Deutsche Forschungsgemeinschaft
Priority Programme 1962

Solving Quadratic Multi-Leader-Follower Games by Smoothing the Follower's Best Response

Michael Herty, Sonja Steffensen, Anna Thünen

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-070

received on August 23, 2018

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Solving Quadratic Multi-Leader-Follower Games by Smoothing the Follower's Best Response

Michael Herty, Sonja Steffensen, and Anna Thünen

August 23, 2018

Abstract

We derive Nash-s-stationary equilibria for a class of quadratic multi-leader-follower games using the nonsmooth best response function. To overcome the challenge of nonsmoothness, we pursue a smoothing approach resulting in a reformulation as smooth Nash equilbrium problem. We prove existence and uniqueness for all smoothing parameters. For a decreasing sequence of these smoothing parameters accumulation points of Nash equilibria exist and we show that they fullfill the conditions of s-stationarity. Finally, we propose an update on the leader variables for efficient computation and numerically compare nonsmooth Newton and subgradient method.

Keywords: Multi-Leader-Follower Games, Nash Equilibria, Nonsmooth Newton, Subgradient Method, Game Theory, Equilibrium Problems with Equilibrium Constraints

AMS-MSC2010: 91A06, 91A10, 90C33, 91A65, 49J52

1 Introduction and Background

The multi-leader-follower game (MLFG) is a particular class of problems in classical game theory. These models serve as an analytical tool to study strategic behavior of individuals in a noncooperative manner. In particular, the individuals (so-called players) are divided into two groups, namely leaders and followers, according to their position in the game. Mathematically, this yields a hierarchical Nash game, where further minimization problems appear in the participants' optimization problems as constraints. An equilibrium is then given by a multistrategy vector of all players, where no player has the incentive to change his chosen strategy unilaterally.

Most recently, such type of models gained an increasing interest among mathematicians as well as scientists in other fields such as operations research, robotics, and computer science [2, 21, 14]. However, compared to the knowledge of other classical game models so far little is known concerning existence and uniqueness theory as well as suitable numerical solution methods.

The structure of MLFG can be seen as equilibrium problem with equilibrium constraints (EPEC), as the optimization problems of the followers might be replaced by the corresponding optimality conditions. Recently, the competition on electric power market is described by EPEC [1, 3, 4, 13, 15]. Here, the leaders are power generators and consumers, who bid their cost and utility functions. The single follower plays the role of an independent system operator coordinating dispatch and minimizing social costs subject to network constraints.

So far, there exist only a few recent theoretical results for MLFG or EPEC, analyzing the existence, the uniqueness and characterizations of equilibria: Pang and Fukushima [17] present a reformulation as generalized Nash equilibrium problem and quasi-variational inequality. In another work, Hu and Ralph [13] formulate a sufficient condition for the existence of pure strategy Nash equilibria in a particular case of a MLFG related to an electricity market model. Leyffer and Munson [15] describe various reformulations in terms of MPEC, nonlinear programs, and nonlinear complementarity problems. Further discussions on MLFG and EPEC can be found in [6, 11, 20]. More recent work has been done by Hu and Fukushima in [12]. Therein, they discuss the existence of robust Nash equilibria of a class of quadratic MLFG. Further, they propose a uniqueness result for a MLFG with two leaders.

In this paper we study a quadratic MLFG with similarities to the model studied in [12] and generalize the follower's strategy set by allowing inequality constraints. This generalization translates into equilibrium conditions or to a nonsmooth Nash game formulation. With suitable convexity assumptions, existence of Nash equilibria can be proved. Furthermore, we propose a smoothed Nash game formulation and prove uniqueness of Nash equilibria for an arbitrary number of players. Besides the theoretical results, we propose an algorithm to numerically compute Nash equilibria. Therein, we combine an update based on a Taylor expansion of the parameter dependent solution and the computation of Nash equilibria of approximating problems.

This article is structured as follows: In Section 2, we introduce the quadratic MLFG and give two motivating examples. Then we develop an equivalent nonsmooth Nash equilibrium problem (NEP) for which we can prove existence of equilibria and uniqueness. In Section 3, smoothing of the best response of the follower leads to a differentiable NEP formulation. In Section 4, we characterize the Nash equilibria of the smoothed problem by KKT conditions. For decreasing smoothing parameter, the limit of the Nash equilibria fulfills a necessary optimality condition. In Section 5, we introduce our general approach to compute Nash equilibria. In Section 6, we present numerical results of the proposed methods.

2 Existence of Nash Equilibria of MLFG

We consider a MLFG, where the follower's game is given by:

$$\min_{y \in \mathbb{R}^m} \Theta(y, x) = \frac{1}{2} y^\top Q_y y - b(x)^\top y \quad \text{s.t.} \quad y \geq l(x) \tag{1}$$

where $Q_y \in \mathbb{R}^{m \times m}$ is a positive definite diagonal matrix and $b, l : \mathbb{R}^n \to \mathbb{R}^m$ convex and differentiable functions.

Furthermore, the leader problems are given for $\nu = 1, \ldots, N$ by:

$$\min_{x_\nu \in \mathbb{R}^{n_\nu}} \theta_\nu(x_\nu, x_{-\nu}) = \frac{1}{2} x_\nu^\top Q_\nu x_\nu + a^\top y \quad \text{s.t.} \quad x_\nu \in X_\nu \tag{2}$$

with nonempty, convex, and closed strategy sets X_ν which we assume to be described by smooth functions $g_\nu : \mathbb{R}^{n_\nu} \to \mathbb{R}^{m_\nu}$ such that $X_\nu = \{x_\nu \in \mathbb{R}^{n_\nu} | g_\nu(x_\nu) \le 0\}$. The objective θ_ν is strictly convex quadratic with $Q_\nu \in \mathbb{R}^{n_\nu \times n_\nu}$ symmetric positive definite and $a \in \mathbb{R}^m_+$. We denote $x = (x_\nu)_{\nu=1}^N \in \mathbb{R}^n$ and $x_{-\nu} = (x_i)_{i=1,i\neq\nu}^N \in \mathbb{R}^{n-n_\nu}$.

2.1 Motivating Examples

Harks et al. [10] modeled a routing game with parallel links, streets, connecting a common source to a common destination. The street owners aiming to maximize their revenue by tolls are modeled as the set of leaders. In our notation, we write for each leader $\nu = 1, \dots, N$:

$$\max_{x_\nu \in \mathbb{R}} \frac{1}{2} x_\nu^\top Q_\nu x_\nu \quad \text{s.t. } g_\nu(x_\nu) = x_\nu - c \le 0$$

where the Hessian in [10] is $Q_\nu = 2$ for all ν and $c \in \mathbb{R}$ is a toll cap set by the authorities. In [10], the toll cap is determined by an outer Stackelberg game. This maximization is different to (1) could be modeled as cost minimization in a similar spirit. The commuters are modeled as follower. Their goal is to minimize linear latency cost plus toll cost. This translates to a quadratic follower problem:

$$\min_{y \in \mathbb{R}^m} \frac{1}{2} y^\top Q_y y - b(x)^\top y \quad \text{s.t. } y \ge l(x) = 0, \ \sum_{i=1}^m y_i = 1$$

where $b(x) = -d - x$ and $Q_y = \operatorname{diag}(2q_1, \dots, 2q_n)$. The data d and q are given by the latency functions of the commuters $c(y_i) = q_i y_i + d_i$. In contrast to our problem formulation, b is assumed to be affine instead of linear. This example shows that a diagonal positive definite Q_y is a reasonable assumption.

A second example is presented as electricity market model in [3, 4]. Here, the leader problems represent profit maximization of electricity producers and are given for $\nu = 1, \dots, N$:

$$\min_{x_\nu \in \mathbb{R}^2} \begin{bmatrix} -y_\nu \\ -y_\nu^2 \end{bmatrix}^\top x_\nu + a^\top \begin{bmatrix} y_\nu \\ y_\nu^2 \end{bmatrix} \quad \text{s.t. } g_\nu(x_\nu) = -x_\nu \le 0$$

These leader problems are different to our model since its objective is linear. There is also a bilinear term for a single leader and a follower variable, they do not allow for dependence on other leaders. The follower problem represents an independent system operator aiming to minimize the total cost and assigning the quantity required. Therefore the dimension of the follower variable is equal to the number of leaders.

$$\min_{y \in \mathbb{R}^N} \frac{1}{2} y^\top Q_y y - b(x)^\top y \quad \text{s.t. } y \ge l(x) = 0, \ \sum_{i=1}^N y_i \ge D$$

where b and Q_y are as in the toll model. The last constraints matches the demand D to the required electricity produced.

Overall, these examples illustrate the importance of investigations on quadratic MLFG.

2.2 Theoretical Results

Since the follower's problem has a strictly convex objective $\Theta(y, x)$ and a convex strategy set, we state its unique solution in the following lemma:

Lemma 2.1 (Follower's Best Response)**.** *The follower's optimization problem* (1) *has a unique solution y^* for any given leader strategy vector $x \in \mathbb{R}^n$ if in y^* Guignard constraint qualification holds. In particular, this best response function is:*

$$y^*(x) = \max\left\{Q_y^{-1} b(x), l(x)\right\} \tag{3}$$

which is convex.

Proof. The objective and the feasible set are convex for any $x \in \mathbb{R}^n$ and we assume that Guignard constraint qualification [9] holds in y^*. Therefore the KKT conditions are necessary and sufficient for a global minimizer. Since the objective is strictly convex the minimizer is unique for all leader strategies x.

To derive the structure, we apply the KKT conditions: There exist Lagrange multipliers $\lambda \in \mathbb{R}^m$ such that:

$$\begin{aligned} 0 &= Q_y y - b(x) - \lambda \\ 0 &\leq \lambda \perp y - l(x) \geq 0 \end{aligned}$$

Combining these expressions yields: $0 \leq \lambda = Q_y y - b(x) \perp y - l(x) \geq 0$. We assumed Q_y to be positive definite and diagonal, therefore we can write equivalently: $0 \leq y - Q_y^{-1} b(x) \perp y - l(x) \geq 0$ and $0 = \min\left\{y - Q_y^{-1} b(x), y - l(x)\right\}$, respectively. We obtain (3) by extracting y and changing min to max. The best response function $y^*(x)$ is convex in x as a maximum of convex functions. $\square$

We remark that the min and max operator are applied componentwise to a vector.

