SC^1 OPTIMIZATION REFORMULATIONS OF THE GENERALIZED NASH EQUILIBRIUM PROBLEM

Anna von Heusinger and Christian Kanzow

Preprint 278 May 2007

University of Würzburg
Institute of Mathematics
Am Hubland
97074 Würzburg
Germany

e-mail: heusinger@mathematik.uni-wuerzburg.de
kanzow@mathematik.uni-wuerzburg.de

May 15, 2007

Abstract. The generalized Nash equilibrium problem is a Nash game which, in contrast to the standard Nash equilibrium problem, allows the strategy sets of each player to depend on the decision variables of all other players. It was recently shown by the authors that this generalized Nash equilibrium problem can be reformulated as both an unconstrained and a constrained optimization problem with continuously differentiable objective functions. This paper further investigates these approaches and shows, in particular, that the objective functions are SC^1-functions. Moreover, conditions for the local superlinear convergence of a semismooth Newton method being applied to the unconstrained optimization reformulation are also given. Some numerical results indicate that this method works quite well on a number of problems coming from different application areas.

Key Words: Generalized Nash equilibrium, normalized Nash equilibrium, joint constraints, regularized Nikaido-Isoda function, implicit function, semismooth function, superlinear convergence.

1 Introduction

In the recent paper [17], the authors introduced reformulations of the generalized Nash equilibrium problem (GNEP for short) as continuously differentiable optimization problems. However, these optimization reformulations are, in general, not twice differentiable. This paper investigates some further properties of these reformulations and, in particular, shows that they are sufficiently smooth so that locally superlinearly convergent Newton-type methods can be applied in order to solve the underlying GNEP.

To be more specific, let us first recall the definition of a standard Nash equilibrium problem (NEP for short): There are N players, each player $\nu \in \{1, \ldots, N\}$ controls the variables $x^\nu \in \mathbb{R}^{n_\nu}$. Let $x = (x^1, \ldots, x^N)^T \in \mathbb{R}^n$ be the vector comprised by all these decision variables, where $n := n_1 + \ldots + n_N$. To emphasize the νth player's variables within the vector x, we sometimes write $x = (x^\nu, x^{-\nu})^T$, where $x^{-\nu}$ subsumes all the other players' variables.

Let $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ be the νth player's payoff (or loss or utility) function, and let $X_\nu \subseteq \mathbb{R}^{n_\nu}$ be the strategy set of player ν. Then $x^* = (x^{*,1}, \ldots, x^{*,N})^T \in \mathbb{R}^n$ is called a *Nash equilibrium* or a *solution of the NEP* if each block component $x^{*,\nu}$ is a solution of the optimization problem

$$\min_{x^\nu} \ \theta_\nu(x^\nu, x^{*,-\nu}) \quad \text{s.t.} \quad x^\nu \in X_\nu,$$

i.e., x^* is a Nash equilibrium if no player can improve his situation by unilaterally changing his strategy.

The characteristic feature of a standard NEP is therefore that the utility function θ_ν of player ν depends both on the decision variables x^ν of this player and on the decision variables $x^{-\nu}$ of all other players, whereas the strategy sets X_ν depend on x^ν only. In fact, this is the main difference to the GNEP, where the strategy set is allowed to depend on the choice of the other players, too. More precisely, the GNEP is defined by the utility functions $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ of player ν and a common strategy set $X \subseteq \mathbb{R}^n$. A vector $x^* = (x^{*,1}, \ldots, x^{*,N})^T \in \mathbb{R}^n$ is then called a *generalized Nash equilibrium* or a *solution of the GNEP* if each block component $x^{*,\nu}$ is a solution of the optimization problem

$$\min_{x^\nu} \ \theta_\nu(x^\nu, x^{*,-\nu}) \quad \text{s.t.} \quad (x^\nu, x^{*,-\nu}) \in X.$$

Throughout this paper, we assume that the set X can be represented as

$$X = \{x \in \mathbb{R}^n \mid g(x) \leq 0\} \tag{1}$$

for some function $g : \mathbb{R}^n \to \mathbb{R}^m$. Additional equality constraints are also allowed, but for notational simplicity, we prefer not to include them explicitly. In many cases, a player ν might have some additional constraints of the form $h^\nu(x^\nu) \leq 0$ depending on his decision variables only. However, these additional constraints may simply be viewed as part of the joint constraints $g(x) \leq 0$, with some of the component functions g_i of g depending on the block component x^ν of x only.

Throughout this paper, we make the following blanket assumptions.

