constant at this level for at least ten subsequent

iterations. According to our experience, this improves the numerical behaviour and does not change the convergence theory at all since either  $\gamma^k$  reaches the bound  $\Upsilon$  and the theory from Theorem 4.6 applies, or (23) is satisfied for almost all iterations and the theory from Proposition 5.2 holds.

Finally, we initialize Algorithm 5.1 using  $\gamma^0 := 0.1$  and the starting point  $(x^0, \mu^0) := (0, 0)$ .

## 6.1 Elliptic Optimal Control GNEPs

In this section we want to solve a class of examples that was previously used and discussed in [22, 24, 25], where different solution methods are considered.

In order to describe the example, we slightly change our notation in this section to be consistent with the standard notation used in the optimal control setting. The players' strategies  $x_\nu \in \mathcal{X}_\nu$  are now called the controls and denoted by  $u_\nu \in L^2(\Omega)$ ,  $\Omega$  being a suitable and sufficiently smooth domain in  $\mathbb{R}^d$ . The so-called state variable  $y \in H_0^1(\Omega)$  is the solution of an elliptic partial differential equation that depends on the players' strategies  $u = (u_1, \dots, u_N) \in L^2(\Omega)^N$ . We then consider the optimal control generalized Nash problem

$$\min_{u_\nu \in L^2(\Omega)} \left\{ \frac{1}{2} \|y(u_\nu, u_{-\nu}) - y_\nu^d\|_{L^2(\Omega)}^2 + \frac{\alpha_\nu}{2} \|u_\nu\|_{L^2(\Omega)}^2 \right\} \quad (30a)$$

$$\begin{aligned} \text{s.t.} \quad & -\Delta y = \sum_{\nu=1}^N u_\nu && \text{in } \Omega, \\ & y = 0 && \text{in } \partial\Omega, \\ & u_\nu(x) \in [u_\nu^a(x), u_\nu^b(x)] && \text{f.a.a. } x \in \Omega, \\ & y(x) \leq \psi(x) && \text{f.a.a. } x \in \Omega. \end{aligned} \quad (30b)$$

Hence we have a tracking-type objective function for each player  $\nu = 1, \dots, N$ , pointwise lower and upper bounds on the controls  $u_\nu$ , and an additional upper bound on the state.

Using the control-to-state-map

$$S : L^2(\Omega)^N \rightarrow H_0^1(\Omega) \cap C(\overline{\Omega}), \quad u \mapsto y, \quad Su = \sum_{\nu=1}^N S_\nu u_\nu,$$

where the last expression uses the linearity of the solution mapping  $S$ , we can rewrite (30) as

$$\min_{u_\nu \in L^2(\Omega)} \left\{ \frac{1}{2} \left\| \sum_{\nu=1}^N S_\nu u_\nu - y_\nu^d \right\|_{L^2(\Omega)}^2 + \frac{\alpha_\nu}{2} \|u_\nu\|_{L^2(\Omega)}^2 \right\} \quad \text{s.t.} \quad u_\nu \in [u_\nu^a, u_\nu^b], \quad \sum_{\nu=1}^N S_\nu u_\nu \leq \psi \quad (31)$$

and therefore obtain a GNEP of the form (29) by taking

$$\begin{aligned} \theta_\nu(u) &= \frac{1}{2} \left\| \sum_{\nu=1}^N S_\nu u_\nu - y_\nu^d \right\|_{L^2(\Omega)}^2, & \varphi_\nu(u_\nu) &= \frac{\alpha_\nu}{2} \|u_\nu\|_{L^2(\Omega)}^2, \\ \mathcal{X}_\nu &= \{u_\nu \in L^2(\Omega) \mid u_\nu(x) \in [u_\nu^a(x), u_\nu^b(x)] \text{ f.a.a. } x \in \Omega\}, \\ B_\nu &= S_\nu, & b &= \psi. \end{aligned}$$



It was shown in [24, Lem. 6.2] that the operator  $\partial\varphi + \hat{P}_\theta$  is strongly monotone with modulus  $\min_\nu \alpha_\nu$ . If  $\Omega$  is contained in a cube with edges of length  $r$  then in [4, Prop. 7.2] the operator-norm of  $S_\nu : L^2(\Omega) \rightarrow H_0^1(\Omega)$  was estimated as  $\|S_\nu\| \leq r^2$  where  $r$  is the length of one side of this cube. (Notice that we need the norm of  $S_\nu : L^2(\Omega) \rightarrow H_0^1(\Omega)$  and not of  $S_\nu : H^{-1}(\Omega) \rightarrow H_0^1(\Omega)$  which is one, since  $S_\nu$  is as isometric isomorphism between these spaces in the  $L^2$ -scalar product.)

