

DFG Deutsche
Forschungsgemeinschaft
Priority Programme 1962

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

Michael Herty, Sonja Steffensen, Anna Thünen



Preprint Number SPP1962-070r

received on October 21, 2019

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

October 21, 2019

Abstract

We derive Nash 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 a smooth Nash equilibrium problem. The existence and uniqueness of solutions are proven for all smoothing parameters. Accumulation points of Nash equilibria exist for a decreasing sequence of these smoothing parameters and we show that these candidates fulfill the conditions of s-stationarity and are Nash equilibria to the multi-leader-follower game. Finally, we propose an update on the leader variables for efficient computation and numerically compare nonsmooth Newton and subgradient methods.

Keywords: Multi-Leader-Follower Games, Nash Equilibria, 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 as a generalization of the Stackelberg game which includes a single leader. These models serve as an analytical tool for studying the strategic behavior of noncooperative individuals. 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 models have increasingly become the focus of interest among mathematicians as well as scientists in other fields such as operations research, robotics, and computer science [3, 30, 16]. However, compared to the knowledge of other classical game models, little is known so far concerning existence and uniqueness theory as well as suitable numerical solution methods.

The structure of MLFGs can be seen as an equilibrium problem with equilibrium constraints (EPEC), as the optimization problems of the followers might be replaced by the corresponding optimality conditions. The admissibility of this approach is discussed in [2]. Recently, the competition

on the electric power market is described by EPEC [1, 4, 5, 11, 13]. Here, the power generators and consumers are the leaders 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 general MLFGs or EPECs, analyzing the existence, the uniqueness and characterizations of equilibria: Early work by Sherali [25] generalizes Stackelberg games in the setting of Cournot competition. Existence theory of equilibria for identical leaders and sufficiency conditions for convex follower reaction functions is provided here. Su [29] extended this work to a two-period forward market model and proposes an snpc algorithm in [28]. Later, further generalizations in the Cournot setting include the incorporation of stochasticity by DeMiguel and Xu and the application to telecommunication industry [7].

In [17, 18], Kulkarni and Shanbhag introduce a more general setting for MLFGs and its reformulation with shared constraints. The relation of Nash equilibria of the MLFG and weaker solution concepts of the reformulation are extensively discussed here, many results rely on potentiality of the players' objectives.

Pang and Fukushima [22] present a reformulation as generalized Nash equilibrium problem and quasi-variational inequality. In another work, Hu and Ralph [13] apply a standard result for the existence of pure strategy Nash equilibria in a particular case of an MLFG related to an electricity market model. Leyffer and Munson [19] describe various reformulations in terms of mathematical programs with equilibrium constraints (MPEC), nonlinear programs, and nonlinear complementarity problems. In a very recent paper [15], Kim and Ferris propose a reformulation of MLFG using extended mathematical programming which allows the usage of the complementarity problem solver PATH [8, 10].

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 MLFGs. Further, they propose a uniqueness result for an 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 modification translates into equilibrium conditions or to a nonsmooth Nash game formulation. The existence of Nash equilibria can be proven with suitable convexity assumptions. Furthermore, we propose a smoothed Nash game formulation and prove uniqueness of Nash equilibria for an arbitrary number of players. The main result of this article is that it can be shown that these smooth problems are indeed approximating Nash equilibria of the MLFG. 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 develop an equivalent nonsmooth Nash equilibrium problem (NEP) for which we can prove existence of equilibria. In Section 3, smoothing of the best response of the follower leads to a differentiable NEP formulation, which allows to characterize solutions by KKT conditions. In Section 4, it is

demonstrated that a limit of Nash equilibria of the smooth problems converges also in the multipliers for decreasing smoothing parameters and that s-stationarity conditions are satisfied. Eventually, this s-stationary point are proven to be Nash equilibria to the MLFG. In Section 5, we introduce our general approach to computing Nash equilibria. In Section 6, we present the numerical results of the proposed methods.

2 Existence of Nash Equilibria for MLFGs

We consider an MLFG, where the follower's game is modeled by the optimization problem

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

where $Q_y \in \mathbb{R}^{m \times m}$ is a positive definite diagonal matrix and $b_i, l_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are differentiable functions for $i = 1, \dots, m$. The leader problems are given for $\nu = 1, \dots, N$ by

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

with nonempty, convex, and closed strategy sets X_ν which we assume to be described by smooth functions $g_\nu : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}^{m_\nu}$ such that $X_\nu = \{x_\nu \in \mathbb{R}^{n_\nu} | g_\nu(x_\nu) \leq 0\}$. The quadratic objective θ_ν is strictly convex with $Q_\nu \in \mathbb{R}^{n_\nu \times n_\nu}$ symmetric positive definite, $c_\nu \in \mathbb{R}^{n_\nu}$, and $a \in \mathbb{R}_+^m$. The multistrategy vector of all players is denoted by $x = (x_\nu)_{\nu=1}^N \in \mathbb{R}^n$ and the rival's strategies to player ν by $x_{-\nu} = (x_i)_{i=1, i \neq \nu}^N \in \mathbb{R}^{n-n_\nu}$. The follower's strategy y in (2) is a best response to the leader strategies and couples implicitly to the leaders' problems.

For simple notation, the model is introduced as a Multi-Leader-(Single-)Follower game. However, this can be extended to multiple followers:

Remark 2.1. (*Multiple Followers*) For $j = 1, \dots, N_F$ let the j -th follower's optimization problem be

$$\min_{y_j \in \mathbb{R}^{m_j}} \frac{1}{2} y_j^\top Q_y^j y_j - b_j(x)^\top y_j \quad \text{s.t.} \quad y_j \geq l_j(x),$$

with $Q_y^j \in \mathbb{R}^{m_j \times m_j}$ a positive definite diagonal matrix and $b_j, l_j : \mathbb{R}^n \rightarrow \mathbb{R}^{m_j}$ convex in every component and differentiable. This structure is called potential game and it can be equivalently reformulated as a single optimization problem by summing the objectives and concatenating the constraints, c.f. [20].

Since the follower's problem has a strictly convex objective and a convex strategy set, we state its unique solution in the following lemma:

Lemma 2.2 (Follower's Best Response). *The follower's optimization problem (1) has a unique solution y^* for any given leader strategy vector $x \in \mathbb{R}^n$. In particular, this best response function is*

$$y^*(x) = \max \{Q_y^{-1} b(x), l(x)\}. \quad (3)$$

Proof. The objective and the feasible set are convex for any $x \in \mathbb{R}^n$, in particular, the constraints are linear. 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 the complementarity expression $0 \leq \lambda = Q_y y - b(x) \perp y - l(x) \geq 0$. We assumed Q_y to be positive definite and diagonal; therefore, we can write equivalently as $0 \leq y - Q_y^{-1} b(x) \perp y - l(x) \geq 0$ and $0 = \min \{y - Q_y^{-1} b(x), y - l(x)\}$. We obtain (3) by extracting y and changing min to max. \square

We remark that the min and max operator are applied componentwise to a vector. If the data b, l are chosen to be componentwise convex, then also the components of best response function $y^*(x)$ are convex in x as a maximum of convex functions.

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

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

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

Theorem 2.3 (Existence of Nash Equilibria for Compact Strategy Sets). *Assume that the non-smooth Nash equilibrium problem in (NEP) has a convex and compact joint strategy set $X = X_1 \times \dots \times X_N$, where all X_ν are nonempty. Further, assume that b, l are componentwise convex functions on X . 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 [21, Theorem 3.1] are fulfilled. \square

Alternatively, Theorem 4.5 includes a constructive existence proof. In both cases, the existence result relies strongly on the compactness of the strategy sets ensure global solutions to the players'

optimization problems. However, this is clearly not a necessary condition for the existence or uniqueness of Nash equilibria to (1,2).