Assumption 1.1 *(a) The utility functions θ_ν are twice continuously differentiable and, as a function of x^ν alone, convex.*

(b) The function g is twice continuously differentiable, its components g_i are convex (in x), and the corresponding strategy space X defined by (1) *is nonempty.*

The convexity assumptions are absolutely standard when considering (generalized) Nash equilibrium problems, and the smoothness assumptions are also very natural since our aim is to develop locally fast convergent methods for the solution of GNEPs. Note that Assumption 1.1 (b) implies that the strategy space X is nonempty, closed, and convex. In contrast to some other papers on GNEPs, we do not assume that X is compact.

The previous description of a GNEP follows the one from the seminal paper by Rosen [35], who also suggested a projected gradient-type method for the solution of GNEPs. Motivated by the fact that a standard NEP can be reformulated as a variational inequality problem (VIP for short), see, for example, [9], both Bensousson [2] and Harker [16] characterize the GNEP as a quasi-variational inequality (abbreviated as QVI in the following). However, since there are essentially no efficient methods for solving QVIs, such a characterization is less interesting from a practical point of view. On the other hand, it was noted in [13, 7], for example, that certain solutions of the GNEP (the normalized Nash equilibria, to be defined later) can be found by solving a suitable standard VIP associated to the GNEP. A discussion of some local issues related to this formulation is given in [8] (in fact, this paper also considers a more general class of GNEPs). A globally convergent augmented Lagrangian-type VIP method is presented in [29].

Another approach for the solution of GNEPs is based on the so-called Nikaido-Isoda-function (see [27]), cf. Section 3 for a formal definition. Relaxation methods using this Nikaido-Isoda-function are investigated in [38, 20] (see also [1, 23]) for some similar ideas), and a proximal-like method on the basis of the Nikaido-Isoda-function is presented in [13]. A regularized version of the Nikaido-Isoda-function, first introduced in [15] for standard NEPs (see also [25, 39] for related approaches in the context of equilibrium programming problems) and then further investigated in [17] for GNEPs, is the basis of suitable reformulations of the GNEP as both an unconstrained and a constrained optimization reformulation of the GNEP in [17].

Here we further investigate the properties of these optimization reformulations. We believe that these reformulations, together with the VIP-based methods, are currently the most promising approaches for the solution of GNEPs. A discussion including the advantages and disadvantages of these two different techniques will be given at the end of this paper, since this requires a better knowledge of certain properties, with some of them to be shown in this paper.

The organization of the paper is as follows: We first recall some basic facts and recent results from nonsmooth analysis in Section 2. We then give the details of our optimization reformulations of the GNEP in Section 3 and show that these are SC^1 reformulations, i.e., the objective functions are continuously differentiable with a semismooth gradient. We then show in Section 4 that this fact can be used in order to get locally superlinearly

convergent Newton-type methods for the solution of GNEPs. Some numerical results are then presented in Section 5, and we close with some final remarks in Section 6.

Notation: Given a differentiable function $g : \mathbb{R}^n \to \mathbb{R}^m$, $g'(x) \in \mathbb{R}^{m\times n}$ or $Dg(x)$ denotes the Jacobian of g at x, whereas $\nabla g(x) \in \mathbb{R}^{n\times m}$ is the transposed Jacobian. In particular, for $m = 1$, the gradient $\nabla g(x)$ is viewed as a column vector. Several times, we also consider partial derivatives of a real-valued function $f : \mathbb{R}^n \to \mathbb{R}$ with respect to certain block components of x only, and this will be denoted by using suitable subscripts, e.g., $\nabla_{x^\nu} f(x)$ denotes the partial gradient of f at x, where the derivatives are taken with respect to the components of the block vector x^ν only. Second-order partial derivatives with respect to certain block components are written in a similar way as $\nabla^2_{x^\nu x^\mu} f(x)$, for example, meaning that we first differentiate with respect to x^ν and then with respect to x^μ. For a matrix $A \in \mathbb{R}^{m\times n}$ and a subset $I \subseteq \{1, \ldots, n\}$ we denote by A_I the submatrix of A consisting of the columns a_i, $i \in I$.