We implemented the elliptic optimal control GNEP presented above with the same functions and parameters as [22, 24, 25] in MATLAB®, i.e.  $N = 4$ ,  $\Omega = (0, 1)^2$ ,  $\alpha = (2.8859, 4.3374, 2.5921, 3.9481)$

$$y_1^d = \xi_1 - \xi_4, \quad y_2^d = \xi_2 - \xi_3, \quad y_3^d = \xi_3 - \xi_2, \quad y_4^d = \xi_4 - \xi_1$$

where

$$\xi_i(x) = 10^3 \max \left( 0, 4 \left( \frac{1}{4} - \max(|x_1 - z_i^1|, |x_2 - z_i^2|) \right) \right),$$

and  $z^1 = (0.25, 0.75, 0.25, 0.75)$ ,  $z^2 = (0.25, 0.25, 0.75, 0.75)$ . The constraint function is

$$\psi(x) := \left( \cos \left( 5 \sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2} \right) + 0.1 \right).$$

The numerical results for different discretization widths are given in Table 1. Notice that, for this example, there was no need to increase  $\gamma^k$  at any iteration, keeping the total number of iterations very small for essentially all levels of discretization.

Table 1: Number of iterations of Algorithm 5.1 for the elliptic optimal control GNEP.

Discretization	16	32	64	128	256
Iterations	94	95	96	142	137
Final $\gamma$	0.1	0.1	0.1	0.1	0.1

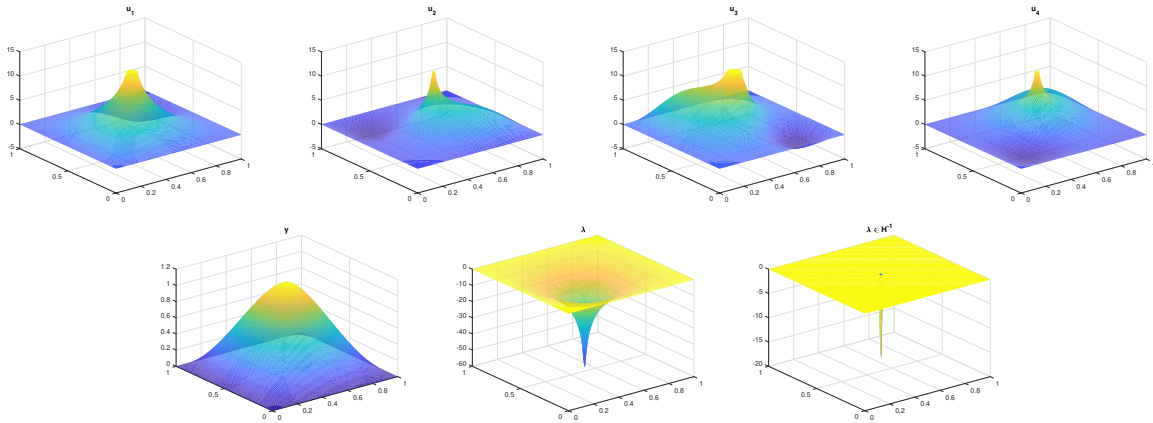


Figure 2: The optimal controls  $u_1, \dots, u_4$ , the state, and the Lagrange multiplier in the isomorphic spaces  $H_0^1(\Omega)$  and  $H^{-1}(\Omega)$ , respectively.

## 6.2 Environmental Differential Games

Another infinite-dimensional GNEP results from a class of differential games, which are quite popular in the literature, cf. [16, 17, 27]. We use an example from [24], where a more detailed discussion can be found. The only modification we made is that we changed some constants such that the problem fits into our framework. The problem of player  $\nu$  is given by

$$\begin{aligned} \min_{u_\nu \in L^2(0,1), y_\nu \in H^1(0,1)} & \int_0^1 \left[ q_\nu(u_\nu(t), y_\nu(t)) + r_\nu(y_1(t), \dots, y_N(t)) \right] dt \\ \text{s.t.} & \sum_{\mu=1}^N e_\mu y_\mu(t) \leq \psi(t), \end{aligned} \quad (32a)$$

$$u_\nu(t) \in [0, u_\nu^{max}(t)], \quad (32b)$$

$$\dot{y}_\nu(t) + b_\nu y_\nu(t) = u_\nu(t), \quad (32c)$$

$$y_\nu(0) = y_\nu^0. \quad (32d)$$

We choose  $N = 2$ ,

$$q_\nu(u_\nu(t), y_\nu(t)) = \frac{a_1}{2} y_\nu(t)^2 + \frac{a_2}{2} u_\nu(t)^2, \quad r_\nu(y_1(t), \dots, y_N(t)) = \frac{-c \cdot y_\nu(t)}{\sum_{i=1}^N y_i(t) + \varepsilon}$$