In the remainder of the section, we exchange the compactness assumption with an coercivity assumption on the best response of the followers exploiting the potentiality of the leader objectives. Therefore, let the generalized potential function be

$$\Theta(x) = \sum_{\nu=1}^N \left[\frac{1}{2} x_{\nu}^{\top} Q_{\nu} x_{\nu} + c_{\nu}^{\top} x_{\nu} \right] + \varphi(x),$$

$$\varphi(x) = \sum_{i=1}^m a_i \max \left\{ (Q_y^{-1} b(x))_i, l_i(x) \right\}.$$

With this, we can formulate the major assumption:

Assumption 2.4. *Assume, it exists $\rho \geq 0$ such that for all $x \in X$ with $\|x\| > \rho$ and for $\omega_1, \omega_2 \in \mathbb{R}$ it holds*

$$\varphi(x) \geq \omega_1 \|x\| + \omega_2.$$

This assumption may be understood as boundedness from below by a linear function.

Lemma 2.5 (Coercivity of the Generalized Potential Θ). *Let Assumption 2.4 hold, then the generalized potential function Θ is coercive.*

Proof. We introduce a short hand for the block diagonal matrix $Q = \text{diag}(Q_1, \dots, Q_N)$ and the concatenation of c_1, \dots, c_N as c . Then we have

$$\Theta(x) = x^{\top} Q x + c^{\top} x + \varphi(x) \geq \mu \|x\|^2 - \|c\| \|x\| + \varphi(x),$$

where μ is the smallest eigenvalue of Q , recall that $\mu > 0$ since all Q_{ν} are positive definite. We rewrite the the right hand side with Assumption 2.4 and get

$$\Theta(x) \geq \mu \|x\|^2 - \|c\| \|x\| + \omega_1 \|x\| + \omega_2,$$

As $\|x\| \rightarrow \infty$, the right hand side is dominated by a quadratic term with $\mu \|x\|^2 \rightarrow \infty$ and we conclude

$$\lim_{\|x\| \rightarrow \infty} \Theta(x) = \infty,$$

i.e. the generalized potential function is coercive. □

The Assumption 2.4 and the concluded coercivity property allows us to formulate an existence result without requiring compactness.

Theorem 2.6 (Existence of Nash Equilibria for Coercive Generalized Potential). *Let Assumption 2.4 hold and x^* be a global minimizer of the generalized potential Θ , then x^* is also a Nash equilibrium of the MLFG (1-2).*

Proof. Since Assumption 2.4 holds, the generalized potential Θ admits a global minimizer on any closed joint strategy set X . We demonstrate that x^* is a Nash equilibrium by contradiction. Therefore, assume the converse is true, i.e. it exists a leader $\hat{\nu}$ and a strategy $\hat{x}_{\hat{\nu}} \in X_{\hat{\nu}}$ such that:

$$\theta_{\hat{\nu}}(\hat{x}_{\hat{\nu}}, x_{-\hat{\nu}}^*) < \theta_{\hat{\nu}}(x_{\hat{\nu}}^*, x_{-\hat{\nu}}^*),$$

or more detailed,

$$\frac{1}{2}\hat{x}_{\hat{\nu}}^\top Q_{\hat{\nu}}\hat{x}_{\hat{\nu}} + c_{\hat{\nu}}^\top \hat{x}_{\hat{\nu}} + \varphi(\hat{x}_{\hat{\nu}}, x_{-\hat{\nu}}^*) < \frac{1}{2}x_{\hat{\nu}}^{*\top} Q_{\hat{\nu}}x_{\hat{\nu}}^* + c_{\hat{\nu}}^\top x_{\hat{\nu}}^* + \varphi(x_{\hat{\nu}}^*, x_{-\hat{\nu}}^*).$$

With this, we can overestimate the following expression

$$\begin{aligned} \Theta(\hat{x}_{\hat{\nu}}, x_{-\hat{\nu}}^*) &= \sum_{\nu=1, \nu \neq \hat{\nu}}^N \left[\frac{1}{2}x_{\nu}^{*\top} Q_{\nu}x_{\nu}^* + c_{\nu}^\top x_{\nu}^* \right] + \frac{1}{2}\hat{x}_{\hat{\nu}}^\top Q_{\hat{\nu}}\hat{x}_{\hat{\nu}} + c_{\hat{\nu}}^\top \hat{x}_{\hat{\nu}} + \varphi(\hat{x}_{\hat{\nu}}, x_{-\hat{\nu}}^*) \\ &< \sum_{\nu=1}^N \left[\frac{1}{2}x_{\nu}^{*\top} Q_{\nu}x_{\nu}^* + c_{\nu}^\top x_{\nu}^* \right] + \varphi(x_{\hat{\nu}}^*, x_{-\hat{\nu}}^*) = \Theta(x_{\hat{\nu}}^*, x_{-\hat{\nu}}^*), \end{aligned}$$

which is a contradiction to the optimality of x^* . \square

In the remainder of this article, we assume b and l to be linear:

Assumption 2.7. *The data is assumed to be linear:*

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

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} .

This linearity combined with Assumption 2.4 directly yields existence of Nash equilibria.

Lemma 2.8 (Existence of Nash Equilibria for Linear Data b, l). *If Assumption 2.4 and 2.7 hold, then it exists a Nash equilibrium to the MLFG.*

Proof. With linear data b, l , φ gets:

$$\varphi(x) = \sum_{i=1}^m a_i \max \left\{ \left(Q_y^{-1} B^\top x \right)_i, \left(L^\top x \right)_i \right\},$$

which is a linear combination of the maxima of two linear functions. Thus, it can be underestimated by an other linear function. Therefore, Assumption 2.4 holds and Theorem 2.6 applies. \square

In [26] there is a counterexample for the structure of MLFG in (1,2) with linear data b, l . The data and X are such that the non-differentiable parts of the objectives of the NEP reformulation lie either on the boundary of X or outside the feasible domain. Or with other words, the max operator can be uniquely evaluated on the interior of the feasible set. In particular, existence and uniqueness of Nash equilibria is derived by a formulation as variational inequality. In case of diagonal Q_{ν} , the Nash equilibrium can be explicitly computed in the presented setting.

3 Existence and Uniqueness 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 2.2, we formulate the follower's KKT conditions with the linear data:

$$0 \leq y - Q_y^{-1}B^\top x \perp y - L^\top x \geq 0, \quad (4)$$

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 $\varepsilon > 0$:

$$\phi_\varepsilon(\alpha, \beta) = \alpha + \beta - \tilde{\phi}_\varepsilon(\alpha - \beta). \quad (5)$$

An example for a smooth NCP function with convex $\tilde{\phi}_\varepsilon$ is $\phi_\varepsilon^p(\alpha, \beta) = \alpha + \beta - \sqrt[p]{(\alpha - \beta)^p + (2\varepsilon)^p}$ for an even number $p \in \mathbb{N}$, this coincides for $p = 2$ with the smooth minimum function. This structure also captures also NCP functions related to the relaxation scheme for mathematical programs with equilibrium constraints (MPEC) introduced by Steffensen and Ulbrich [27]. Besides regularity in α and β , we require the smooth NCP function ϕ_ε to be continuous in the smoothing parameter ε .

If we apply (5) on the KKT system (4), we obtain (component wise) for $\varepsilon > 0$ the expression

$$0 = y - Q_y^{-1}B^\top x + y - L^\top x - \tilde{\phi}_\varepsilon \left(y - Q_y^{-1}B^\top x - y + L^\top x \right). \quad (6)$$

Therefore, we write the best response function as:

$$y_\varepsilon(x) = \frac{1}{2} \left[\left(L^\top + Q_y^{-1}B^\top \right) x + \tilde{\phi}_\varepsilon \left(\left(L^\top - Q_y^{-1}B^\top \right) x \right) \right]. \quad (7)$$

The smoothed best response function y_ε plugged into the leader's objectives yields a smooth Nash equilibrium problem with positive smoothing parameter ε ; for $\nu = 1, \dots, N$, we have

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

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

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

Proof. Due to [9, 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 $\text{VI}(X, \theta^\varepsilon)$, 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 [9, Theorem 2.3.3]. The remainder of the proof demonstrates this.