The MLFG is formulated as a Nash game by plugging the best response (3) in the leader game (2) for $\nu = 1, \ldots, N$:

$$\min_{x_\nu} \frac{1}{2} x_\nu^\top Q_\nu x_\nu + \sum_{i=1}^{m} a_i \max\left\{\left(Q_y^{-1} b(x)\right)_i, l_i(x)\right\} \quad \text{s.t.} \quad x_\nu \in X_\nu \tag{NEP}$$

Each optimization problem has a nonsmooth but convex objective and a convex strategy set. For compact strategy sets, we can prove existence of Nash equilibria in the following theorem:

Theorem 2.2 (Existence of Nash Equilibria for Compact Strategy Sets)**.** *Assume that the nonsmooth Nash equilibrium problem in* (NEP) *has a convex and compact joint strategy set $X = X_1 \times \cdots \times X_N$, where all X_ν are nonempty. Then there exists at least one Nash equilibrium.*

Therefore also the quadratic multi-leader-follower game given by (1) and (2) has at least one Nash equilibrium.

Proof. We formulated the MLFG as a convex Nash equilibrium problem (NEP), especially the objectives are continuous in $(x_\nu, x_{-\nu})$ and convex in x_ν because they are a sum of a strictly convex quadratic term and the maximum of two convex functions. Furthermore, we assumed the admissible strategy sets X_ν to be nonempty, convex, and compact. Therefore the conditions of [16, Theorem 3.1] are fulfilled. $\square$

In the following, we give an example of a MLFG where Nash equilibria can be explicitly computed. This aims to illustrate that particularly the compactness assumption of the previous theorem is a sufficient condition but not necessary.

In the remainder of this article, we assume that b and l are linear, so we can write:

$$b(x) = B^\top x \quad \text{and} \quad l(x) = L^\top x$$

where $B, L \in \mathbb{R}^{n \times m}$, $B_{:,i}, L_{:,i} \in \mathbb{R}^n$ denotes the i-th column, and $B_{\nu,:}, L_{\nu,:}$ denote their submatrices of the rows referring to x_ν.

We consider the MLFG in (1-2) and assume strategy set to be the nonnegative orthant, i.e. $X_\nu = \mathbb{R}^{n_\nu}_+$. Since this strategy set is not compact, existence of Nash equilibria is not immediately obvious. The resulting Nash equilibrium problem is written as follows for $\nu = 1, \ldots, N$:

$$\begin{aligned} \min_{x_\nu} \; & \frac{1}{2} x_\nu^\top Q_\nu x_\nu + \sum_{i=1}^{m} a_i \left(\max \left\{ \left(L^\top - Q_y^{-1} B^\top \right)_{:,i} x, 0 \right\} + (Q_y^{-1} B^\top)_{:,i} x \right) \\ \text{s.t. } & x_\nu \in \mathbb{R}^{n_\nu}_+ \end{aligned} \tag{4}$$

We are interested in the first derivative of the objective function in order to state optimality conditions. Since we assume non-negativity of x, some parts of the sum are differentiable and others are not, depending on the entries of the matrix $\left(L^\top - Q_y^{-1} B^\top\right)$. Therefore we split the sum in the objective to emphasize accordingly:

$$\begin{aligned} I_{\geq} &= \left\{ i \in \{1, \ldots, m\} \,\middle|\, \left(L^\top - Q_y^{-1} B^\top \right)_{:,i} \geq 0 \right\} \\ I_{<} &= \left\{ i \in \{1, \ldots, m\} \,\middle|\, 0 \neq \left(L^\top - Q_y^{-1} B^\top \right)_{:,i} \leq 0 \right\} \\ I_{\text{ns}} &= \{1, \ldots, m\} \setminus (I_{\geq} \cup I_{<}) \end{aligned}$$

These disjoint sets are independent of the decision variables and depend on the data only.

We separate the sum and drop $\max\{\cdot, \cdot\}$ for the index sets $I_{\geq}$ and $I_{<}$. This yields an equivalent formulation because we assumed that for all $i \in I_{\geq}$ it holds $\left(L^\top - Q_y^{-1} B^\top\right)_{:,i} \geq 0$ and together with $x \geq 0$ we conclude that the first argument of the max operator is always nonnegative. Similarly, it holds $\left(L^\top - Q_y^{-1} B^\top\right)_{:,i} \leq 0$ for all $i \in I_{<}$. We write instead of (4) for $\nu = 1, \ldots, N$:

$$\begin{aligned} \min_{x_\nu} \; \hat{\theta}_\nu(x_\nu, x_{-\nu}) = & \frac{1}{2} x_\nu^\top Q_\nu x_\nu + \sum_{i \in I_{\geq}} a_i L_{:,i}^\top x + \sum_{i \in I_{<}} a_i (Q_y^{-1} B^\top)_{:,i} x \\ & + \sum_{i \in I_{\text{ns}}} a_i \max \left\{ (Q_y^{-1} B^\top)_{:,i} x, L_{:,i}^\top x \right\} \\ \text{s.t. } x_\nu \in \mathbb{R}^{n_\nu}_+ & \end{aligned} \tag{5}$$

With this reformulation, we can state a existence and uniqueness result for that non-compact strategy set in case $I_{\mathrm{ns}} = \emptyset$.

Lemma 2.3 (Existence and Uniqueness)**.** *The Nash equilibrium problem* (4) *is reformulated as* (5) *with the index sets* $I_{\geq}, I_{<}, I_{\mathrm{ns}}$. *If* $I_{\mathrm{ns}} = \emptyset$, *there exists a unique Nash equilibrium.*

Proof. We assume $I_{\mathrm{ns}} = \emptyset$ then the objectives are differentiable. Due to [8, Proposition 1.4.2], a strategy $x \in \mathbb{R}^n_+$ is a Nash equilibrium if and only if x solves the variational inequality $\mathrm{VI}(\mathbb{R}^n_+, \hat{\theta}')$ with:

$$\hat{\theta}'(x) = \begin{pmatrix} \nabla_{x_1} \hat{\theta}_1(x_1, x_{-1}) \\ \vdots \\ \nabla_{x_N} \hat{\theta}_N(x_N, x_{-N}) \end{pmatrix}.$$

Since we assume a convex and closed strategy set $\mathbb{R}^n_+$, the variational inequality has a unique solution if $\hat{\theta}'$ is uniformly monotone [8, Theorem 2.3.3]. The remainder of the proof demonstrates this. Let $Q = \mathrm{diag}(Q_1, \ldots, Q_N)$ and $x, \hat{x} \in \mathbb{R}^n_+$, then:

$$\begin{aligned} (x - \hat{x})^\top \left(\hat{\theta}'(x) - \hat{\theta}'(\hat{x})\right) &= (x - \hat{x})^\top \Big[Qx + \sum_{i \in I_{\geq}} a_i L^\top_{:,i} + \sum_{i \in I_{<}} a_i (Q_y^{-1} B^\top)_{:,i} \\ &\quad - Q\hat{x} - \sum_{i \in I_{\geq}} a_i L^\top_{:,i} - \sum_{i \in I_{<}} a_i (Q_y^{-1} B^\top)_{:,i}\Big] \\ &= (x - \hat{x})^\top Q (x - \hat{x}) \geq \mu ||x - \hat{x}||^2 \end{aligned}$$

where μ is the smallest eigenvalue of the symmetric matrix Q which is positive because Q is positive definite. Thus $\hat{\theta}'$ is uniformly monotone. Therefore the variational inequality $\mathrm{VI}(\mathbb{R}^n_+, \hat{\theta}')$ has exactly one solution which is the unique Nash equilibrium of (5). □

With this lemma we illustrate that compactness is not a necessary condition for existence. We conclude this section by showing that the Nash equilibrium for diagonal Q_ν is computed explicitly.

Lemma 2.4. *Assume that the Nash equilibrium problem in* (4) *is reformulated as in* (5) *with the index sets* $I_{\geq}, I_{<}, I_{\mathrm{ns}}$. *Furthermore, let* $I_{\mathrm{ns}} = \emptyset$ *and let* Q_ν *be diagonal. Then the Nash equilibrium is given for* $\nu = 1, \ldots, N$ *by:*

$$x_\nu = -\min \left\{ \sum_{i \in I_{\geq}} a_i Q_\nu^{-1} L^\top_{\nu,i} + \sum_{i \in I_{<}} a_i Q_\nu^{-1} (Q_y^{-1} B^\top)_{\nu,i}, 0 \right\}$$

Proof. We consider the nonnegative orthant as strategy set, therefore Guignard constraint qualification holds in every feasible point and we formulate the KKT system of the leader's problems for

$\nu = 1, \ldots, N$:

$$
\begin{aligned}
& 0 = Q_\nu x_\nu + \sum_{i \in I_\geq} a_i L_{\nu,i}^\top + \sum_{i \in I_<} a_i (Q_y^{-1} B^\top)_{\nu,i} - \lambda_\nu \quad \text{and} \quad 0 \leq \lambda_\nu \perp x_\nu \geq 0 \\
\Leftrightarrow\ & 0 \leq Q_\nu^{-1} \lambda_\nu = x_\nu + \sum_{i \in I_\geq} a_i Q_\nu^{-1} L_{\nu,i}^\top + \sum_{i \in I_<} a_i Q_\nu^{-1} (Q_y^{-1} B^\top)_{\nu,i} \perp x_\nu \geq 0 \\
\Leftrightarrow\ & 0 = x_\nu + \min \left\{ \sum_{i \in I_\geq} a_i Q_\nu^{-1} L_{\nu,i}^\top + \sum_{i \in I_<} a_i Q_\nu^{-1} (Q_y^{-1} B^\top)_{\nu,i}, 0 \right\}
\end{aligned}
$$

The follower's solution is obtained by plugging x in the best response (3). □

3 Existence of Nash Equilibria of Smoothed MLFG

We formulated the MLFG as a convex but nonsmooth Nash game and proved existence of equilibria in case of compact strategy sets. In this section, we relax the nonsmoothness of the follower's best response. For the resulting smooth convex Nash equilibrium problem, we show existence and uniqueness for more general strategy sets.