the constants  $a_1 = 0.7$ ,  $a_2 = 0.6$ ,  $b_1 = 0.2$ ,  $b_2 = 0.6$ ,  $e_1 = 2$ ,  $e_2 = 1$ ,  $\varepsilon = 10^{-9}$ ,  $y_1^0 = 0$ ,  $y_2^0 = 1$ ,  $u_1^{max}(t) = u_2^{max}(t) = 0.2$ , as well as the mapping

$$\psi(t) = \begin{cases} 1.0 & \text{if } t \in [0, 1], \\ 0.8 & \text{if } t \in (1, 2], \\ 0.6 & \text{if } t \in (2, 3]. \end{cases}$$

To satisfy the constraints (32c) and (32d), we use an affine linear solution operator  $S$  such that  $Su_\nu = y_\nu$ , as in Section 6.1. The box constraints (32b) represent our set  $\mathcal{X}_\nu$ , while (32a) will be enforced through the augmented Lagrangian approach.

The numerical results for this example, using different discretization levels of this differential game, are summarized in Table 2. Note that the final value of  $\gamma^k$  is still quite small. The corresponding solutions are depicted in Figure 3.

Table 2: Number of iterations of Algorithm 5.1 applied to the environmental differential game.

Discretization	128	256	512	1024	2048
Iterations	871	944	873	1085	1208
Final $\gamma$	4.1	4.1	2.1	4.1	5.1

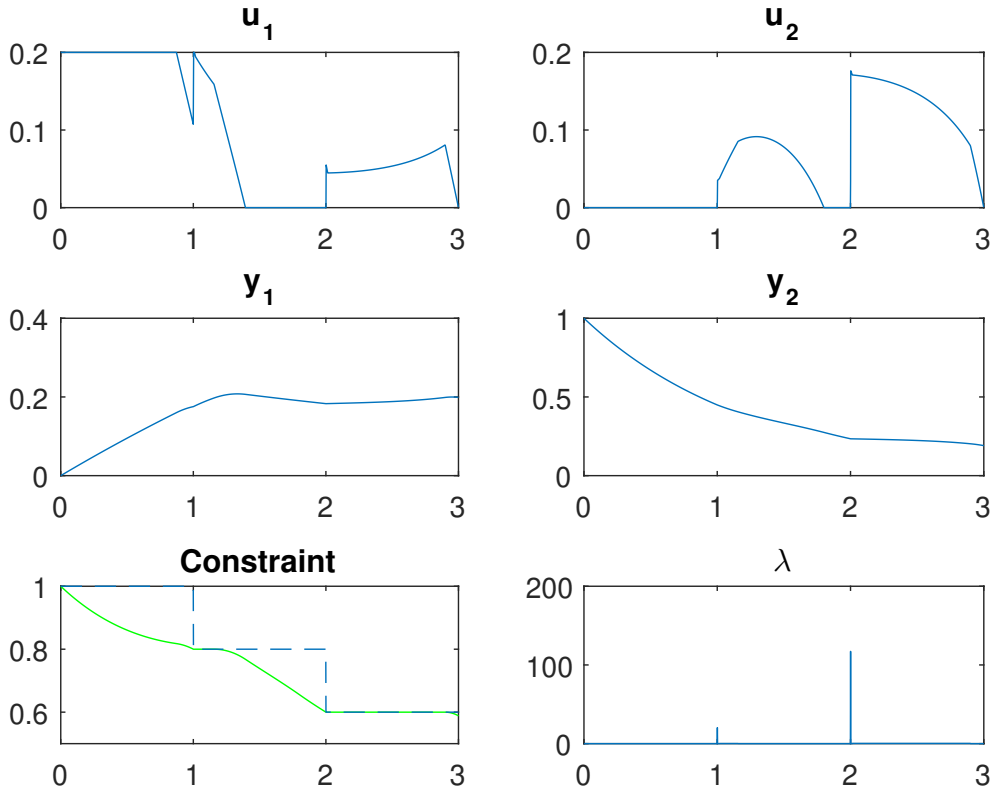


Figure 3: Optimal controls  $u_1, u_2$ , optimal states  $y_1, y_2$ , satisfaction of the constraint, and Lagrange multiplier for the environmental differential game.