We introduce a short hand for the linear term $A = L^\top - Q_y^{-1}B^\top$, the block diagonal matrix $Q = \text{diag}(Q_1, \dots, Q_N)$, and the concatenation of c_1, \dots, c_N as c . Let $x, \hat{x} \in X$ be distinct, then we have

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

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

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

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 $\text{VI}(X, \theta^\varepsilon)$ has exactly one solution which is the unique Nash equilibrium of the smoothed game $\text{NEP}(\varepsilon)$. \square

The remainder of this section is dedicated to the characterization of the unique Nash equilibrium. Therefore, we recall the assumptions made for the smooth Nash equilibrium problem $\text{NEP}(\varepsilon)$ for positive smoothing parameter ε :

Assumption 3.2. *We assume that the data of the MLFG and its smooth Nash game reformulation satisfy for $\nu = 1, \dots, N$ the following properties.*

- $Q_\nu \in \mathbb{R}^{n_\nu \times n_\nu}$ symmetric positive definite,

- $c_\nu \in \mathbb{R}^{n_\nu}$,
- $a \in \mathbb{R}_+^m$,
- $X_\nu = \{x_\nu \in \mathbb{R}^{n_\nu} | g_\nu(x_\nu) \leq 0\} \subseteq \mathbb{R}^{n_\nu}$ nonempty, convex, and closed;
- $g_\nu : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}^{m_\nu}$ at least twice differentiable, and convex;
- $Q_y \in \mathbb{R}^{m \times m}$ positive definite and diagonal,
- $b(x) = B^\top x$, $l(x) = L^\top x$, with $B, L \in \mathbb{R}^{n \times m}$,
- 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 $\varepsilon > 0$.

Based on these assumptions, we characterize the Nash equilibrium of $\text{NEP}(\varepsilon)$ by its KKT conditions in the following lemma:

Lemma 3.3 (KKT of $\text{NEP}(\varepsilon)$). *Given the Assumptions 3.2 and some constraint qualification for X_ν , the KKT conditions are necessary and sufficient for the global minimizer of each leader problem in $\text{NEP}(\varepsilon)$. In particular, we have the KKT conditions of player ν 's optimization problem for $\nu = 1, \dots, N$:*

$$0 = Q_\nu x_\nu + c_\nu + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top_\nu a + \frac{1}{2} \sum_{i=1}^m a_i (L^\top - Q_y^{-1}B^\top)^\top_{\nu,i} \tilde{\phi}'_\varepsilon \left([(L^\top - Q_y^{-1}B^\top)x]_i \right) + \nabla_{x_\nu} g_\nu(x_\nu) \lambda_\nu, \quad (8a)$$

$$0 = \min \{ \lambda_\nu, -g_\nu(x_\nu) \}, \quad (8b)$$

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

Proof. KKT is necessary and sufficient for every leader problem in presence of a constraint qualification, since the objectives are strictly convex and the strategy set are convex and closed. 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. (7). \square

The imposed constraint qualification is specified in the following section.

4 Relation between MLFG/NEP and $\text{NEP}(\varepsilon)$

In this section, we study the relationship between the original MLFG (or its equivalent NEP) and the smoothed problem $\text{NEP}(\varepsilon)$. Our aim is to demonstrate that Nash equilibria of $\text{NEP}(\varepsilon)$ are converging to a Nash equilibrium of the original problem as the smoothing parameter $\varepsilon \rightarrow 0$.

Instead of looking at optimality conditions for the nonsmooth formulation, we are considering stationarity concepts for mathematical programs with complementarity constraints (MPCC) next.

Definition 4.1 (Mathematical Programs with Complementarity Constraints). *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, and $G_1, G_2 : \mathbb{R}^n \rightarrow \mathbb{R}^l$ be smooth functions. Then we call:*

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

a mathematical program with complementarity constraints.

In fact, each leader problem (2) can be formulated as an MPCC:

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

with the complementarity constraints $G_1(x_\nu, x_{-\nu}, y) = y - (Q_y^{-1} B^\top)^\top x$ and $G_2(x_\nu, x_{-\nu}, y) = y - L^\top x$. For $\nu = 1, \dots, N$, the (MPCC) $_\nu$ form together the generalized Nash equilibrium problem (GNEP) formulation of the MLFG, where the complementarity constraints are shared constraints.

Similar to KKT points, there is a variety of stationary concepts for MPCC, we introduce the strongest one here: strongly stationary points. We adapt [24, Theorem 2].

Definition 4.2 (S-Stationarity). *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:*

$$\begin{aligned} 0 &= \nabla_z f(\bar{z}) + \sum_{i=1}^m \lambda_i \nabla_z g_i(\bar{z}) - \sum_{i=1}^l \Gamma_{1,i} \nabla_z G_{1,i}(\bar{z}) - \sum_{i=1}^l \Gamma_{2,i} \nabla_z G_{2,i}(\bar{z}), \\ g(\bar{z}) &\leq 0, \\ \lambda &\geq 0, \\ g_i(\bar{z}) \lambda_i &= 0, \quad i = 1, \dots, m, \\ \min \{G_{1,i}(\bar{z}), G_{2,i}(\bar{z})\} &= 0, \quad i = 1, \dots, l, \\ G_{1,i}(\bar{z}) \Gamma_{1,i} &= 0, \quad i = 1, \dots, l, \\ G_{2,i}(\bar{z}) \Gamma_{2,i} &= 0, \quad i = 1, \dots, l, \\ \Gamma_{1,i}, \Gamma_{2,i} &\geq 0, \quad i : G_{1,i}(\bar{z}) = G_{2,i}(\bar{z}) = 0. \end{aligned} \tag{9}$$

Besides suitable stationarity concepts, MPCCs also require proper constraint qualifications. Here, the Constant Rank Constraint Qualification (CRCQ) for MPCC is defined similarly to its version for MPEC in [27, Def. 2.2]. Originally, the CRCQ was introduced for NLP by [14].

Definition 4.3. (MPCC-CRCQ) *The constraint qualification MPCC-CRCQ holds in the feasible*

point \bar{z} of (MPCC), if for every \mathcal{K}_g , \mathcal{K}_1 , and \mathcal{K}_2 with

$$\begin{aligned}\mathcal{K}_g &\subseteq I_g(\bar{z}) = \{i \in \{1, \dots, m\} | g_i(\bar{z}) = 0\}, \\ \mathcal{K}_1 &\subseteq I_1(\bar{z}) = \{i \in \{1, \dots, l\} | G_{1,i}(\bar{z}) = 0\}, \\ \mathcal{K}_2 &\subseteq I_2(\bar{z}) = \{i \in \{1, \dots, l\} | G_{2,i}(\bar{z}) = 0\},\end{aligned}$$

there exists a neighborhood $\mathcal{U}(\bar{z})$ such that for every $z \in \mathcal{U}(\bar{z})$ the family of gradient vectors

$$\{\nabla g_i(z) | i \in \mathcal{K}_g\} \cup \{\nabla G_{1,i}(z) | i \in \mathcal{K}_1\} \cup \{\nabla G_{2,i}(z) | i \in \mathcal{K}_2\},$$

has the same rank as the family

$$\{\nabla g_i(\bar{z}) | i \in \mathcal{K}_g\} \cup \{\nabla G_{1,i}(\bar{z}) | i \in \mathcal{K}_1\} \cup \{\nabla G_{2,i}(\bar{z}) | i \in \mathcal{K}_2\}.$$

In the following theorem we show that a limit of the Nash equilibria of the smoothed problems is indeed strongly stationary with respect to the leader problems.

Theorem 4.4 (S-stationarity and Convergence of Multipliers). *Let $(\varepsilon_k)_k$ be a positive sequence with $\varepsilon_k \rightarrow 0$ and let $(x^*(\varepsilon_k))_k$ be the associated sequence of the unique Nash equilibria to NEP(ε_k). Assume this sequence converges with $x^*(\varepsilon_k) \rightarrow x^*(0)$ and every subvector of the limit $x^*(0)$ satisfies MPCC-CRCQ for its leader's MPCC $_\nu$.*

Then, it exists a bounded sequence of multipliers associated to $(x^(\varepsilon_k))_k$ and the limit $x^*(0)$ is a strongly stationary point for (MPCC $_\nu$) for $\nu = 1, \dots, N$.*

Proof. The proof consists of three major parts: First, we introduce notation and modest conclusions. Second, we demonstrate that there exists a bounded sequence of multipliers associated to $(x^*(\varepsilon_k))_k$. Third, we verify the conditions of strong stationarity by constructing suitable multipliers.