Similarly to Lemma 2.1, we formulate the follower's KKT conditions:

$$
0 \leq y - Q_y^{-1} B^\top x \perp y - L^\top x \geq 0 \tag{6}
$$

where we replace the complementarity expression by a formulation with a nonlinear complementarity (NCP) function. We consider smooth NCP functions of the following type with smoothing parameter $\varepsilon > 0$:

$$
\phi_\varepsilon(\alpha, \beta) = \alpha + \beta - \tilde{\phi}_\varepsilon(\alpha - \beta). \tag{7}
$$

We remark that the smoothed minimum function is one example of this class, where $\tilde{\phi}_\varepsilon^{\min}(\alpha - \beta) = \sqrt{(\alpha - \beta)^2 + 4\varepsilon^2}$. If we apply (7) on the KKT system (6), we obtain (component wise) for $\varepsilon > 0$:

$$
0 = y - Q_y^{-1} B^\top x + y - L^\top x - \tilde{\phi}_\varepsilon \left(y - Q_y^{-1} B^\top x - y + L^\top x \right) \tag{8}
$$

Therefore, we write the best response function as:

$$
y_\varepsilon(x) = \frac{1}{2} \left[\left(L^\top + Q_y^{-1} B^\top \right) x + \tilde{\phi}_\varepsilon \left(\left(L^\top - Q_y^{-1} B^\top \right) x \right) \right] \tag{9}
$$

The smoothed best response function y_ε in the leader's objectives yields a smooth Nash equilibrium problem with positive smoothing parameter ε for $\nu = 1, \ldots, N$:

$$
\begin{aligned}
\min_{x_\nu \in \mathbb{R}^{n_\nu}} \theta_\nu^\varepsilon(x_\nu, x_{-\nu}) = & \frac{1}{2} x_\nu^\top Q_\nu x_\nu \\
& + \frac{1}{2} \sum_{i=1}^m a_i \left[\left(L^\top + Q_y^{-1} B^\top \right) x + \tilde{\phi}_\varepsilon \left(\left(L^\top - Q_y^{-1} B^\top \right) x \right) \right]_i \\
\text{s.t. } x_\nu \in X_\nu &
\end{aligned}
\tag{NEP(ε)}
$$

For this game, we can state an existence and uniqueness theorem.

Theorem 3.1 (Existence and Uniqueness). *Assume that the Nash equilibrium problem* (NEP(ε)) *has a convex and closed strategy set $X = X_1 \times \cdots \times X_N$, where all X_ν are nonempty, and $\tilde{\phi}_\varepsilon$ is convex. Then the Nash equilibrium problem has a unique equilibrium for every smoothing parameter $\varepsilon > 0$.*

Proof. Due to [8, Proposition 1.4.2] and the convexity assumptions, a strategy $x \in X$ is a Nash equilibrium if and only if x solves the variational inequality VI(X, θ'^{ε}), where:

$$\theta'^{\varepsilon}(x) = \begin{pmatrix} \nabla_{x_1}\theta_1^{\varepsilon}(x_1, x_{-1}) \\ \vdots \\ \nabla_{x_N}\theta_N^{\varepsilon}(x_N, x_{-N}) \end{pmatrix}.$$

Since we assumed a convex and closed strategy set X, the variational inequality has a unique solution if θ'^{ε} is uniformly monotone [8, Theorem 2.3.3]. The remainder of the proof demonstrates this.

We introduce a short hand for the linear term $A = L^\top - Q_y^{-1}B^\top$ and the block diagonal matrix $Q = \operatorname{diag}(Q_1, \ldots, Q_N)$. Let $x, \hat{x} \in X$, then:

$$\begin{aligned}
&(x-\hat{x})^\top \left(\theta'^{\varepsilon}(x) - \theta'^{\varepsilon}(\hat{x})\right) \\
=&(x-\hat{x})^\top \Bigg[Qx + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top a + \frac{1}{2}\sum_{i=1}^m a_i A_{i,:}^\top \tilde{\phi}_\varepsilon'((Ax)_i) \\
&- \left(Q\hat{x} + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top a + \frac{1}{2}\sum_{i=1}^m a_i A_{i,:}^\top \tilde{\phi}_\varepsilon'((A\hat{x})_i)\right)\Bigg] \\
=&(x-\hat{x})^\top \left[Q(x-\hat{x}) + \frac{1}{2}\sum_{i=1}^m a_i A_{i,:}^\top \left(\tilde{\phi}_\varepsilon'((Ax)_i) - \tilde{\phi}_\varepsilon'((A\hat{x})_i)\right)\right]
\end{aligned}$$

We apply the mean value theorem for $\tilde{\phi}_\varepsilon'$, then there exists a $t \in [0,1]$ such that we derive from the last equation with $\xi_i = t(Ax)_i + (1-t)(A\hat{x})_i$:

$$\begin{aligned}
&(x-\hat{x})^\top \left(\theta'^{\varepsilon}(x) - \theta'^{\varepsilon}(\hat{x})\right) \\
=&(x-\hat{x})^\top Q(x-\hat{x}) + \frac{1}{2}\sum_{i=1}^m a_i \underbrace{\tilde{\phi}_\varepsilon''(\xi_i)}_{\geq 0} \underbrace{(x-\hat{x})^\top A_{i,:}^\top A_{i,:}(x-\hat{x})}_{\geq 0} \\
\geq&(x-\hat{x})^\top Q(x-\hat{x}) \geq \mu ||x-\hat{x}||^2
\end{aligned}$$

where μ is the smallest eigenvalue of Q. It is positive because Q is symmetric positive definite. Thus θ'^{ε} is uniformly monotone. Therefore the variational inequality VI($X, \hat{\theta}^{\varepsilon}$) has exactly one solution which is the unique Nash equilibrium of the smoothed game (NEP(ε)). □

An example for a smooth NCP function with convex $\tilde{\phi}_\varepsilon$ is the minimum function $\phi_\varepsilon^{\min}(\alpha, \beta) = \alpha + \beta - \sqrt{(\alpha-\beta)^2 + 4\varepsilon^2}$.

4 Characterization of Nash Equilibria

We recall the assumptions made for the smooth Nash equilibrium problem (NEP(ε)) for positive smoothing parameter ε:

Assumption 4.1. *We assume that the data of the MLFG and its smooth Nash game reformulation satisfy for $\nu = 1, \dots, N$:*

- $Q_\nu \in \mathbb{R}^{n_\nu \times n_\nu}$ *symmetric positive definite*
- $a \in \mathbb{R}^m_+$
- $X_\nu = \{x_\nu \in \mathbb{R}^{n_\nu} | g_\nu(x_\nu) \le 0\} \subseteq \mathbb{R}^{n_\nu}$ *nonempty, convex, and closed*
- $g_\nu : \mathbb{R}^{n_\nu} \to \mathbb{R}^{m_\nu}$ *at least twice differentiable, and convex*
- $Q_y \in \mathbb{R}^{m \times m}$ *positive definite and diagonal*
- $B, L \in \mathbb{R}^{n \times m}$
- *smooth NCP function of the form* $\phi_\varepsilon(\alpha, \beta) = \alpha + \beta - \tilde{\phi}_\varepsilon(\alpha - \beta)$ *where* $\tilde{\phi}_\varepsilon$ *is at least twice differentiable and convex for every* $\varepsilon > 0$

We state the KKT conditions of each optimization problem. For $\nu = 1, \dots N$, the KKT conditions of player ν's optimization problem are:

$$\begin{aligned}
0 =& Q_\nu x_\nu + \frac{1}{2}(L^\top + Q_y^{-1} B^\top)^\top_{\nu,:} a \\
&+ \frac{1}{2} \sum_{i=1}^m a_i (L^\top - Q_y^{-1} B^\top)^\top_{\nu,i} \; \tilde{\phi}_\varepsilon{}' \left([(L^\top - Q_y^{-1} B^\top) x]_i \right) + \nabla_{x_\nu} g_\nu(x_\nu) \lambda_\nu \\
0 =& \min \{\lambda_\nu, -g_\nu(x_\nu)\}
\end{aligned}$$

with Lagrange multiplier $\lambda_\nu \in \mathbb{R}^{m_\nu}_+$ and the Jacobian of the constraints $\nabla_{x_\nu} g_\nu(x_\nu) = (\nabla_{x_\nu} g_{\nu_1}(x_\nu), \dots, \nabla_{x_\nu} g_{\nu_{m_\nu}}(x_\nu))$ $\mathbb{R}^{n_\nu \times m_\nu}$.

Remark 4.2 (KKT point is equilibrium)**.** *We consider the smooth Nash equilibrium problem* (NEP(ε)) *for* $\nu = 1, \dots, N$ *and the Assumptions 4.1. Then the KKT conditions are necessary and sufficient for the global minimizer of each leader problem* (NEP(ε))*, because each leader problem has a strictly convex objective and a convex and closed strategy set. Therefore the joint KKT system is* necessary and sufficient *for the unique Nash equilibrium. The follower's solution can be explicitly computed by the leader's solutions, c.f. (9).*

We summarize the KKT systems of all $\nu = 1, \dots, N$ as a single system using the notation

$Q = \operatorname{diag}(Q_1, \ldots, Q_N)$.

$$0 = Qx + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top a + \frac{1}{2}\sum_{i=1}^{m} a_i (L^\top - Q_y^{-1}B^\top)_{:,i}^\top \; \tilde{\phi}_\varepsilon{}' \left([(L^\top - Q_y^{-1}B^\top)x]_i\right) + \begin{bmatrix} \nabla_{x_1} g_1(x_1)\lambda_1 \\ \vdots \\ \nabla_{x_N} g_N(x_N)\lambda_N \end{bmatrix} \tag{11a}$$

$$0 = \min\left\{ \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix}, \begin{bmatrix} -g_1(x_1) \\ \vdots \\ -g_N(x_N) \end{bmatrix} \right\} \tag{11b}$$

This system characterizes the Nash equilibrium for a fixed relaxation parameter ε. In the following, we analyze the limit $\varepsilon \to 0$.

4.1 Limit System

Theorem 3.1 states an existence and uniqueness result for Nash equilibrium of the relaxed NEP(ε) for every $\varepsilon > 0$. Therefore, we know that there exists a map $\varepsilon \mapsto x^*(\varepsilon)$. If we assume that it holds LICQ in $x_\nu^*(\varepsilon)$ for $\nu = 1, \ldots, N$, the uniqueness of the Lagrange multipliers implies that there is also a map $\varepsilon \mapsto \lambda^*(\varepsilon)$.