## 7 Final Remarks

We presented two methods for solving linearly constrained generalized Nash equilibrium problems. These methods were motivated by the very popular alternating direction method of multipliers (ADMM) for the solution of certain structured optimization problems. The convergence rate of ADMM methods is usually slow (sublinear), but they can often be applied to large-scale problems since they typically have to solve only small-dimensional optimization problems at each iteration.

There are several possible modifications of Algorithm 4.1 or Algorithm 5.1 which might be of interest for their own. For example, the following changes of step (11a) or (22a) might be investigated in a way similar to the analysis given in this paper:

- A linearized Gauss-Seidel-ADMM-type modification

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \langle \nabla_{x_\nu} \theta_\nu(x_1^{k+1}, \dots, x_{\nu-1}^{k+1}, x_\nu^k, x_{\nu+1}^k, \dots, x_N^k) \mid x_\nu - x_\nu^k \rangle \right. \\ \left. + \langle \mu^k \mid A_\nu x_\nu \rangle_{\mathcal{K}} + \frac{\gamma_\nu}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{i=1}^{\nu-1} A_i x_i^{k+1} + \sum_{i=\nu+1}^N A_i x_i^k - b\|_{\mathcal{K}}^2 \right\}.$$

This method differs from (11a) by replacing the function  $\theta_\nu$  by its (Gauss-Seidel-) linearization, hence the subproblems to be solved at each iteration become even simpler.

- A linearized Gauss-Seidel-Jacobi-ADMM-type modification

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \langle \nabla_{x_\nu} \theta_\nu(x^k) \mid x_\nu - x_\nu^k \rangle \right. \\ \left. + \langle \mu^k \mid A_\nu x_\nu \rangle_{\mathcal{K}} + \frac{\gamma_\nu}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{i=1}^{\nu-1} A_i x_i^{k+1} + \sum_{i=\nu+1}^N A_i x_i^k - b\|_{\mathcal{K}}^2 \right\}.$$

This method differs from the previous one by using the old iterate  $x^k$  in the linearization of  $\theta_\nu$ , whereas the penalty term includes the new information  $x_i^{k+1}$  for all  $i = 1, \dots, \nu-1$ .

- The Jacobi-ADMM-type modification

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \theta_\nu(x_\nu, x_{-N}^k) + \langle \mu^k \mid A_\nu x_\nu \rangle_{\mathcal{K}} \right. \\ \left. + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{\substack{i=1 \\ i \neq \nu}}^N A_i x_i^k - b\|_{\mathcal{K}}^2 + \frac{\gamma_\nu}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 \right\}.$$

This updating corresponds to (11a) except that  $x_i^{k+1}$  is replaced everywhere by  $x_i^k$  for  $i = 1, \dots, \nu-1$ , leading to a method that is fully parallel.

- The linearized Jacobi-ADMM-type modification

$$x_\nu^{k+1} := \arg \min_{x_\nu \in \mathcal{X}_\nu} \left\{ \varphi_\nu(x_\nu) + \langle \nabla_{x_\nu} \theta_\nu(x^k) \mid x_\nu - x_\nu^k \rangle + \langle \mu^k \mid A_\nu x_\nu \rangle_{\mathcal{K}} \right. \\ \left. + \frac{\beta}{2} \|A_\nu x_\nu + \sum_{\substack{i=1 \\ i \neq \nu}}^N A_i x_i^k - b\|_{\mathcal{K}}^2 + \frac{\gamma_\nu}{2} \|x_\nu - x_\nu^k\|_{\mathcal{H}_\nu}^2 \right\}$$

which results from the previous method by linearization of the mapping  $\theta_\nu$ .

We believe that it is possible to extend our convergence theory to the above modifications, but leave the details for our future research. We further note that the linearized Jacobi-ADMM-type method is essentially the method discussed in [3], where convergence was shown under suitable assumptions by an interpretation of this approach as a forward-backward splitting method.