(i) In order to keep the notation simple, let $x_\nu^k = x_\nu^*(\varepsilon_k)$ and the limit $\bar{x}_\nu = x_\nu^*(0)$ and respectively the concatenations x^k and \bar{x} , both without the player index. Further we introduce the short forms for $k \in \mathbb{N}$ and $i = 1, \dots, m$:

$$z_i^k = \left((L^\top - Q_y^{-1} B^\top) x^k \right)_i, \bar{z}_i = \left((L^\top - Q_y^{-1} B^\top) \bar{x} \right)_i, \text{ and } \xi_i^k = \tilde{\phi}'_{\varepsilon_k}(z_i^k),$$

and the concatenations are denoted by z^k , \bar{z} , and ξ^k , respectively. Recall, it holds that $z_i^k \rightarrow \bar{z}_i$ and $|\xi_i^k| \leq 1$ for $i = 1, \dots, m$ and all $k \in \mathbb{N}$.

Due to stationarity of x_ν^k for all ν and $k \in \mathbb{N}$, we use Lemma 3.3, i.e. (8), and conclude that there exists multipliers $\lambda_\nu^k \in \mathbb{R}^{m_\nu}$ such that:

$$0 = Q_\nu x_\nu^k + c_\nu + \frac{1}{2} (L^\top + Q_y^{-1} B^\top)_{\nu, \cdot}^\top a + \frac{1}{2} \sum_{i=1}^m a_i (L^\top - Q_y^{-1} B^\top)_{\nu, i}^\top \xi_i^k + \nabla_{x_\nu} g_\nu(x_\nu^k) \lambda_\nu^k, \quad (10a)$$

$$0 = \min \left\{ \lambda_\nu^k, -g_\nu(x_\nu^k) \right\}, \quad (10b)$$

Hence, by rearranging the sum in (10a) we have:

$$0 = Q_\nu x_\nu^k + c_\nu + \sum_{i=1}^m L_{\nu,i}^\top \frac{a_i}{2} (1 + \xi_i^k) + \sum_{i=1}^m Q_y^{-1} B_{\nu,i}^\top \frac{a_i}{2} (1 - \xi_i^k) + \sum_{j=1}^{m_\nu} \nabla_{x_\nu} g_{\nu,j}(x_\nu^k) \lambda_{\nu,j}^k, \quad (11)$$

Next, we construct a multiplier vector $\hat{\lambda}_\nu^k$ for all ν and sufficiently large k such that the following inclusion holds:

$$\text{supp} \left(\hat{\lambda}_\nu^k \right) \subseteq \text{supp} \left(\lambda_\nu^k \right) \subseteq I_{g_\nu}(\bar{x}_\nu) = \{j \mid g_{\nu,j}(\bar{x}_\nu) = 0\},$$

where $\text{supp}(\cdot)$ denotes the support of a vector, i.e. the set of indices of its nonzero entries, the multiplier $\hat{\lambda}_\nu^k$ shall satisfy (11) and the family of vectors

$$\left\{ \nabla_{x_\nu} g_{\nu,j}(x_\nu^k) \mid j \in \text{supp} \left(\hat{\lambda}_\nu^k \right) \right\} \quad (12)$$

is linearly independent.

(ii) Now, we demonstrate by contradiction that the multiplier sequence $\left(\hat{\lambda}_\nu^k \right)_k$ is bounded for all ν . Therefore, it is assumed that the converse is true, i.e. there exists $\hat{\nu}$ such that $\left(\hat{\lambda}_{\hat{\nu}}^k \right)_k$ is unbounded. By assumption, we have $\|\hat{\lambda}_{\hat{\nu}}^k\| \rightarrow \infty$ as $k \rightarrow \infty$ and thus it holds, that $\|\hat{\lambda}_{\hat{\nu}}^k\| > 0$ for sufficiently large k . Therefore, we can define the auxiliary sequence

$$\tilde{\lambda}_{\hat{\nu}}^k = \frac{\hat{\lambda}_{\hat{\nu}}^k}{\|\hat{\lambda}_{\hat{\nu}}^k\|}. \quad (13)$$

Clearly, all elements of this sequence are normalized, i.e. $\|\tilde{\lambda}_{\hat{\nu}}^k\| = 1$. This in turn implies that the sequence is bounded and admits a convergent subsequence $\left(\tilde{\lambda}_{\hat{\nu}}^l \right)_{l \in K} \rightarrow \tilde{\lambda}_{\hat{\nu}}$ with $K \subseteq \mathbb{N}$. It follows, that also the limit of the subsequence satisfies $\|\tilde{\lambda}_{\hat{\nu}}\| = 1$ and there exists an index $j_0 \in \{1, \dots, m_{\hat{\nu}}\}$ with $\tilde{\lambda}_{\hat{\nu},j_0} > 0$ such that $\text{supp} \left(\tilde{\lambda}_{\hat{\nu}} \right) \neq \emptyset$. Now, divide (11) by norm of the unbounded sequence $\|\hat{\lambda}_{\hat{\nu}}^k\|$ and take the limit $k \rightarrow \infty$. Note, that the first four terms of (11) are bounded and thus they vanish in the limit. It remains to compute the limit of the last term:

$$0 = \lim_{k \rightarrow \infty, l \in K} \frac{1}{\|\hat{\lambda}_{\hat{\nu}}^k\|} \sum_{j \in \text{supp}(\hat{\lambda}_{\hat{\nu}}^k)} \nabla_{x_{\hat{\nu}}} g_{\hat{\nu},j}(\bar{x}_{\hat{\nu}}) \hat{\lambda}_{\hat{\nu},j}^k,$$

with (13) and (for sufficiently large k) $\text{supp} \left(\tilde{\lambda}_{\hat{\nu}} \right) \subseteq \text{supp} \left(\tilde{\lambda}_{\hat{\nu}}^k \right) = \text{supp} \left(\hat{\lambda}_{\hat{\nu}}^k \right)$, we get

$$0 = \sum_{j \in \text{supp}(\tilde{\lambda}_{\hat{\nu}})} \nabla_{x_{\hat{\nu}}} g_{\hat{\nu},j}(\bar{x}_{\hat{\nu}}) \tilde{\lambda}_{\hat{\nu},j}.$$

This implies with $\text{supp} \left(\tilde{\lambda}_{\hat{\nu}} \right) \neq \emptyset$, that the set of vectors

$$\left\{ \nabla_{x_{\hat{\nu}}} g_{\hat{\nu},j}(\bar{x}_{\hat{\nu}}) \mid j \in \text{supp} \left(\tilde{\lambda}_{\hat{\nu}} \right) \right\},$$

is linearly dependent. Thus, by MPCC-CRCQ and $\text{supp}(\tilde{\lambda}_\nu) \subseteq \text{supp}(\hat{\lambda}_\nu^k)$, also

$$\left\{ \nabla_{x_\nu} g_{\nu,j}(x_\nu^k) \mid j \in \text{supp}(\hat{\lambda}_\nu^k) \right\},$$

is linearly dependent, which is a contradiction to the linear independence of (12) for sufficiently large k . Therefore the assumption $(\hat{\lambda}_\nu^k)_k$ being unbounded, because (11) could not be satisfied by such a multiplier sequence. Therefore, the sequence of multipliers associated to $(x^k)_k$ is bounded, i.e. it exist a multiplier vector $\bar{\lambda}$ associated to \bar{x} .