In order to emphasize the dependence of (x^*, λ^*) on the parameter ε, we state (11) as:

$$\begin{aligned} 0 =& Qx(\varepsilon) + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top a \\ &+ \frac{1}{2}\sum_{i=1}^{m} a_i (L^\top - Q_y^{-1}B^\top)_{:,i}^\top \; \tilde{\phi}_\varepsilon{}' \left([(L^\top - Q_y^{-1}B^\top)x(\varepsilon)]_i\right) \\ &+ \begin{bmatrix} \nabla_{x_1} g_1(x_1(\varepsilon))\lambda_1(\varepsilon) \\ \vdots \\ \nabla_{x_N} g_N(x_N(\varepsilon))\lambda_N(\varepsilon) \end{bmatrix} \\ 0 =& \min\left\{ \begin{bmatrix} \lambda_1(\varepsilon) \\ \vdots \\ \lambda_N(\varepsilon) \end{bmatrix}, \begin{bmatrix} -g_1(x_1(\varepsilon)) \\ \vdots \\ -g_N(x_N(\varepsilon)) \end{bmatrix} \right\} \end{aligned} \tag{12}$$

We are interested in the solution of the system in the limit $\varepsilon \to 0$. One can show the following Lemma with elementary arguments on convergence of sequences in compact spaces.

Lemma 4.3. *Let $x^*(\varepsilon) \in X \subset \mathbb{R}^n$ and $\lambda^*(\varepsilon) \in \Lambda \subset \mathbb{R}^{\bar{m}}$ for all $\varepsilon > 0$ where $(x^*(\varepsilon), \lambda^*(\varepsilon))$ solves* (12) *for a $\varepsilon > 0$. Assume that X and Λ are nonempty, closed, and bounded.*

Then the sequence of KKT solutions $(x^*(\varepsilon_k), \lambda^*(\varepsilon_k))_{k\in\mathbb{N}}$ for a sequence $\varepsilon_k \to 0$ has convergent subsequence $(x^*(\varepsilon_k), \lambda^*(\varepsilon_k))_{k\in K}$, i.e. $(x^*(\varepsilon_k), \lambda^*(\varepsilon_k)) \to (x^*(0), \lambda^*(0))$ for $k \in K$ and $k \to \infty$.

We note, that compactness of Λ follows from the existence of an accumulation point $x^*(0)$ and from the assumption of LICQ, the proof is analogous to [19, Theorem 5.2].

Next, we analyze the limit system for $(x^*(0), \lambda^*(0))$. The critical part of (12) is $\tilde{\phi}_\varepsilon$ which is nonsmooth for $\varepsilon = 0$. In the remainder of this section, we consider the smoothed minimum function:

$$\tilde{\phi}_\varepsilon(z) = \tilde{\phi}_\varepsilon^{\min}(z) = \sqrt{z^2 + 4\varepsilon^2}.$$

We note that the argument of $\tilde{\phi}_0$ also depends on the smoothing parameter ε.

$$\lim_{\varepsilon \to 0} \tilde{\phi}_\varepsilon{}' \left([(L^\top - Q_y^{-1} B^\top) x(\varepsilon)]_i \right) = \begin{cases} 1, & [(L^\top - Q_y^{-1} B^\top) x(\varepsilon)]_i > 0 \ \forall \varepsilon > 0 \\ -1, & [(L^\top - Q_y^{-1} B^\top) x(\varepsilon)]_i < 0 \ \forall \varepsilon > 0 \end{cases}$$

The derivative of $\tilde{\phi}_\varepsilon$ is not defined if the argument is zero for $\varepsilon = 0$, however the Clarke subdifferential exists:

$$\partial^C \tilde{\phi}_0(z) = \partial^C |z| = \begin{cases} 1, & z > 0 \\ -1, & z < 0 \\ [-1, 1], & z = 0 \end{cases}$$

In the following, we state necessary optimality conditions which are weaker than KKT but require locally Lipschitz continuity only. Subdifferentials are available for convex local Lipschitz functions.

Definition 4.4 (Slater Condition). *A set $X = \{x \in \mathbb{R}^n | g_i(x) \leq 0, g_i$ convex $, i = 1, \ldots, m\}$ fulfills the Slater condition if and only if there exists a $\hat{x} \in X$ such that $g_i(\hat{x}) < 0$ for $i = 1, \ldots, m$. ($\hat{x}$ is called a strictly feasible point.)*

Theorem 4.5 (Fritz-John conditions of Clarke for nonsmooth objectives). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz-continuous and $g : \mathbb{R}^n \to \mathbb{R}^m$ differentiable. Let $\bar{x} \in \mathbb{R}^n$ be a local minimizer of f on $X = \{x \in \mathbb{R}^n | g(x) \leq 0\}$.*

Then there exists multipliers $\bar{\alpha} \geq 0$ and $\bar{\lambda} \in \mathbb{R}^m_+$ such that $(\bar{\alpha}, \bar{\lambda}) \neq 0$, $\lambda^\top g(x^*) = 0$, and:

$$0 \in \bar{\alpha} \partial^C f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(\bar{x}).$$

In case Slater condition holds, it is guaranteed that $\bar{\alpha} \neq 0$.

Proof. In [7, Theorem 6.1.1] a more general setting is proved. If the Slater condition holds, $\bar{\alpha} \neq 0$ [7, Proposition 6.3.1]. □

If Slater condition holds, we can assume w.l.o.g. that $\bar{\alpha} = 1$. With Theorem 4.5, we state necessary optimality conditions of the nonsmooth Nash equilibrium problem formulation (NEP). Let $\bar{x}$ be a vector of local minimizer $\bar{x}_\nu$ of each leader problem and Slater condition holds for every strategy set X_ν, then there exists for $\nu = 1, \dots, N$ a multiplier $\lambda_\nu \in \mathbb{R}_+^{m_\nu}$ such that:

$$
\begin{aligned}
0 \in\ & Q_\nu \bar{x}_\nu + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)_{\nu,:}^\top a \\
& + \frac{1}{2}\sum_{i=1}^m a_i (L^\top - Q_y^{-1}B^\top)_{\nu,i}^\top\, \partial^C \tilde{\phi}_0 \left([(L^\top - Q_y^{-1}B^\top)\bar{x}]_i \right) + \nabla_{x_\nu} g_\nu(\bar{x}_\nu)\lambda_\nu
\end{aligned} \tag{13a}
$$

$$
0 = \min\{\lambda_\nu, -g_\nu(\bar{x}_\nu)\} \tag{13b}
$$

Due to convexity of the objective, it is regular and its smooth parts are differentiated in the classical sense and only the nonsmooth part is differentiated in the Clarke sense.

In case $[(L^\top - Q_y^{-1}B^\top)x^*(0)]_i = 0$ does not appear for any i, the necessary optimality conditions of Fritz-John type in (13) coincide with the classical KKT conditions which necessary and sufficient in our setting, thus $\bar{x} = x^*(0)$. The more interesting case is if there is at least one i with $[(L^\top - Q_y^{-1}B^\top)x^*(0)]_i = 0$. Here, we define the limit $\Phi = \lim_{\varepsilon\to 0} \tilde{\phi}_\varepsilon{}' \left([(L^\top - Q_y^{-1}B^\top)x(\varepsilon)]_i\right)$. It is clear that $\Phi \in [-1, 1]$ because $\left|\tilde{\phi}_\varepsilon{}'(z)\right| = \left|\frac{z}{\sqrt{z^2+4\varepsilon^2}}\right| \leq \left|\frac{z}{\sqrt{z}}\right| = 1$ for any $z \neq 0$ and $\tilde{\phi}_\varepsilon{}'(0) = 0$. Therefore, $\Theta \in \partial^C \tilde{\phi}_0 \left([(L^\top - Q_y^{-1}B^\top)x(0)]_i\right)$ and we conclude that $x^*(0)$ satisfies the necessary optimality conditions in (13).

Instead of looking at optimality conditions for the nonsmooth formulation, we are considering stationarity concepts for mathematical programs with complementarity constraints (MPCC) next. For that, we adapt [18, Theorem 2]:

Definition 4.6 (S-Stationarity)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $G_1, G_2 : \mathbb{R}^n \to \mathbb{R}^l$ be smooth functions. Then we call:*

$$
\begin{aligned}
\min_z\ & f(z) \\
\text{s.t. } & g(z) \leq 0 \\
& 0 = \min\{G_1(z), G_2(z)\}
\end{aligned} \tag{MPCC}
$$

a mathematical program with complementarity constraints. Furthermore, we call $\bar{z}$ a strongly (s-)stationary point of (MPCC) *if there exist multipliers $(\lambda, \Gamma_1, \Gamma_2) \in \mathbb{R}^{m+l+l}$ with:*

$$
\begin{aligned}
0 = \nabla_z f(\bar{z}) + \sum_{i=1}^m \lambda_i \nabla_z g_i(\bar{z}) - \sum_{i=1}^l \Gamma_{1,i}\nabla_z G_{1,i}(\bar{z}) - \sum_{i=1}^l \Gamma_{2,i}\nabla_z G_{2,i}(\bar{z})& \\
g(\bar{z}) \leq 0& \\
\lambda \geq 0& \\
g_i(\bar{z})\lambda_i = 0,& \quad i = 1, \dots, m \\
G_{1,i}(\bar{z})\Gamma_{1,i} = 0,& \quad i = 1, \dots, l \\
G_{2,i}(\bar{z})\Gamma_{2,i} = 0,& \quad i = 1, \dots, l \\
\Gamma_{1,i}, \Gamma_{2,i} \geq 0,& \quad i : G_{1,i}(\bar{z}) = G_{2,i}(\bar{z}) = 0
\end{aligned} \tag{14}
$$

Each leader problem can be formulated as a MPCC, which form together the generalized Nash equilibrium problem (GNEP) formulation of the MLFG. In the following theorem we verify that $x_\nu^*(0)$ is a s-stationary point for every leader $\nu = 1, \dots, N$.

Theorem 4.7. *Under the given assumptions the Nash equilibrium $x_\nu^*(0)$ is a strongly stationary point for each leader $\nu = 1, \dots, N$.*

Proof. We begin with the MPCC formulation of each leader problem for $\nu = 1, \dots, N$, we have:

$$\begin{aligned}\min_{x_\nu, y}\ & \theta_\nu(x_\nu, x_{-\nu}) = \frac{1}{2} x_\nu^\top Q_\nu x_\nu + a^\top y \\ \text{s.t. } & g_\nu(x_\nu) \leq 0 \\ & 0 = \min\{G_1(x_\nu, x_{-\nu}, y), G_2(x_\nu, x_{-\nu}, y)\}\end{aligned}$$

with $G_1(x_\nu, x_{-\nu}, y) = y - (Q_y^{-1} B^\top)^\top x$ and $G_2(x_\nu, x_{-\nu}, y) = y - L^\top x$.