2 Preliminaries

In this section, we first recall some basic definitions and results from nonsmooth analysis, and then state some preliminary results that will be used in our subsequent analysis. To this end, let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitzian mapping. According to Rademacher's theorem (see [33]), it follows that F is almost everywhere differentiable. Let D_F denote the set of all differentiable points of F. Then we call

$$\partial_B F(x) := \left\{ H \in \mathbb{R}^{m\times n} \,\middle|\, \exists \{x^k\} \subseteq D_F : x^k \to x, F'(x^k) \to H \right\}$$

the *B-subdifferential* of F at x. Its convex hull

$$\partial F(x) := \text{conv}\partial_B F(x)$$

is Clarke's *generalized Jacobian* of F at x, see [3]. In case of $m = 1$, we call this set also the *generalized gradient* of F at x which, therefore, is a set of row vectors. Furthermore, we call the set

$$\partial_C F(x) := \left(\partial F_1(x)^T \times \ldots \times \partial F_m(x)^T\right)^T$$

the *C-subdifferential* of F at x, i.e., the C-subdifferential is the set of matrices whose ith rows consist of the elements of the generalized gradient of the ith component functions F_i. According to [3, Proposition 2.6.2], the following inclusions hold:

$$\partial_B F(x) \subseteq \partial F(x) \subseteq \partial_C F(x). \tag{2}$$

Based on the generalized Jacobian, we next recall the definition of a semismooth function.

Definition 2.1 *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz continuous. Then F is called* semismooth *at x if F is directionally differentiable at x and*

$$\|Hd - F'(x; d)\| = o(\|d\|)$$

holds for all $d \to 0$ *and all* $H \in \partial F(x+d)$.

In the following, we often call a mapping $F : \mathbb{R}^n \to \mathbb{R}^m$ semismooth if it is semismooth at every point $x \in \mathbb{R}^n$. The notion of a semismooth function was originally introduced by Mifflin [26] for functionals, and later extended by Qi and Sun [32] to vector-valued mappings.

Note that there are many different notions of semismooth functions available in the literature, and we would like to give some comments here. First of all, our definition of a semismooth function is not the original one from [32], however, it follows from [32, Theorem 2.3] that it is equivalent to the original definition (note that the assumption of directional differentiability is missing in that result). Another very popular reformulation of the semismoothness of a locally Lipschitz and directionally differentiable function is that it satisfies

$$\|F(x+d) - F(x) - Hd\| = o(\|d\|) \tag{3}$$

for all $d \to 0$ and all $H \in \partial F(x+d)$. Sun [37] calls this the *superlinear approximation property* of F at x since it is central in order to prove local superlinear convergence of certain Newton-type methods, see also the general scheme in Kummer [21, 22]. The equivalence of this superlinear approximation property to our definition of semismoothness can be found, e.g., in [10, Theorem 7.4.3] and is based on the fact that a locally Lipschitz and directionally differentiable function is automatically B-differentiable, see [36] for details. On the other hand, property (3) can be defined also for mappings that are not necessarily directionally differentiable. In fact, Gowda [14] takes this property of a locally Lipschitz function as the definition of semismoothness. In order to avoid confusion with the existing definition of semismoothness, Pang et al. [30] suggested the name *G-semismoothness* (with the 'G' referring to Gowda).

We stress that the previous discussion on semismoothness is somewhat crucial for our later analysis since we want to apply a suitable implicit function theorem for semismooth functions. However, there are different implicit function theorems available in the literature, and they are based on different notions of a semismooth (or related) function, see, [37, 14] and, in particular, the corresponding discussion in [30].

We next state a simple result that will play an important role in later sections, in particular, the equivalence between statements (a) and (d).

Lemma 2.2 *Let* $F : \mathbb{R}^n \to \mathbb{R}^m$ *be locally Lipschitz continuous and directionally differentiable, and let* $x \in \mathbb{R}^n$ *be an arbitrary point. Then the following statements are equivalent:*

(a) F *is semismooth at* x*, i.e.,* $\|Hd - F'(x;d)\| = o(\|d\|)$ *for all* $d \to 0$ *and all* $H \in \partial F(x+d)$.

(b) $\|Hd - F'(x;d)\| = o(\|d\|)$ *for all* $d \to 0$ *and all* $H \in \partial_B F(x+d)$.

(c) $\|Hd - F'(x;d)\| = o(\|d\|)$ *for all* $d \to 0$ *and all* $H \in \partial_C F(x+d)$.

(d) F_i *is semismooth for all components* $i = 1, \ldots, m$*, i.e.,* $\|h_i d - F_i'(x;d)\| = o(\|d\|)$ *for all* $d \to 0$*, all* $h_i \in \partial F_i(x+d)$*, and all* $i = 1, \ldots, m$.

Proof. The implications $(c) \Longrightarrow (a) \Longrightarrow (b)$ follow directly from the fact that $\partial_B F(x+d) \subseteq \partial F(x+d) \subseteq \partial_C F(x+d)$, cf. (2).