(iii) In the following, we verify the conditions of s-stationarity of Definition 4.2, i.e. (9) exem-
plary for on leader ν , and begin with applying them to (MPCC $_\nu$):

$$0 = \begin{pmatrix} Q_\nu x_\nu + c_\nu \\ a \end{pmatrix} + \sum_{i=1}^m \lambda_i \begin{pmatrix} \nabla_{x_\nu} g_{\nu,i}(x_\nu) \\ 0 \end{pmatrix} - \sum_{i=1}^l \Gamma_{1,i} \begin{pmatrix} -(Q_y^{-1} B^\top)_{\nu,i} \\ e_i \end{pmatrix} - \sum_{i=1}^l \Gamma_{2,i} \begin{pmatrix} -L_{\nu,i} \\ e_i \end{pmatrix}, \quad (14a)$$

$$g_\nu(x_\nu) \leq 0, \quad (14b)$$

$$\lambda_\nu \geq 0, \quad (14c)$$

$$g_{\nu,i}(x_\nu) \lambda_{\nu,i} = 0, \quad i = 1, \dots, m_\nu, \quad (14d)$$

$$\min \left\{ (y - (Q_y^{-1} B^\top)^\top x)_i, (y - L^\top x)_i \right\} = 0, \quad i = 1, \dots, m, \quad (14e)$$

$$(y - (Q_y^{-1} B^\top)^\top x)_i \Gamma_{1,i} = 0, \quad i = 1, \dots, m, \quad (14f)$$

$$(y - L^\top x)_i \Gamma_{2,i} = 0, \quad i = 1, \dots, m, \quad (14g)$$

$$\Gamma_{1,i}, \Gamma_{2,i} \geq 0, \quad i : (y - (Q_y^{-1} B^\top)^\top x)_i = (y - L^\top x)_i. \quad (14h)$$

The remainder of the proof demonstrates that (14) is satisfied for the limit strategy \bar{x}_ν and its multiplier $\bar{\lambda}_\nu$, and we construct the additional multipliers $\Gamma_{1,i}, \Gamma_{2,i}$. We begin with the limit of (11) for $k \rightarrow \infty$ in the sense of a suitable subsequence and get the expression:

$$0 = Q_\nu \bar{x}_\nu + c_\nu + \sum_{i=1}^m L_{\nu,i}^\top \frac{a_i}{2} (1 + \bar{\xi}_i) + \sum_{i=1}^m Q_y^{-1} B_{\nu,i}^\top \frac{a_i}{2} (1 - \bar{\xi}_i) + \sum_{j=1}^{m_\nu} \nabla_{x_\nu} g_{\nu,j}(\bar{x}_\nu) \bar{\lambda}_{\nu,j}, \quad (15)$$

where $\bar{\xi}_i = \lim_{k \rightarrow \infty} \xi_i^k = \lim_{k \rightarrow \infty} \tilde{\phi}'_{\varepsilon_k}(z_k^i)$, note that $\bar{\xi}_i \in \{-1, 0, 1\}$. If we choose the multipliers to the complementarity constraints to be:

$$\bar{\Gamma}_{1,i} = \frac{a_i}{2} (1 + \bar{\xi}_i), \bar{\Gamma}_{2,i} = \frac{a_i}{2} (1 - \bar{\xi}_i),$$

and since $a_i - \bar{\Gamma}_{1,i} - \bar{\Gamma}_{2,i} = a_i - a_i/2(1 + \bar{\xi}_i) - a_i/2(1 - \bar{\xi}_i) = 0$ for all $i = 1, \dots, m_\nu$ and (15), then the condition in (14a) follows.

Feasibility (14b-14d) is due to continuity of g_ν and the convergent subsequence of the multipliers, which gives us $\bar{\lambda}$. The feasibility of the complementarity constraint (14e) is due to choice of $\tilde{\phi}_\varepsilon$, which belongs to a smooth NCP function, c.f. (4-7).

It remains to demonstrate (14f-14h):

In case $G_{1,i}(\bar{x}_\nu, \bar{x}_{-\nu}) > 0$, then by feasibility of \bar{x} we have $G_{2,i}(\bar{x}_\nu, \bar{x}_{-\nu}) = 0$ and thus $\bar{\xi}_i = -1$ such that $\Gamma_{1,i} = 0$. If otherwise $G_{2,i}(\bar{x}_\nu, \bar{x}_{-\nu}) > 0$, then by the feasibility of \bar{x} it holds that $G_{1,i}(\bar{x}_\nu, \bar{x}_{-\nu}) = 0$ and thus $\bar{\xi}_i = -1$, i.e. $\Gamma_{2,i} = 0$. Since both arguments hold for all $i = 1, \dots, m$, this yields (14f-14g).

Moreover, both multipliers satisfy $\bar{\Gamma}_{1,i}, \bar{\Gamma}_{2,i} \geq 0$ for all i because $a_i > 0$ and $|\bar{\xi}_i| \leq 1$; therefore, (14h) holds.

Hence, the strategy \bar{x}_ν is strongly stationary for MPCC_ν and since we derived (14) for an arbitrary leader ν , the proof is complete. \square

In the preceding theorem, it is demonstrated that the sequence of multipliers has accumulation points in presence of a suitable constraint qualification if the primal variables converge. We have seen that the limit of the Nash equilibria of $\text{NEP}(\varepsilon)$ are in fact strongly stationary to the original MLFG. However, the following theorem goes even further by demonstrating that a limit of these Nash equilibria is in fact a Nash Equilibrium of the MLFG.

Theorem 4.5 (A Nash Equilibrium of the MLFG). *Let $(\varepsilon_k)_k$ be a positive sequence with $\varepsilon_k \rightarrow 0$ and let $(x^*(\varepsilon_k))_k$ be the associated sequence of the unique Nash equilibria to $\text{NEP}(\varepsilon_k)$.*

Then any accumulation point of $(x^(\varepsilon_k))_k$ for a positive sequence $\varepsilon_k \rightarrow 0$ is a Nash equilibrium to NEP and therefore of the MLFG.*

Proof. Recall, since $x^*(\varepsilon)$ is the unique Nash equilibrium of $\text{NEP}(\varepsilon)$ for any $\varepsilon > 0$, it holds by the definition of a Nash equilibrium that for all $\nu = 1, \dots, N$:

$$\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) \leq \theta_\nu^\varepsilon(x_\nu, x_{-\nu}^*(\varepsilon)) \text{ for all } x_\nu \in X_\nu.$$

We prove that a limit strategy $x^*(0)$ is a Nash equilibrium to NEP by contradiction. Assume, $x^*(0)$ is not a Nash equilibrium to NEP, then it exists $\hat{x} \in X$ such that:

$$\theta_\nu(x_\nu^*(0), x_{-\nu}^*(0)) > \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0)) \text{ for a leader } \nu.$$

We define the distance as

$$\varrho = \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0)) \tag{16}$$

Recall that all objectives θ_ν^ε are continuous in the strategies $x_\nu, x_{-\nu}$ and the smoothing parameter ε , in particular in $\varepsilon = 0$. Then it exists $\hat{\varepsilon}$ such that for all $\varepsilon \in (0, \hat{\varepsilon})$, the following relations hold for any ν :

$$|\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon))| \leq \frac{\varrho}{6} \quad (\text{continuity in } x_\nu) \tag{17a}$$

$$|\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0))| \leq \frac{\varrho}{6} \quad (\text{continuity in } x_{-\nu}) \tag{17b}$$

$$|\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0)) - \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0))| \leq \frac{\varrho}{6} \quad (\text{continuity in } \varepsilon) \tag{17c}$$

$$|\theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon))| \leq \frac{\varrho}{6} \quad (\text{continuity in } \varepsilon) \tag{17d}$$

$$|\theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0))| \leq \frac{\varrho}{6} \quad (\text{continuity in } x_{-\nu}) \tag{17e}$$

It follows with (17a-17c) and triangle inequality, that

$$\begin{aligned}
\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) &\geq \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0)) - |\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) - \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0))| \\
&\geq \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0)) - |\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon))| \\
&\quad - |\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0))| \\
&\quad - |\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0)) - \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0))|,
\end{aligned}$$

and similar with (17d-17e), we have

$$\begin{aligned}
\theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) &\leq \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0)) + |\theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0))|, \\
&\leq \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0)) + |\theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon))| \\
&\quad + |\theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0))|,
\end{aligned}$$

Subtracting these inequality expressions yields with (16) and (17a-17e)

$$\begin{aligned}
\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) &\geq \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0)) \\
&\quad - |\theta_\nu^\varepsilon(x_\nu^*(\varepsilon), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon))| \\
&\quad - |\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(\varepsilon)) - \theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0))| \\
&\quad - |\theta_\nu^\varepsilon(x_\nu^*(0), x_{-\nu}^*(0)) - \theta_\nu(x_\nu^*(0), x_{-\nu}^*(0))| \\
&\quad - |\theta_\nu^\varepsilon(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon))| \\
&\quad - |\theta_\nu(\hat{x}_\nu, x_{-\nu}^*(\varepsilon)) - \theta_\nu(\hat{x}_\nu, x_{-\nu}^*(0))|, \\
&\geq \varrho - \frac{5}{6}\varrho = \frac{\varrho}{6} > 0,
\end{aligned}$$

which contradicts the assumption that $x^*(\varepsilon)$ is the Nash equilibrium of $\text{NEP}(\varepsilon)$. Therefore the assumption that $x^*(0)$ is not a Nash equilibrium to NEP , which completes the proof. \square

We remark, that this can also be understood as a constructive existence proof of Nash equilibria of (NEP) if X_ν is compact for all $\nu = 1, \dots, N$, as alternative to Theorem 2.3.