Before verifying the statement, we apply the s-stationarity conditions of (14):

$$0 = \begin{pmatrix} Q_\nu x_\nu \\ a \end{pmatrix} + \sum_{i=1}^{m} \lambda_i \begin{pmatrix} \nabla_{x_\nu} g_{\nu,i}(x_\nu) \\ 0 \end{pmatrix} - \sum_{i=1}^{l} \Gamma_{1,i} \begin{pmatrix} -(Q_y^{-1} B^\top)_{\nu,i} \\ e_i \end{pmatrix} - \sum_{i=1}^{l} \Gamma_{2,i} \begin{pmatrix} -L_{\nu,i} \\ e_i \end{pmatrix} \tag{15a}$$

$$g_\nu(x_\nu) \leq 0 \tag{15b}$$

$$\lambda_\nu \geq 0 \tag{15c}$$

$$g_{\nu,i}(x_\nu)\lambda_{\nu,i} = 0, \quad i = 1, \dots, m_\nu \tag{15d}$$

$$(y - (Q_y^{-1} B^\top)^\top x)_i \Gamma_{1,i} = 0, \quad i = 1, \dots, m \tag{15e}$$

$$(y - L^\top x)_i \Gamma_{2,i} = 0, \quad i = 1, \dots, m \tag{15f}$$

$$\Gamma_{1,i}, \Gamma_{2,i} \geq 0, \quad i : (y - (Q_y^{-1} B^\top)^\top x)_i = (y - L^\top x)_i = 0 \tag{15g}$$

We derive the conditions of s-stationarity (15) from (13) which can be equivalently expressed that there exists $c \in [0,1]^m$ such that:

$$0 = Q_\nu x_\nu + \sum_{i=1}^{m} a_i \left(c_i (Q_y^{-1} B^\top)_{\nu,i} + (1 - c_i) L_{\nu,i} \right) + \nabla_{x_\nu} g_\nu(\bar{x}_\nu)\lambda_\nu \tag{16a}$$

$$0 = \min\{\lambda_\nu, -g_\nu(\bar{x}_\nu)\} \tag{16b}$$

Where we remark, that c_i is unique only if $(L^\top x)_i \neq ((Q_y^{-1} B^\top)x)_i$ because the NCP function approximates the absolute value such that $\left|[(L^\top - Q_y^{-1} B^\top)x]_i\right| = (L^\top - Q_y^{-1} B^\top)_{\nu,i}^\top \, \partial^C \tilde{\phi}_0 \left([(L^\top - Q_y^{-1} B^\top)x]_i\right)$.

We choose $\Gamma_{1,i} = a_i c_i$ and $\Gamma_{2,i} = a_i(1 - c_i)$, with that choice $\Gamma_{1,i}, \Gamma_{2,i} \geq 0$ for all i because $a_i > 0$, therefore (15g) holds. The equation (16a) yields the upper part of (15a), the lower part is obtained with the choice of the multipliers, since $a_i - \Gamma_{1,i} - \Gamma_{2,i} = a_i - a_i c_i - a_i(1 - c_i) = 0$ for all i. Furthermore, second expression (16b) is equivalent to (15b-15d).

It remains to demonstrate (15e-15f):

(i) Assume $G_{1,i}(x_\nu, x_{-\nu}) \geq 0$, then $c_i = 0$ because of the Clarke derivative in (13a) and this yields $\Gamma_{1,i} = 0$. Therefore $\Gamma_{2,i} = a_i > 0$ and $G_{2,i}(x_\nu, x_{-\nu}) = 0$, which is consistent to the feasibility.

(ii) Assume $G_{1,i}(x_\nu, x_{-\nu}) = G_{2,i}(x_\nu, x_{-\nu}) = 0$, then $\Gamma_{1,i}, \Gamma_{2,i}$ are arbitrary.

(iii) Assume $G_{1,i}(x_\nu, x_{-\nu}) = 0$ and $G_{2,i}(x_\nu, x_{-\nu}) \geq 0$, then $c_i = 1$ similar to (i) and this yields $\Gamma_{2,i} = 0$. Therefore $\Gamma_{1,i} = a_i > 0$, which is consistent to the feasibility.

Since this holds for all $i = 1, \dots, m$, (i)-(iii) yield (15e-15f) and the proof is complete. □

5 Numerical Algorithms

In the previous section, we reformulated the MLFG in (1,2) as smooth Nash game (NEP(ε)) with a smoothing parameter $\varepsilon > 0$. We apply a gradient type method and recall corresponding convergence theory. As alternative we propose a Newton like method.

5.1 The Method

For a fixed smoothing parameter ε, the KKT system in (11) characterizes the unique Nash equilibrium of (NEP(ε)), therefore, we aim to find a primal dual pair (x, λ) which satisfies the KKT conditions. Let $F_1^\varepsilon(x, \lambda) = F_1^\varepsilon(z) = 0$ be a short hand of (11a) and $F_2^\varepsilon(x, \lambda) = F_2^\varepsilon(z) = 0$ of (11b), respectively. We abbreviate the concatenation with $F^\varepsilon(z)$.

With this notation, the KKT system can be equivalently expressed as the minimization of the auxiliary function $\Psi_\varepsilon : \mathbb{R}^{n+\bar{m}} \to \mathbb{R}_+$ where $\bar{m} = m_1 + \cdots + m_N$ and:

$$\Psi_\varepsilon(z) = \frac{1}{2}||F^\varepsilon(z)||_2^2 = \frac{1}{2}\left(||F_1^\varepsilon(z)||_2^2 + ||F_2^\varepsilon(z)||_2^2\right)$$

The global minimum is obtained for an z^* satisfying $\Psi_\varepsilon(z^*) = 0$. For convergence theory, the Lipschitz property of Ψ_ε is crucial, therefore we prove it in the following lemma:

Lemma 5.1. *Ψ_ε is locally Lipschitz and directionally differentiable.*

Proof. We verify the properties for each part of the sum separately.
(i) $\frac{1}{2}||F_1(z)||_2^2 \in C^1$, as a composition of C^1 functions because ϕ_ε is assumed to be twice differentiable. Therefore this part is locally Lipschitz and directionally differentiable.
(ii) $\frac{1}{2}||F_2(z)||_2^2 = \frac{1}{2}\sum_{i=1}^{m} \min^2\{\lambda_i, -g_i(x)\} = \frac{1}{8}\sum_{i=1}^{m} \left(\lambda_i - g_i(x) - |\lambda_i + g_i(x)|\right)^2$ is locally Lipschitz as a composition of locally Lipschitz functions. It is also directionally differentiable as it is also a composition of directionally differentiable functions. □

We are interested in the solution of the system for ε close to zero. However, the problem characteristics are poor for very small ε and we expect bad numerical performance with arbitrary initial values. Therefore, we propose to solve a sequence of minimization problems:

$$\min \ \Psi_\varepsilon(z) \quad \text{s.t. } z \in \mathbb{R}^{n+\bar{m}}$$

with decreasing sequence of $(\varepsilon_i)_{i\in\mathbb{N}}$. This approach returns a sequence of KKT points $(z^*(\varepsilon_i))_{i\in\mathbb{N}} = (x^*(\varepsilon_i), \lambda^*(\varepsilon_i))_{i\in\mathbb{N}}$ whose primal part $(x^*(\varepsilon_i))_{i\in\mathbb{N}}$ is the Nash equilibrium of NEP(ε_i). We use the solution $z^*(\varepsilon_i)$ as initial value for the subsequent solving for ε_{i+1}.

To further increase the quality of the initial values, we propose an update for the primal variables x based on formal Taylor expansion of the map $\varepsilon \mapsto x^*(\varepsilon)$. We compute the derivative of the objectives of the Nash game with respect to ε. For $\nu = 1, \ldots, N$:

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\nabla_{x_\nu}\theta_\nu^\varepsilon(x_\nu(\varepsilon), x_{-\nu}(\varepsilon))\right) = 0$$

which leads to the following system:

$$E\frac{\partial x}{\partial \varepsilon}(\varepsilon) = h$$

Here, we denote $\tilde{\phi}_\varepsilon(t) = \Phi(t,\varepsilon)$ to emphasize the explicit dependence on ε, then the linear system has the following coefficient matrix:

$$E = Q + \frac{1}{2}\sum_{i=1}^{m} a_i (L^\top - Q_y^{-1}B^\top)_{:,i}^\top (L^\top - Q_y^{-1}B^\top)_{:,i} \frac{\partial^2 \Phi}{\partial t^2}((L^\top - Q_y^{-1}B^\top)_{:,i}^\top x, \varepsilon)$$

and the right-hand-side:

$$h = \frac{1}{2}\sum_{i=1}^{m} a_i (L^\top - Q_y^{-1}B^\top)_{:,i}^\top \frac{\partial \Phi}{\partial \varepsilon}((L^\top - Q_y^{-1}B^\top)_{:,i}^\top x, \varepsilon)$$

We remark that E is nonsingular since it is the second derivative of the strictly convex objectives. We summarize the general approach in the following algorithm:

Algorithm 1

1: **Initialize** Choose $z^0(\varepsilon_0) = (x^0(\varepsilon_0), \lambda^0(\varepsilon_0)) \in \mathbb{R}^{n+\bar{m}}$, $tol > 0$, $\varepsilon_0 \in (1,2)$, $\gamma \in (0,1)$

2: **for** $i = 0, 1, \dots$ **do**

3: Compute Nash equilibrium of (NEP(ε_i)) with initial guess $z^0(\varepsilon_i) = (x^0(\varepsilon_i), \lambda^0(\varepsilon_i))$ by

$$z^*(\varepsilon_i) = (x^*(\varepsilon_i), \lambda^*(\varepsilon_i)) = \arg\min_{z \in \mathbb{R}^{n+\bar{m}}} \Psi_{\varepsilon_i}(z)$$

4: Decrease $\varepsilon_{i+1} = \gamma\varepsilon_i$

5: Compute Taylor update

$$d^i = \frac{\partial x}{\partial \varepsilon}(\varepsilon_{i+1})$$

6: Update initial guess $x^0(\varepsilon_{i+1}) = x^*(\varepsilon_i) - (\varepsilon_i - \varepsilon_{i+1})d^i$ and $\lambda^0(\varepsilon_{i+1}) = \lambda^*(\varepsilon_i)$

7: **end for**

In Step 5 of the algorithm, we use a forward evaluation of $\frac{\partial x}{\partial \varepsilon}$ but also $\frac{\partial x}{\partial \varepsilon}(\varepsilon_i)$ is a valid choice. In the remainder of this section, we propose two algorithms for computation of the Nash equilibria in Step 3.