The implication $(b) \Longrightarrow (a)$ is a consequence of Carathéodory's theorem. To see this, let $d^k \to 0$ and $H^k \in \partial F(x+d^k)$ be given arbitrarily. Then, for all $k \in \mathbb{N}$, we can find at most $r := nm+1$ matrices $H_j^k \in \partial_B F(x+d^k)$ and numbers $\lambda_j^k \geq 0$ satisfying

$$\sum_{j=1}^{r} \lambda_j^k = 1 \quad \text{and} \quad H^k = \sum_{j=1}^{r} \lambda_j^k H_j^k.$$

Using (b), we therefore obtain

$$\begin{aligned} \|H^k d^k - F'(x;d^k)\| &= \Big\| \sum_{j=1}^{r} \lambda_j^k H_j^k d^k - F'(x;d^k) \Big\| \\ &\leq \sum_{j=1}^{r} \lambda_j^k \|H_j^k d^k - F'(x;d^k)\| = o(\|d^k\|) \end{aligned}$$

in view of the boundedness of λ_j^k.

The implication $(a) \Longrightarrow (d)$ can be verified in the following way: Using the chain rule from [3, Theorem 2.6.6], the composite mapping $f := g \circ F$ with the continuously differentiable function $g(z) := z_i$ has the generalized gradient

$$\begin{aligned} \partial F_i(x) &= \partial f(x) = \partial g\big(F(x)\big)\partial F(x) = e_i^T \partial F(x) \\ &= \{h_i \mid h_i \text{ is the } i\text{th row of some } H \in \partial F(x)\}. \end{aligned}$$

Therefore, if we assume that (a) holds, and if we take an arbitrary $d \in \mathbb{R}^n$ as well as any component $i \in \{1,\ldots,m\}$, it follows that for any $h_i \in \partial F_i(x+d)$, we can choose an element $H \in \partial F(x+d)$ such that its ith row is equal to h_i. Then we get

$$\big|F_i'(x;d) - h_i d\big| = \big|e_i^T (F'(x;d) - Hd)\big| \leq \|F'(x;d) - Hd\| = o(\|d\|),$$

hence F_i is semismooth at x.

Finally, (d) $\Longrightarrow$ (c) is an immediate consequence of the definition of the C-subdifferential.

Altogether, we have shown that $(c) \Longrightarrow (a) \Longrightarrow (d) \Longrightarrow (c)$ and $(a) \Longleftrightarrow (b)$, implying that all four statements are indeed equivalent. □

Some parts of the previous result are known, for example, [32, Corollay 2.4] shows that the semismoothness of all component functions implies the semismoothness of F itself. The fact that the converse also holds seems to be around in the community, but we were not able to find an explicit reference. Furthermore, [14, page 447] already observed the equivalence of statements (a) and (b) in Lemma 2.2, albeit in the slightly different context of G-semismoothness.

We next want to state an implicit function theorem for semismooth mappings that will be used in order to show local fast convergence of our Newton-type method for generalized

Nash equilibrium problems. To this end, consider a mapping $H : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, $(x, y) \mapsto H(x, y)$. Then $\pi_y \partial H(x, y)$ denotes the set of all $n \times n$ matrices M such that, for some $n \times m$ matrix N, the $n \times (m + n)$ matrix $[N, M]$ belongs to $\partial H(x, y)$. The set $\pi_x \partial H(x, y)$ is defined in a similar way.

Theorem 2.3 *Suppose that $H : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz and semismooth in a neighbourhood of a point $(\bar{x}, \bar{y})$ satisfying $H(\bar{x}, \bar{y}) = 0$, and assume that all matrices in $\pi_y \partial H(\bar{x}, \bar{y})$ are nonsingular. Then there exists an open neighborhood X of $\bar{x}$ and a function $g : X \to \mathbb{R}^n$ which is Lipschitz and semismooth on X such that $g(\bar{x}) = \bar{y}$ and $H(x, g(x)) = 0$ for all $x \in X$.*

Proof. Since this particular implicit function theorem does not seem to be available in the literature, we derive it from a suitable inverse function theorem. To this end, consider the mapping $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$ defined by

$$F(x, y) := \begin{pmatrix} x - \bar{x} \\ H(x, y) \end{pmatrix}.$$

Then

$$\partial F(\bar{x}, \bar{y}) \subseteq \begin{pmatrix} I_m & 0 \\ \pi_x \partial H(\bar{x}, \bar{y}) & \pi_y \partial H(\bar{x}, \bar{y}) \end{pmatrix},$$