## References

- [1] H. Attouch and M. Soueycatt. Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control. *Pac. J. Optim.*, 5(1):17–37, 2009.
- [2] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Second edition.
- [3] E. Börgens and C. Kanzow. A distributed regularized Jacobi-type ADMM-method for generalized Nash equilibrium problems in Hilbert spaces. *Numer. Funct. Anal. Optim.*, 39(12):1316–1349, 2018.
- [4] E. Börgens and C. Kanzow. Regularized Jacobi-type ADMM-methods for a class of separable convex optimization problems in Hilbert spaces. *Comput. Optim. Appl.*, 73(3):755–790, 2019.
- [5] A. Borzì and C. Kanzow. Formulation and numerical solution of Nash equilibrium multiobjective elliptic control problems. *SIAM J. Control Optim.*, 51(1):718–744, 2013.
- [6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Machine Learning*, 3(1):1–122, 2011.
- [7] L. M. Briceño Arias and P. L. Combettes. Monotone operator methods for Nash equilibria in non-potential games. In *Computational and analytical mathematics*, volume 50 of *Springer Proc. Math. Stat.*, pages 143–159. Springer, New York, 2013.
- [8] C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Program.*, 155(1-2, Ser. A):57–79, 2016.
- [9] W. Deng, M.-J. Lai, Z. Peng, and W. Yin. Parallel multi-block ADMM with  $o(1/k)$  convergence. *J. Sci. Comput.*, 71(2):712–736, 2017.
- [10] A. Dreves. A Nash equilibrium approach for multiobjective optimal control problems with elliptic partial differential equations. *Control Cybernet.*, 45(4):457–482, 2016.
- [11] A. Dreves and J. Gwinner. Jointly convex generalized Nash equilibria and elliptic multiobjective optimal control. *J. Optim. Theory Appl.*, 168(3):1065–1086, 2016.
- [12] J. Eckstein and W. Yao. Understanding the convergence of the alternating direction method of multipliers: theoretical and computational perspectives. *Pac. J. Optim.*, 11(4):619–644, 2015.
- [13] F. Facchinei and C. Kanzow. Generalized Nash equilibrium problems. *Ann. Oper. Res.*, 175:177–211, 2010.

- [14] F. Facchinei, V. Piccialli, and M. Sciandrone. Decomposition algorithms for generalized potential games. *Comput. Optim. Appl.*, 50(2):237–262, 2011.
- [15] A. Fischer, M. Herrich, and K. Schönefeld. Generalized Nash equilibrium problems - recent advances and challenges. *Pesquisa Operacional*, 34(3):521–558, 2014.
- [16] A. Friedman. *Differential Games*. Wiley-Interscience, New York-London, 1971.
- [17] T. L. Friesz. *Dynamic Optimization and Differential Games*, volume 135 of *International Series in Operations Research & Management Science*. Springer Science & Business Media, 2010.
- [18] B. He, L. Hou, and X. Yuan. On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming. *SIAM J. Optim.*, 25(4):2274–2312, 2015.
- [19] B. He, M. Tao, and X. Yuan. Alternating direction method with Gaussian back substitution for separable convex programming. *SIAM J. Optim.*, 22(2):313–340, 2012.
- [20] B. He, H.-K. Xu, and X. Yuan. On the proximal Jacobian decomposition of ALM for multiple-block separable convex minimization problems and its relationship to ADMM. *J. Sci. Comput.*, 66(3):1204–1217, 2016.
- [21] B.-S. He. PPA-like contraction methods for convex optimization: a framework using variational inequality approach. *J. Oper. Res. Soc. China*, 3(4):391–420, 2015.
- [22] M. Hintermüller and T. Surowiec. A PDE-constrained generalized Nash equilibrium problem with pointwise control and state constraints. *Pac. J. Optim.*, 9(2):251–273, 2013.
- [23] M. Hintermüller, T. Surowiec, and A. Kämmler. Generalized Nash equilibrium problems in Banach spaces: theory, Nikaido-Isoda-based path-following methods, and applications. *SIAM J. Optim.*, 25(3):1826–1856, 2015.
- [24] C. Kanzow, V. Karl, D. Steck, and D. Wachsmuth. The multiplier-penalty method for generalized Nash equilibrium problems in Banach spaces. *SIAM J. Optim.*, 29(1):767–793, 2019.
- [25] C. Kanzow and D. Steck. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.*, 26(4):2034–2058, 2016.
- [26] C. Kanzow and D. Steck. Quasi-variational inequalities in Banach spaces: theory and augmented Lagrangian methods. *ArXiv e-prints*, December 2018. ArXiv:1810.00406.
- [27] K. M. Ramachandran and C. P. Tsokos. *Stochastic Differential Games, Theory and Applications*. Atlantis Press, Paris, 2012.
- [28] F. Salehisadaghiani and L. Pavel. Generalized Nash equilibrium problem by the alternating direction method of multipliers. *arXiv preprint arXiv:1703.08509*, 2017.

- [29] M. Tao and X. Yuan. Convergence analysis of the direct extension of ADMM for multiple-block separable convex minimization. *Adv. Comput. Math.*, 44(3):773–813, 2018.
- [30] X. Wang, M. Hong, S. Ma, and Z.-Q. Luo. Solving multiple-block separable convex minimization problems using two-block alternating direction method of multipliers. *Pac. J. Optim.*, 11(4):645–667, 2015.