5.2 Subgradient Method

To generate the sequence of Nash equilibria, we propose a method which is based on subgradient decent. We apply the method of [5] for a fixed relaxation parameter $\varepsilon > 0$. Stationary points of Ψ_ε are computed as the limit of a sequence of h-δ-stationary points. Bagirov et. al. [5] showed that the limit is a Clarke stationary point. With this method we obtain the unique Nash Equilibrium of the smoothed game.

Before stating the algorithm and the inherent convergence results, we introduce some terms:

Definition 5.2 (h-δ stationary point)**.** *Let $W_h(x)$ denote the closed convex hull of all possible quasisecants of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ at the point $x \in \mathbb{R}^n$ with length $h > 0$:*

$$W_h(x) = \overline{\text{conv}}\,\{w \in \mathbb{R}^n : \exists d \in \mathbb{R}^n \text{ with } ||d|| = 1 : w = v(x, d, h)\}$$

Then a point x is called a h-δ stationary point of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ if and only if:

$$\min\{||v|| : v \in W_h(x)\} < \delta.$$

Lemma 5.3. *(1) If $\max\{||v|| : v \in W_h(z)\} < \infty$ for all iterates $z^k \in \mathbb{R}^{n+\bar{m}}$, the loop in Lines 7-15 terminates after finitely many iterations with a decent direction. (2) The loop in Lines 4-20 terminates after finitely many iterations with a h-δ-stationary point.*

Proof. (1) Since Ψ_ε is locally Lipschitz with Lemma 5.1, [5, Proposition 4.1] is applicable.
(2) The function Ψ_ε is bounded from below as it takes nonnegative values only, therefore [5, Proposition 5.1] is applicable. □

Theorem 5.4. *Assume $\mathcal{L}(z^0) = \left\{z \in \mathbb{R}^{n+\bar{m}} : \Psi_\varepsilon(z) \leq \Psi_\varepsilon(z^0)\right\}$ is bounded and Assumption A.2 is fulfilled. Then there exists at least one accumulation point of the sequence $(z^k)_{k\in\mathbb{N}}$ generated by Alg. 2 and any accumulation point is a stationary point of Ψ_ε.*

Proof. Due to Lemma 5.1, Ψ_ε is locally Lipschitz and therefore [5, Proposition 5.2] is applicable. The boundedness of $\mathcal{L}(z^0)$ implies that it exists at least one accumulation point. □

Bagirov et. al. [5] state that subgradients are in particular quasisecants and therefore we limit ourselves to the usage of subgradients as decent directions and to $h = 0$ in the implementations. The algorithm is stated as Algorithm 2 below.

Algorithm 2 Subgradient Method

Initialize $h_0 > 0$, $\delta_0 > 0$, $\gamma \in (0,1)$, $z^0 \in \mathbb{R}^{n+\bar{m}}$, $d_0 \in \mathbb{R}^{n+\bar{m}}$ with $||d_0|| = 1$, $0 < c_2 \leq c_1 \leq 1$, $\varepsilon > 0$
for $k = 0, \ldots$ **do**
  $\bar{z}_1 = z^k$
  for $j = 1, \ldots$ **do** ▷ compute h-δ-stationary point
    Compute quasisecant $v_0 = v(\bar{z}_j, d_0, h)$
    $\tilde{v}_0 = v_0$
    for $i = 0, 1, \ldots$ **do** ▷ find decent direction
      $c_i = \arg\min\{||cv_i + (1-c)\tilde{v}_i||_2^2 \; |c \in (0,1)\}$
      $\bar{v}_i = c_i v_i + (1-c_i)\tilde{v}_i$
      if $||\bar{v}_i|| \leq \delta_k$ **then** return $v^j = \bar{v}_i$
      $d_i = -\frac{\bar{v}_i}{||\bar{v}_i||}$
      if $\Psi_\varepsilon(\bar{z}_j + hd_i) - \Psi_\varepsilon(\bar{z}_j) \leq -c_1 h||\bar{v}_i||$ **then** return $v^j = \bar{v}_i$
      Compute quasisecant $v_{i+1} = v(x, d_i, h)$
      $\tilde{v}_{i+1} = \bar{v}_i$
    end for
    if $||v^j|| \leq \delta_k$ **then** Stop
    $d^j = -\frac{v^j}{||v^j||}$
    Compute step length such that

$$\sigma_j = \arg\max\{\Psi_\varepsilon(\bar{z}_j + \sigma d^j) - \Psi_\varepsilon(\bar{z}_j) \leq -c_2\sigma||v^j|| \; |\sigma > 0\}$$

    Update $\bar{z}_{j+1} = \bar{z}_j + \sigma_j d^j$
  end for
  $z^{k+1} = \bar{z}_j$
  $h_{k+1} = \gamma h_k$
  $\delta_{k+1} = \gamma \delta_k$
end for

5.3 Nonsmooth Newton

Next, we present an improved method. The joint KKT system (11) leads to the problem to find the unique $z^*(\varepsilon) = (x^*(\varepsilon), \lambda^*(\varepsilon))$ that satisfy:

$$F^\varepsilon(z) = 0$$

This is a nonlinear and nonsmooth system of equations which depend on the parameter $\varepsilon > 0$. The generalized Newton method can be written as the solving of a sequence of the following linear systems:

$$H\left(z^{k+1} - z^k\right) = -F^\varepsilon(z^k)$$

for an element $H \in \partial F^{\varepsilon}(z^k)$ of the Clarke subdifferential of F^{ε}. The explicit structure of a generalized Jacobian H can be found in A.2.

Since H is not necessarily regular, we verify this property in Step 5 of Algorithm 3 and use a first order decent direction if necessary. This subgradient decent also serves as globalization strategy.

Algorithm 3 Nonsmooth Newton

Initialize Choose $z^0 = (x^0, \lambda^0) \in \mathbb{R}^{n+\bar{m}}$, $\beta \in (0,1)$, $\sigma \in (0, 0.5)$, $tol > 0$, $\varepsilon > 0$

for $k = 0, \dots$ **do**

  if $\Psi_\varepsilon \leq tol$ **then** Stop

  Let $H \in \partial F^{\varepsilon}(z^k)$

  if H singular **then** do steepest decent of Ψ_ε, thus

$$s^k \in -\partial \Psi_\varepsilon(z^k)$$

  with the step length $t_k = \max\{\beta^l | l = 0, 1, \dots\}$ which fulfills the Armijo condition

$$\Psi_\varepsilon(z^k + t^k s^k) \leq \Psi_\varepsilon(z^k) + t^k \sigma {s^k}^\top s^k$$

  else let $t_k = 1$ and compute Newton step by solving

$$H s^k = -F^{\varepsilon}(z^k)$$

  Update $z^{k+1} = z^k + t_k s^k$

end for

6 Numerical Results

In the previous sections, we proposed an algorithm with gradient updates of the primal variables. This included the computation of Nash equilibria for a sequence of smoothing parameter $(\varepsilon_i)_{i \in \mathbb{N}}$. For this computation, we introduced a subgradient and a Newton method. The presented numerical results are obtained for the data sets in A.3 which are adapted from [12]. All plots are generated for the Data Set 1, however experiments with Data Set 2 produced similar graphics.

The naive approach of computing a sequence of Nash equilibria is to use the Nash equilibrium of a larger smoothing parameter as initial for the subsequent computation with the smaller smoothing parameter. The main purpose of the outer Taylor expansion based update in Algorithm 1 (Step 5 and 6) is to improve the quality of the initials in order to reduce the computational effort in Step 3.

In the upper left part of Figure 6, we observe the quadratic decent of the error for decreasing smoothing parameter. In the upper right part, the Taylor update is exemplary illustrated for one component of the leader variables. The blue dots indicate each the Nash equilibrium of a (NEP(ε_i)), $x^*(\varepsilon_i)$. A black line represents the Taylor update and the lower end of a black line indicates the updated initial values $x^0(\varepsilon_{i+1})$ for the subsequent Nash equilibrium computation.

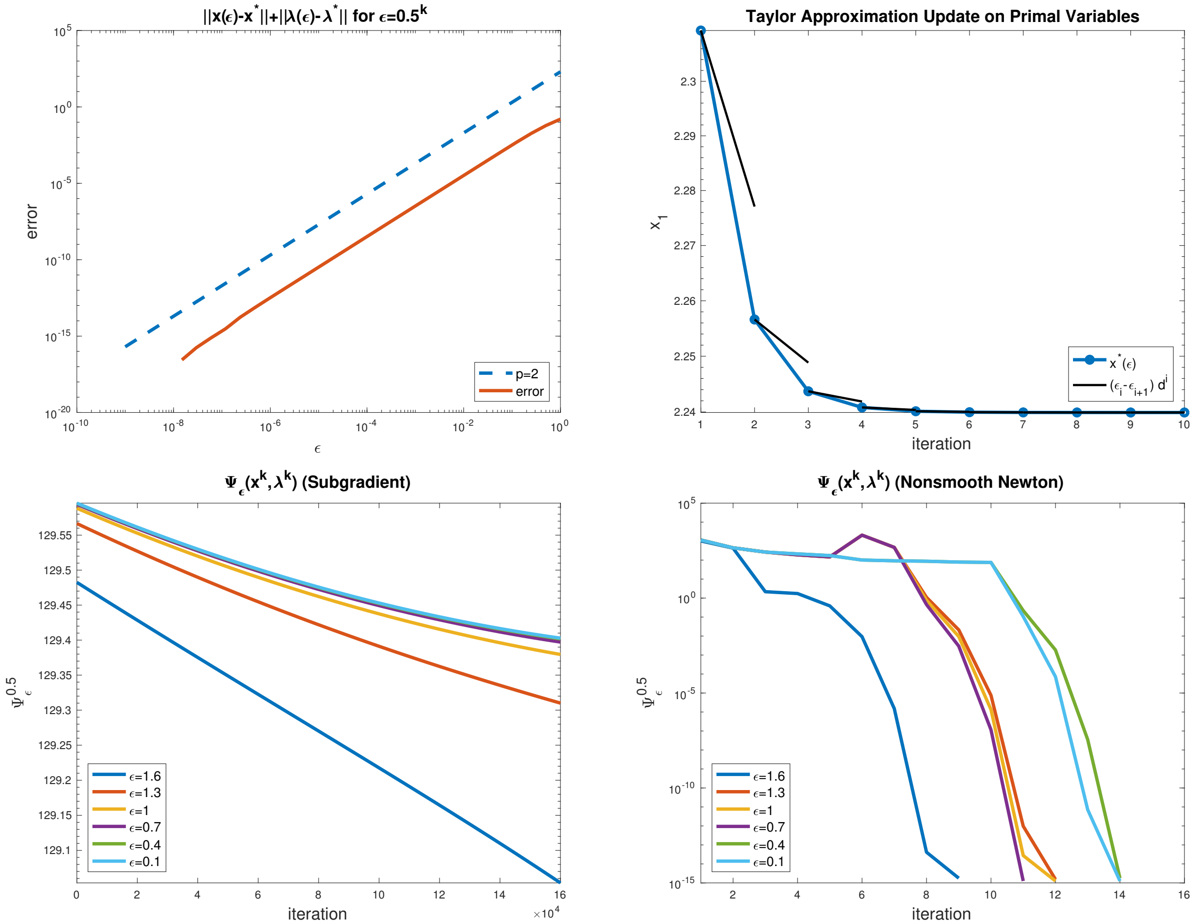

Figure 1: Upper left: Quadratic convergence to the limit Nash equilibrium $x^*(0)$, right: Taylor expansion based update on primal variables, exemplary for first leader variable; Lower: comparison of Subgradient and Nonsmooth Newton method for varying smoothing parameter.