Note, that we needed to assume convergence of the Nash equilibria. This strong requirement is weakened on the following two corollaries:

Corollary 4.6 (Compact Strategy Sets). *If in addition X_ν is compact for all $\nu = 1, \dots, N$, the convergence of x_k^* and $x^*(0)$ is not an assumption, because it exists at least one accumulation point.*

Corollary 4.7 (Accumulation Points are Nash Equilibria). *If the limit of x_k^* is non unique, every accumulation point of the sequence is a Nash equilibrium of the MLFG.*

5 Numerical Algorithms

In the previous sections, we reformulated the MLFG in (1,2) as smooth Nash game $(\text{NEP}(\varepsilon))$ with a smoothing parameter $\varepsilon > 0$ and developed theory confirming the validity of this approach. In this section, a computational method is provided which is consistent to the developed theory.

In particular, we apply a gradient type method and recall corresponding convergence theory. As alternative we propose a Newton like method.

5.1 The Method

Due to Lemma 3.3, every leaders unique optimal solution is characterized by its KKT system 8 for a fixed smoothing parameter ε . The multistrategy vector of these solutions characterizes the unique Nash equilibrium of (NEP(ε)); therefore, we aim to find a primal dual pair $z = (x, \lambda)$ which satisfies the concatenated KKT conditions. We abbreviate the concatenation with $F^\varepsilon(z) = (F_1^\varepsilon(z), F_2^\varepsilon(z))^\top$ such that

$$F_1^\varepsilon(z) = Qx + c + \frac{1}{2}(L^\top + Q_y^{-1}B^\top)^\top a + \frac{1}{2} \sum_{i=1}^m a_i (L^\top - Q_y^{-1}B^\top)_{:,i}^\top \tilde{\phi}_\varepsilon' \left([(L^\top - Q_y^{-1}B^\top)x]_i \right) + \begin{bmatrix} \nabla_{x_1} g_1(x_1) \lambda_1 \\ \vdots \\ \nabla_{x_N} g_N(x_N) \lambda_N \end{bmatrix} \quad (18a)$$

$$F_2^\varepsilon(z) = \min \left\{ \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix}, \begin{bmatrix} -g_1(x_1) \\ \vdots \\ -g_N(x_N) \end{bmatrix} \right\} \quad (18b)$$

using the notation $Q = \text{diag}(Q_1, \dots, Q_N)$ and $c = (c_1^\top, \dots, c_N^\top)^\top$. The roots of this system characterize the Nash equilibrium for a fixed relaxation parameter ε . With this notation, the KKT system can be equivalently expressed as the minimization of the auxiliary function $\Psi_\varepsilon : \mathbb{R}^{n+\bar{m}} \rightarrow \mathbb{R}_+$ where $\bar{m} = m_1 + \dots + m_N$ and

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

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. *The function 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) $\frac{1}{2} \|F_2(z)\|_2^2 = \frac{1}{2} \sum_{i=1}^m \min^2\{\lambda_i, -g_i(x)\} = \frac{1}{8} \sum_{i=1}^m (\lambda_i - g_i(x) - |\lambda_i + g_i(x)|)^2$ 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. \square

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:

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

for a decreasing sequence of positive numbers $(\varepsilon_i)_{i \in \mathbb{N}}$. 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 $\text{NEP}(\varepsilon_i)$. 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 which implicitly characterize $\frac{\partial x}{\partial \varepsilon}$. For $\nu = 1, \dots, N$, we have

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

which leads to the following system

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

Here, we denote $\tilde{\phi}_\varepsilon(\eta) = \Phi(\eta, \varepsilon)$ 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 composed of the second derivatives of the strictly convex objectives. We summarize the general approach in the following algorithm.

Algorithm 1

- 1: **Initialize** Choose $z^0(\varepsilon_0) = (x^0(\varepsilon_0), \lambda^0(\varepsilon_0)) \in \mathbb{R}^{n+m}$, $tol > 0$, $\varepsilon_0 \in (1, 2)$, $\gamma \in (0, 1)$.
- 2: **for** $i = 0, 1, \dots$ **do**
- 3: Compute Nash equilibrium of $(\text{NEP}(\varepsilon_i))$ with initial guess $z^0(\varepsilon_i) = (x^0(\varepsilon_i), \lambda^0(\varepsilon_i))$ by

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

- 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, but other approaches are conceivable, e.g. diagonalization methods as in [13] or path following techniques [8, 10].

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 [6] for a fixed smoothing parameter $\varepsilon > 0$. Stationary points of Ψ_ε are computed as the limit of a sequence of h - δ -stationary points. Bagirov et. al. [6] 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ if and only if

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

Lemma 5.3 (Termination). (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, [6, Proposition 4.1] is applicable.

(2) The function Ψ_ε is bounded from below as it takes nonnegative values only, therefore [6, Proposition 5.1] is applicable. \square

Theorem 5.4 (Convergence). Assume $\mathcal{L}(z^0) = \{z \in \mathbb{R}^{n+\bar{m}} : \Psi_\varepsilon(z) \leq \Psi_\varepsilon(z^0)\}$ 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, [6, Proposition 5.2] is applicable. The boundedness of $\mathcal{L}(z^0)$ implies that it exists at least one accumulation point. \square

Bagirov et. al. [6] 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

1: **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 $\|d_0\| = 1$, $0 < c_2 \leq c_1 \leq 1$,
 $\varepsilon > 0$.

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

3: $\bar{z}_1 = z^k$,

4: **for** $j = 1, \dots$ **do** ▷ compute h - δ -stationary point

5: Compute quasisecant $v_0 = v(\bar{z}_j, d_0, h)$,

6: $\tilde{v}_0 = v_0$,

7: **for** $i = 0, 1, \dots$ **do** ▷ find decent direction

8: $c_i = \arg \min\{\|cv_i + (1 - c)\tilde{v}_i\|_2^2 \mid c \in (0, 1)\}$,

9: $\bar{v}_i = c_i v_i + (1 - c_i)\tilde{v}_i$,

10: **if** $\|\bar{v}_i\| \leq \delta_k$ **then** return $v^j = \bar{v}_i$.

11: $d_i = -\frac{\bar{v}_i}{\|\bar{v}_i\|}$,

12: **if** $\Psi_\varepsilon(\bar{z}_j + hd_i) - \Psi_\varepsilon(\bar{z}_j) \leq -c_1 h \|\bar{v}_i\|$ **then** return $v^j = \bar{v}_i$.

13: Compute quasisecant $v_{i+1} = v(x, d_i, h)$,

14: $\tilde{v}_{i+1} = \bar{v}_i$.

15: **end for**

16: **if** $\|v^j\| \leq \delta_k$ **then** Stop.

17: $d^j = -\frac{v^j}{\|v^j\|}$,

18: Compute step length such that

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

19: Update $\bar{z}_{j+1} = \bar{z}_j + \sigma_j d^j$.

20: **end for**

21: $z^{k+1} = \bar{z}_j$,

22: $h_{k+1} = \gamma h_k$,

23: $\delta_{k+1} = \gamma \delta_k$.

24: **end for**

5.3 Nonsmooth Newton Method

Next, we present an improved method. The joint KKT system (18) 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 linear systems

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

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.3.

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 Method

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

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

3: **if** $\Psi_\varepsilon \leq tol$ **then** Stop.

4: Let $H \in \partial F^\varepsilon(z^k)$,

5: **if** H singular **then** do subgradient decent of Ψ_ε , thus choose

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

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

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

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

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

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

8: **end for**

For further discussions and convergence analysis we refer to e.g. [23].