The lower part of Figure 6 is dedicated to illustrate the importance of large smoothing parameter for the first computations of Nash equilibria. Since the problem gets closer to it original nonsmooth formulation as ε decreases, the problem is also more challenging to solve for both Subgradient and Nonsmooth Newton method. We observe this expected behavior, in particular if we compare the number of iterations for $\varepsilon = 1.6$ and $\varepsilon = 0.1$.

As already seen in Figure 6, the subgradient based method suffers from characteristically slow convergence for our instances. In Figure 6 on the left, all iterations for a sequence of decreasing smoothing parameter are shown. On the right, we observe the performance for a single fixed ε but using multiple random initial values.

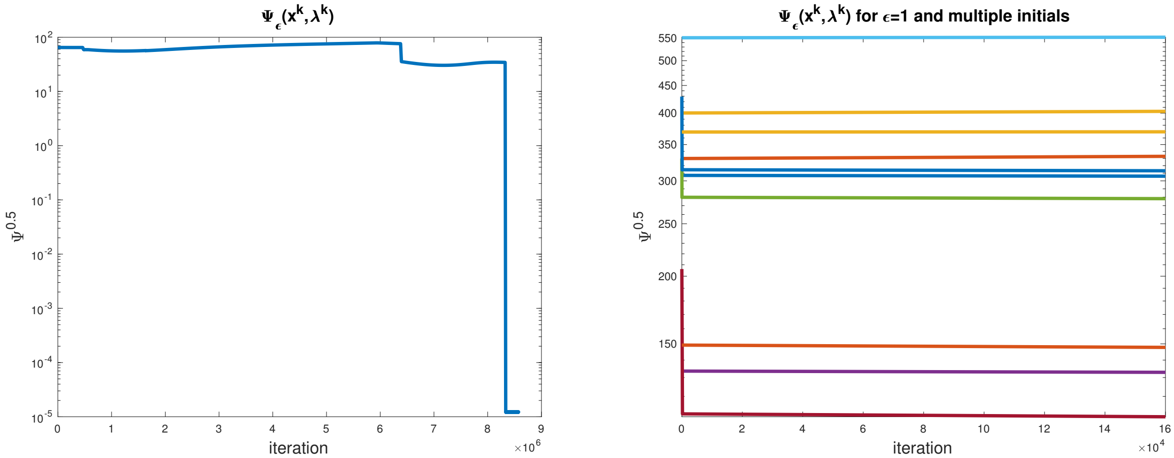

Figure 2: Subgradient Method; left all iterations for decreasing sequence of smoothing parameters, right for one smoothing parameter and multiple initials.

Similarly to Figure 6, left in Figure 6, all iterations for a sequence of decreasing smoothing parameter are shown for the Nonsmooth Newton. The alternating behavior is due to the decreasing parameter changing the minimization problem. The right part of Figure 6 illustrates the decrease in Ψ_ε for different random initial values but fixed smoothing parameter.

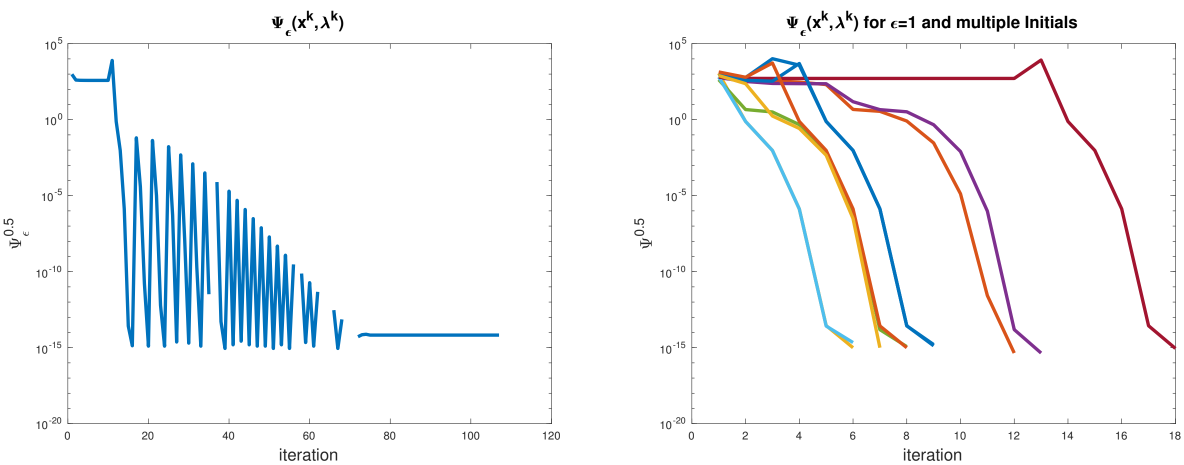

Figure 3: Nonsmooth Newton Method; left all iterations for decreasing sequence of smoothing parameters, right for one smoothing parameter and multiple initials.

7 Conclusion

We presented a quadratic MLFG and explicitly computed best response of the follower. With the best response we derived a Nash game formulation where existence theory is available. Furthermore, we smoothed the best response function and formulate the MLFG as smooth Nash game and proved existence and uniqueness of the Nash equilibrium for all smoothing parameters. We followed an all KKT approach to characterize the corresponding Nash equilibrium. For decreasing positive smoothing parameter, we showed that the limit of Nash equilibria satisfies conditions of s-stationarity. Numerically, we computed Nash equilibria with a globalized nonsmooth Newton and compare with a standard methods based on subgradients. For efficient computation, we updated the primal variables by a Taylor approximation before a subsequent computation of the Nash equilibrium for a smaller smoothing parameter.

A Appendix

A.1 Subgradient Method

In the following, we state the definition of quasisecants and a theorem which relates quasisecants and subgradients. Furthermore, we state a assumption which is needed in the convergence theorem of the subgradient method. All is adapted from [5] and can be found there in an extended form.

Definition A.1 (Quasisecant). *A vector $v = v(x, d, h) \in \mathbb{R}^n$ is called a quasisecant of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$ at the point $x \in \mathbb{R}^n$ in direction $d \in \mathbb{R}^n$ with $||d|| = 1$ with the length $h > 0$ if and only if:*

$$f(x + hd) - f(x) \leq h\langle v, d\rangle$$

and:

$$v \in \partial_{d,h} f(x) + B_{O(h)}$$

where $\partial_{d,h} f = \cup_{t\in[0,h]} \partial f(x + td)$ denotes the union of all Clarke subdifferentials over the set $\mathrm{conv}(x, x + hd)$.

For the convergence proof it is necessary to study the relation of $W_h(x)$ and the subdifferential $\partial f(x)$ and therefore the following assumption is crucial.

Assumption A.2. *At any given point $x \in \mathbb{R}^n$ there exists $\delta = \delta(x) > 0$ such that $O(y, h) \downarrow 0$ uniformly as $h \downarrow 0$ for all $y \in B_\delta(x)$ that is for any $\eta > 0$ there exists $h(\eta) > 0$ such that $O(y, h) < \eta$ for all $h \in (0, h(\eta))$ and $y \in B_\delta(x)$.*

In particular, this assumptions guarantees a certain relation between quasisecants and subgradients.

Theorem A.3. *Assume that a function f satisfies Assumption A.2. Then at a given point $x \in \mathbb{R}^n$ for any $\eta > 0$ there exists $\delta = \delta(\eta)$ and $h(\eta) > 0$ such that:*

$$W_h(y) \subset \partial f + B_\eta$$

for all $h \in (0, h(\eta))$ and $y \in B_\delta(x)$. Furthermore, it holds for locally Lipschitz function that the limit $h \to 0$ of the $W_h(x)$ lies in the subdifferential, i.e.:

$$W_0(x) \subset \partial f(x)$$

A.2 Nonsmooth Newton

We propose a nonsmooth Newton method to compute Nash equilibria. In order to keep the readability of the paper, we specify the structure of the generalized Jacobian here.