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.4 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 1, 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),

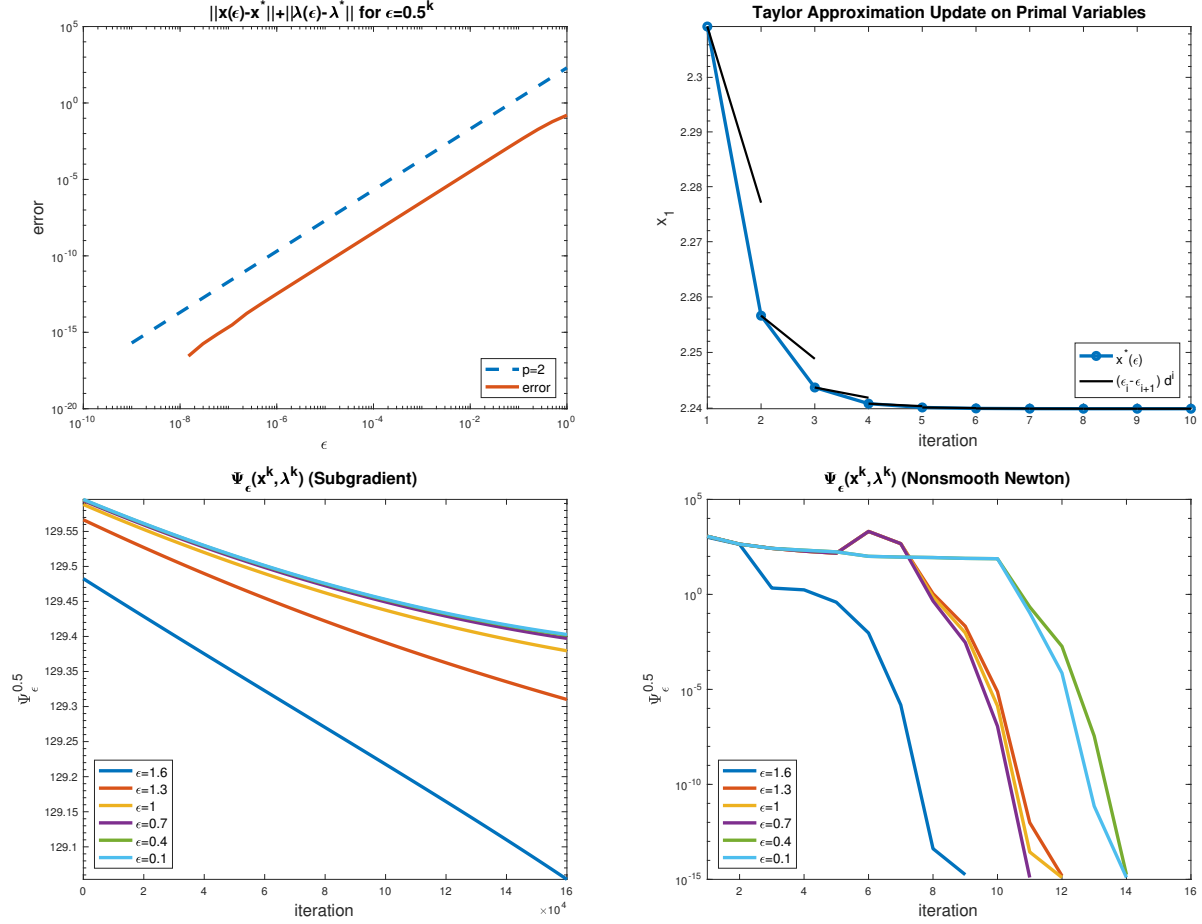


Figure 1: Upper left: Quadratic convergence to the limit Nash equilibrium $x^*(0)$, right: Taylor expansion based update on primal variables, exemplary for first leader variable; Lower: comparison of Subgradient and Nonsmooth Newton method for varying smoothing parameter.

$x^*(\epsilon_i)$. A black line represents the Taylor update and the lower end of a black line indicates the updated initial values $x^0(\epsilon_{i+1})$ for the subsequent Nash equilibrium computation.

The lower part of Figure 1 is dedicated to illustrate the importance of large smoothing parameter for the first computations of Nash equilibria. Since the problem gets closer to its 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 $\epsilon = 1.6$ and $\epsilon = 0.1$.

As already seen in Figure 1, the subgradient based method suffers from characteristically slow convergence for our instances. In Figure 2 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.

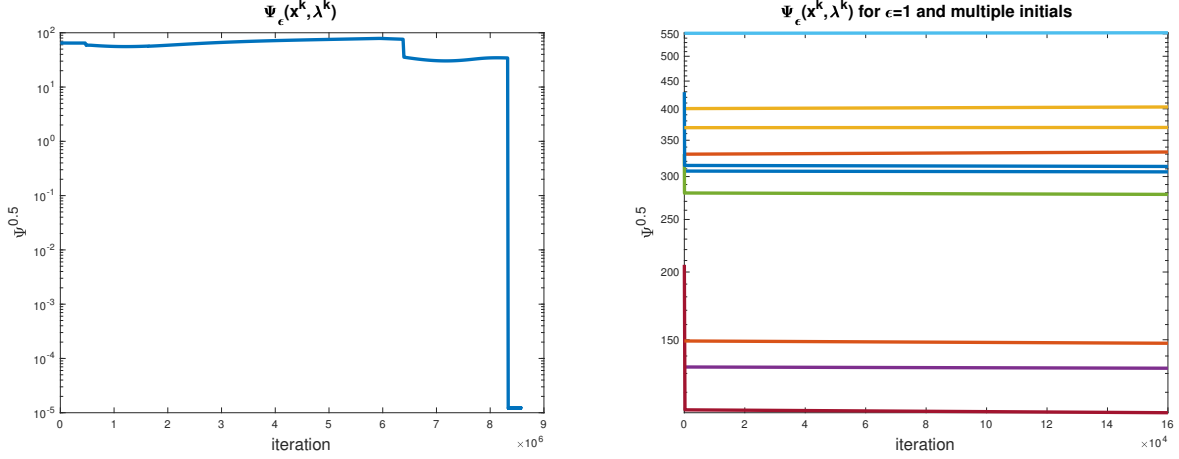


Figure 2: Subgradient Method; left all iterations for decreasing sequence of smoothing parameters, right for one smoothing parameter and multiple initials.

Similarly to Figure 2, 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.

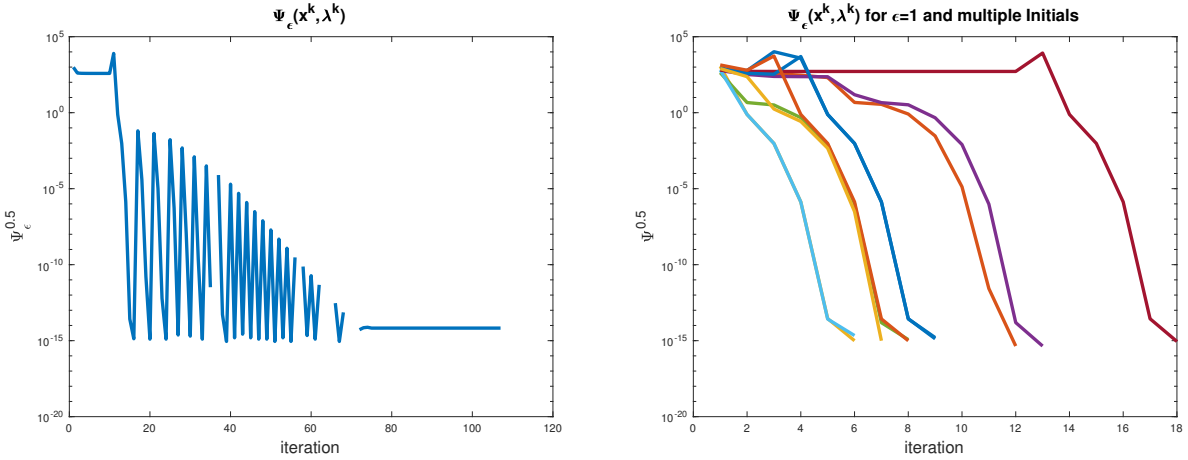


Figure 3: Nonsmooth Newton Method; left all iterations for decreasing sequence of smoothing parameters, right for one smoothing parameter and multiple initials.

7 Conclusion and Outlook

We presented a quadratic MLFG and explicitly computed the best response of the follower player. With this 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 de-

creasing positive smoothing parameter, we showed that the limit of Nash equilibria satisfies the conditions of s-stationarity. Further, we demonstrated that s-stationary points are eventually Nash equilibria of the MLFG. 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.

Also, the extension possibilities discussed in appendix A.1 are subject of further investigation.

A Appendix

A.1 Remarks on Extensions to the Model

In the following, we mention two generalizations to the follower problem in (1), which seem obvious to include. We explain challenges to motivate future research.

In (1), it is assumed that the matrix Q_y is positive definite and diagonal. A popular approach in quadratic programming is to enforce diagonality of symmetric positive definite matrices by introducing an auxiliary variable $z = D^\top y$, where we have the Cholesky decomposition $Q_y = DD^\top$, which exists for symmetric positive definite Q_y . The follower's problem can be equivalently formulated as a minimization problem in z :

$$\min_{z \in \mathbb{R}^m} \frac{1}{2} z^\top z - b(x)^\top D^{-\top} z \quad \text{s.t.} \quad z \geq D^\top l(x).$$

As in Lemma 2.2, we derive the solution to this optimization problem explicitly:

$$z(x) = \max\{D^{-1}b(x), D^\top l(x)\},$$

and we recover the follower's best response in this setting:

$$y(x) = D^{-\top} \max\{D^{-1}b(x), D^\top l(x)\}.$$

However, unlike (3), this function is not necessarily componentwise convex. But this property is essential to guarantee convexity of the leaders' objectives. This property is required for the proof of existence of Nash equilibria in Theorem 2.3 as it is based on Kakutani fixed-point theorem. An other obvious extension to the follower problem is to incorporate upper bounds besides the discussed lower bounds, i.e. $l(x) \leq y \leq u(x)$. We can also derive an explicit representation of the best response, e.g. via projection of the objective's gradient into the feasible set. For that let $\tilde{y}_i(x) = (Q_y)_{ii}^{-1} b_i(x)$, then the best response is for $i = 1, \dots, m$

$$y_i(x) = \text{median}(l_i(x), \tilde{y}_i(x), u_i(x)) = \begin{cases} l_i(x), & \tilde{y}_i(x) \leq l_i(x), \\ u_i(x), & \tilde{y}_i(x) \geq u_i(x), \\ \tilde{y}_i(x), & \text{else.} \end{cases}$$

Like in the previous case, this best response is not necessarily convex; therefore, the existence result in Theorem 2.3 does not apply.

A.2 Subgradient Method

In the following, we state the definition of quasisecants and a theorem which relates quasisecants and subgradients. Furthermore, we state an assumption which is needed in the convergence theorem of the subgradient method. All is adapted from [6] 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 $\text{conv}(x, x + hd)$ and $B_{O(h)}$ denotes a ball for which $O(h) \rightarrow 0$ for $h \rightarrow 0$.

For the convergence proof it is necessary to study the relation of $W_h(x)$ (Definition 5.2) 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 assumption 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 as $h \rightarrow 0$ of the $W_h(x)$ lies in the subdifferential, i.e.

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

A.3 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 the block $\boxed{A} \in \mathbb{R}^{n \times n}$ is

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

the block $\boxed{B} \in \mathbb{R}^{n \times \bar{m}}$ is

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

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

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

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

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

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:

$$\left(\boxed{D_\nu} \right)_i = \partial_{\lambda_i^C} \min \{ \lambda_\nu^i, -g_\nu^i(x_\nu) \} = \begin{cases} 1, & \lambda_\nu^i < -g_\nu^i(x_\nu), \\ 0, & \lambda_\nu^i > -g_\nu^i(x_\nu), \\ [1, 0], & \lambda_\nu^i = -g_\nu^i(x_\nu). \end{cases}$$

A.4 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.4.1 Data Set 1

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

$$Q_1 = \begin{bmatrix} 1.7 & 1.6 \\ 1.6 & 2.8 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 2.7 & 1.3 \\ 1.3 & 3.6 \end{bmatrix}, \quad c_1 = c_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad a = \begin{bmatrix} 1.4 \\ 2.6 \\ 2.1 \end{bmatrix}.$$

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

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

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.4.2 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

$$Q_1 = \begin{bmatrix} 2.5 & 1.6 \\ 1.6 & 3.8 \end{bmatrix}, \quad Q_2 = \begin{bmatrix} 2.9 & 1.3 \\ 1.3 & 1.8 \end{bmatrix}, \quad Q_3 = \begin{bmatrix} 3.2 & 2.3 \\ 2.3 & 2.6 \end{bmatrix}, \quad c_1 = c_2 = c_3 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad a = \begin{bmatrix} 0.4 \\ 1.6 \\ 2.6 \end{bmatrix}.$$

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

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

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

Special thanks to Michael Ferris and Olivier Huber for fruitful discussions on model extensions. This work was supported by the DFG under Grant 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 and A. Svensson. Some remarks about existence of equilibria, and the validity of the epcc reformulation for multi-leader-follower games. *Journal of nonlinear and convex analysis*, 19(7):1141–1162, 2018.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] A. M. Bagirov, L. Jin, N. Karmita, 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.
- [7] V. DeMiguel and H. Xu. A Stochastic Multiple-Leader Stackelberg Model: Analysis, Computation, and Application. *Operations Research*, 57(5):1220–1235, oct 2009. doi: 10.1287/opre.1080.0686.
- [8] S. P. Dirkse and M. C. Ferris. The path solver: a nonmonotone stabilization scheme for mixed complementarity problems. *Optimization Methods and Software*, 5(2):123–156, jan 1995. doi: 10.1080/10556789508805606.
- [9] 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.
- [10] M. C. Ferris and T. S. Munson. Interfaces to PATH 3.0: Design, Implementation and Usage. In *Computational Optimization*, pages 207–227. Springer US, 1999.

- [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] R. Janin. *Directional derivative of the marginal function in nonlinear programming*, pages 110–126. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984. ISBN 978-3-642-00913-6. doi: 10.1007/BFb0121214. URL <https://doi.org/10.1007/BFb0121214>.
- [15] Y. Kim and M. C. Ferris. Solving equilibrium problems using extended mathematical programming. *Mathematical Programming Computation*, 11(3):457–501, mar 2019. doi: 10.1007/s12532-019-00156-4.
- [16] A. Koh and S. Shepherd. Tolling, collusion and equilibrium problems with equilibrium constraints. 2010. ISSN 1825-3997.
- [17] A. A. Kulkarni and U. V. Shanbhag. A Shared-Constraint Approach to Multi-Leader Multi-Follower Games. *Set-Valued and Variational Analysis*, 22(4):691–720, aug 2014. doi: 10.1007/s11228-014-0292-5.
- [18] A. A. Kulkarni and U. V. Shanbhag. An Existence Result for Hierarchical Stackelberg v/s Stackelberg Games. *IEEE Transactions on Automatic Control*, 60(12):3379–3384, dec 2015. doi: 10.1109/tac.2015.2423891.
- [19] S. Leyffer and T. Munson. Solving multi-leader-common-follower games. *Optimization Methods and Software*, 25(4):601–623, 2010.
- [20] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124 – 143, 1996. ISSN 0899-8256.
- [21] H. Nikaidô and K. Isoda. Note on non-cooperative convex game. *Pacific Journal of Mathematics*, 5(5):807–815, 1955.
- [22] 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.
- [23] L. Qi and J. Sun. A nonsmooth version of newton’s method. *Mathematical Programming*, 58(1):353–367, Jan 1993. ISSN 1436-4646.

- [24] 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.
- [25] H. D. Sherali. A multiple leader stackelberg model and analysis. *Operations Research*, 32(2): 390–404, 1984.
- [26] A. T. Sonja Steffensen. An explicit nash equilibrium to a multi-leader-follower game. In *PAMM-Proc. Appl. Math. Mech.* Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2019.
- [27] S. Steffensen and M. Ulbrich. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010.
- [28] C.-L. Su. A snpc method for solving equilibrium problems with equilibrium constraints. *Computing in Economics and Finance* 150, 2005.
- [29] C.-L. Su. Analysis on the forward market equilibrium model. *Operations Research Letters*, 35 (1):74–82, jan 2007. doi: 10.1016/j.orl.2006.01.006.
- [30] 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.