We look at the elements of ∂F_ε as a block matrix:

$$H = \begin{bmatrix} \boxed{A} & \boxed{B} \\ \boxed{C} & \boxed{D} \end{bmatrix}$$

Where $\boxed{A} \in \mathbb{R}^{n \times n}$:

$$\boxed{A} = Q + \begin{bmatrix} \nabla_{x_{(1)}}(\nabla_{x_1} g_1(x_1)\lambda_1) & & \\ & \ddots & \\ & & \nabla_{x_N}(\nabla_{x_N} g_N(x_N)\lambda_N) \end{bmatrix}$$
$$+ \frac{1}{2}\sum_{i=1}^{m} a_i (L^\top - Q_y^{-1}B^\top)^\top_{:,i}(L^\top - Q_y^{-1}B^\top)_{:,i}$$
$$\left(\frac{1}{\sqrt{[(L^\top - Q_y^{-1}B^\top)x]_i^2 + 4\varepsilon^2}} - \frac{[(L^\top - Q_y^{-1}B^\top)x]_i^2}{\sqrt{[(L^\top - Q_y^{-1}B^\top)x]_i^2 + 4\varepsilon^2}^3} \right)$$

and $\boxed{B} \in \mathbb{R}^{n \times \bar{m}}$:

$$\boxed{B} = \begin{bmatrix} \nabla_{x_1} g_1(x_1) & & \\ & \ddots & \\ & & \nabla_{x_N} g_N(x_N) \end{bmatrix}$$

and the block diagonal $\boxed{C} \in \mathbb{R}^{\bar{m} \times n}$:

$$\boxed{C} = \begin{bmatrix} \boxed{C_1} & & \\ & \ddots & \\ & & \boxed{C_N} \end{bmatrix}$$

with the blocks $\boxed{C_\nu} \in \mathbb{R}^{m_\nu \times n_\nu}$ and the entries:

$$\left(\boxed{C_\nu}\right)_{i,j} = \partial^C_{x_\nu^j} \min\left\{\lambda_\nu^i, -g_\nu^i(x_\nu)\right\} = \begin{cases} 0, & \lambda_\nu^i < -g_\nu^i(x_\nu) \\ -\frac{\partial}{\partial x_\nu^j} g_\nu^i(x_\nu), & \lambda_\nu^i > -g_\nu^i(x_\nu) \\ \left[0, -\frac{\partial}{\partial x_\nu^j} g_\nu^i(x_\nu)\right], & \lambda_\nu^i = -g_\nu^i(x_\nu) \end{cases}$$

and the diagonal matrix:

$$\boxed{D} = \begin{bmatrix} \boxed{D_1} & & \\ & \ddots & \\ & & \boxed{D_N} \end{bmatrix} \in \mathbb{R}^{\bar{m} \times \bar{m}}$$

with its blocks $\boxed{D_\nu} \in \mathbb{R}^{m_\nu \times m_\nu}$ with the entries:

$$\left(\boxed{D_\nu}\right)_{i} = \partial^C_{\lambda_\nu^i} \min\left\{\lambda_\nu^i, -g_\nu^i(x_\nu)\right\} = \begin{cases} 1, & \lambda_\nu^i < -g_\nu^i(x_\nu) \\ 0, & \lambda_\nu^i > -g_\nu^i(x_\nu) \\ [1, 0], & \lambda_\nu^i = -g_\nu^i(x_\nu) \end{cases}$$

A.3 The Data

In the following, we specify the data used for the experiments presented in Section 5. We adapted the data used in [12].

A.3.1 Data Set 1

We consider $N = 2$ leader with each $n_1 = n_2 = 2$ variables. The objectives of the leader are given by

$$Q_1 = \begin{bmatrix} 1.7 & 1.6 \\ 1.6 & 2.8 \end{bmatrix} \quad Q_2 = \begin{bmatrix} 2.7 & 1.3 \\ 1.3 & 3.6 \end{bmatrix} \quad a = \begin{bmatrix} 1.4 \\ 2.6 \\ 2.1 \end{bmatrix}$$

Each leader has $m_1 = m_2 = 3$ linear constraints $g_\nu = A_\nu^T x_\nu + b_\nu \leq 0$ with

$$A_1 = \begin{bmatrix} 1.6 & 0.8 & 1.3 \\ 2.6 & 2.2 & 1.7 \end{bmatrix} \quad b_1 = \begin{bmatrix} 1.6 \\ 1.2 \\ 0.4 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 1.8 & 1.6 & 1.4 \\ 1.3 & 1.2 & 2.7 \end{bmatrix} \quad b_2 = \begin{bmatrix} 1.6 \\ 1.5 \\ 2.6 \end{bmatrix}$$

The follower has $M = 3$ variables and its objective and constraints are given by

$$Q_y = \begin{bmatrix} 2.5 & 0 & 0 \\ 0 & 3.6 & 0 \\ 0 & 0 & 4.6 \end{bmatrix} \quad B = \begin{bmatrix} 2.3 & 1.4 & 2.6 \\ 1.3 & 2.1 & 1.7 \\ 2.5 & 1.9 & 1.4 \\ 1.3 & 2.4 & 1.6 \end{bmatrix} \quad L = \begin{bmatrix} 1.3 & 2.4 & 1.8 \\ 1.3 & 2.4 & 1.8 \\ 1.3 & 2.4 & 1.8 \\ 1.3 & 2.4 & 1.8 \end{bmatrix}$$

A.3.2 Data Set 2

We consider $N = 3$ leader with each $n_1 = n_2 = n_3 = 2$ variables. The objectives of the leader are given by

$$Q_1 = \begin{bmatrix} 2.5 & 1.6 \\ 1.6 & 3.8 \end{bmatrix} \quad Q_2 = \begin{bmatrix} 2.9 & 1.3 \\ 1.3 & 1.8 \end{bmatrix} \quad Q_3 = \begin{bmatrix} 3.2 & 2.3 \\ 2.3 & 2.6 \end{bmatrix} \quad a = \begin{bmatrix} 0.4 \\ 1.6 \\ 2.6 \end{bmatrix}$$

Each leader has $m_1 = m_2 = m_3 = 3$ linear constraints $g_\nu = A_\nu^T x_\nu + b_\nu \leq 0$ with

$$\begin{aligned} A_1 &= \begin{bmatrix} 1.6 & 0.8 & 1.3 \\ 2.6 & 2.2 & 1.7 \end{bmatrix} & A_2 &= \begin{bmatrix} 1.8 & 1.6 & 1.4 \\ 1.3 & 1.2 & 2.7 \end{bmatrix} & A_3 &= \begin{bmatrix} 2.3 & 1.9 & 1.6 \\ 1.3 & 1.7 & 2.7 \end{bmatrix} \\ b_1 &= \begin{bmatrix} 1.6 & 1.2 & 0.4 \end{bmatrix}^\top & b_2 &= \begin{bmatrix} 1.6 & 1.5 & 2.6 \end{bmatrix}^\top & b_3 &= \begin{bmatrix} 1.5 & 0.3 & 1.8 \end{bmatrix}^\top \end{aligned}$$

The follower has $M = 3$ variables and its objective and constraints are given by

$$Q_y = \begin{bmatrix} 3.7 & 0 & 0 \\ 0 & 2.6 & 0 \\ 0 & 0 & 0.7 \end{bmatrix} \quad B = \begin{bmatrix} 0.8 & 2.1 & 1.3 \\ 1.5 & 2.3 & 0.7 \\ 1.5 & 0.9 & 2.4 \\ 1.8 & 2.3 & 3.6 \\ 1.3 & 1.7 & 1.7 \\ 1.1 & 2.6 & 1.6 \end{bmatrix} \quad L = \begin{bmatrix} 0.8 & 2.1 & 1.3 \\ 1.5 & 2.3 & 0.7 \\ 1.5 & 0.9 & 2.4 \\ 1.8 & 2.3 & 3.6 \\ 0.5 & 1.1 & 2.1 \\ 1.2 & 1.5 & 1.8 \end{bmatrix}$$

Acknowledgements

This project was supported by DFG (STE2063/2-1).

References

[1] E. Allevi, D. Aussel, and R. Riccardi. On an equilibrium problem with complementarity constraints formulation of pay-as-clear electricity market with demand elasticity. *Journal of Global Optimization*, 70(2):329–346, Feb 2018. ISSN 1573-2916.

[2] D. Aussel, M. Cervinka, and M. Marechal. Deregulated electricity markets with thermal losses and production bounds: models and optimality conditions. *RAIRO-Oper. Res.*, 50(1):19–38, 2016.

[3] D. Aussel, P. Bendotti, and M. Pištěk. Nash equilibrium in a pay-as-bid electricity market: Part 1 - existence and characterization. *Optimization*, 66(6):1013–1025, 2017.

[4] D. Aussel, P. Bendotti, and M. Pištěk. Nash equilibrium in a pay-as-bid electricity market: Part 2 - best response of a producer. *Optimization*, 66(6):1027–1053, 2017.

[5] A. M. Bagirov, L. Jin, N. Karmitsa, A. AlNuaimat, and N. Sultanova. Subgradient method for nonconvex nonsmooth optimization. *Journal of Optimization Theory and Applications*, 157 (2):416–435, May 2013. ISSN 1573-2878.

[6] M. Cervinka. *Hierachical Structures in Equilibrium Problems.* PhD thesis, CharlesUniversity Prag, 2008.

[7] F. Clarke. *Optimization and Nonsmooth Analysis.* Society for Industrial and Applied Mathematics, 1990.

[8] F. Facchinei and J. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer Series in Operations Research and Financial Engineering. Springer New York, 2007. ISBN 9780387218151.

[9] M. Guignard. Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space. *SIAM Journal on Control*, 7(2):232–241, 1969.

[10] T. Harks, M. Schroder, and D. Vermeulen. Toll Caps in Privatized Road Networks. *ArXiv e-prints*, Feb. 2018.

[11] R. Henrion, J. Outrata, and T. Surowiec. Analysis of m-stationary points to an epec modeling oligopolistic competition in an electricity spot market. *ESAIM: Control, Optimisation and Calculus of Variations*, 18(2):295317, 2012.

[12] M. Hu and M. Fukushima. Existence, uniqueness, and computation of robust nash equilibria in a class of multi-leader-follower games. *SIAM Journal on Optimization*, 23(2):894–916, 2013.

[13] X. Hu and D. Ralph. Using epecs to model bilevel games in restructured electricity markets with locational prices. *Operations Research*, 55(5):809–827, 2007.

[14] A. Koh and S. Shepherd. Tolling, collusion and equilibrium problems with equilibrium constraints. 2010. ISSN 1825-3997.

[15] S. Leyffer and T. Munson. Solving multi-leader-common-follower games. *Optimization Methods and Software*, 25(4):601–623, 2010.

[16] H. Nikaidô and K. Isoda. Note on non-cooperative convex game. *Pacific Journal of Mathematics*, 5(5):807–815, 1955.

[17] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized nash equilibria, and multi-leader-follower games. *Computational Management Science*, 2(1):21–56, Jan 2005. ISSN 1619-6988.

[18] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000. ISSN 0364765X, 15265471.

[19] S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.

[20] C.-L. Su. A sncp method for solving equilibrium problems with equilibrium constraints. *Computing in Economics and Finance 150*, 2005.

[21] J. H. Wang, D. M. Chiu, and J. C. S. Lui. A game-theoretic analysis of the implications of overlay network traffic on ips peering. *Computer Networks*, 2008.