and our assumptions imply that all elements from the generalized Jacobian $\partial F(\bar{x}, \bar{y})$ are nonsingular. Noting that a continuously differentiable function is always semismooth and recalling that the mapping H is semismooth by assumption, it follows from Lemma 2.2 that F is also semismooth. Hence we can apply the inverse function theorem from [30, Theorem 6] and obtain open neighbourhoods U of $(\bar{x}, \bar{y})$ and W of $(0, 0) = F(\bar{x}, \bar{y})$ such that $F : U \to W$ is a homeomorphism and has a locally Lipschitz and semismooth inverse $G : W \to U$. Since W is open, the set

$$X := \{x \in \mathbb{R}^m \mid (x - \bar{x}, 0) \in W\}$$

is also open as a subset of $\mathbb{R}^m$. We now show that there is a locally Lipschitz and semismooth function $g : X \to \mathbb{R}^n$ such that $g(\bar{x}) = \bar{y}$ and $H(x, g(x)) = 0$ for all $x \in X$.

To this end, let $x \in X$ be arbitrarily given. Then $(x - \bar{x}, 0) \in W$, and because $F : U \to W$ is a homeomorphism, the definition of the mapping F implies that there is a unique vector y such that $(x, y) \in U$ and $F(x, y) = (x - \bar{x}, 0)$. Consequently, we have $H(x, y) = 0$. Note that this unique vector y depends on x. Setting $g(x) := y$ then gives us a mapping $g : X \to \mathbb{R}^n$ such that $H(x, g(x)) = 0$ for each $x \in X$. This implies

$$F(x, g(x)) = \begin{pmatrix} x - \bar{x} \\ H(x, g(x)) \end{pmatrix} = \begin{pmatrix} x - \bar{x} \\ 0 \end{pmatrix} \quad \forall x \in X.$$

Applying the inverse mapping G on both sides gives

$$\begin{pmatrix} x \\ g(x) \end{pmatrix} = G(x - \bar{x}, 0) \quad \forall x \in X.$$

This shows that g coincides with certain component functions of G. Since the inverse function G is semismooth, it therefore follows from Lemma 2.2 that g is also semismooth. This completes the proof of our implicit function theorem. □

A related implicit function theorem was stated in Sun [37]. However, he only assumes that H has the local superlinear approximation property, and states that the implicit function has the superlinear approximation property, too. A similar result was also stated by Gowda [14] in the framework of H-differentiable functions. Note also that the assumption on the nonsingularity of all elements from $\pi_y \partial H(\bar{x}, \bar{y})$ (corresponding to the strongest possible condition in the inverse function theorem from [30]) can be weakened, but that this (relatively strong) condition will be satisfied in our context.

We close this section with the definition of an SC^1-function that will become important in the next section.

Definition 2.4 *A mapping $f : \mathbb{R}^n \to \mathbb{R}$ is called an* SC^1-function *if it is continuously differentiable and its gradient ∇f is semismooth.*

3 SC^1-Optimization Reformulations

Consider the GNEP from Section 1 with utility functions θ_ν and a joint strategy set X satisfying the requirements from Assumption 1.1. Our aim is to show that the GNEP can then be reformulated as an SC^1 optimization problem (both constrained and unconstrained). This SC^1-reformulation is based on the approach given by the authors in [17]. To this end, we first recall the basic ingredients from that approach, and then present our new results.

A basic tool for both the theoretical analysis and the numerical solution of (generalized) Nash equilibrium problems is the corresponding *Nikaido-Isoda-function* (sometimes also called the *Ky-Fan-function*)

$$\Psi(x, y) := \sum_{\nu=1}^{N} \big[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})\big],$$

see, e.g., [27, 12, 13]. We first use this function in order to define an important subclass of all the solutions of a GNEP.

Definition 3.1 *A vector $x^* \in X$ is called a* normalized Nash equilibrium *of the GNEP if $\sup_{y \in X} \Psi(x^*, y) = 0$ holds, where Ψ denotes the Nikaido-Isoda-function.*

The above definition of a normalized Nash equilibrium corresponds to one given in, e.g., [13, 38]. Note that it is slightly different from the original definition of a normalized equilibrium given in [35], see, however, the corresponding results in [13, 7]. It is not difficult to see that a normalized Nash equilibrium is always a solution of the GNEP, whereas the converse is not true in general. In fact, simple examples show that a GNEP might have

infinitely many solutions, but just a single normalized Nash equilibrium, see, e.g., [7]. On the other hand, for a standard NEP, there is no difference between Nash equilibria and normalized Nash equilibria.

By adding a suitable regularization term depending on a parameter $\gamma > 0$, we obtain the *regularized Nikaido-Isoda-function*

$$\begin{aligned} \Psi_\gamma(x,y) &:= \sum_{\nu=1}^{N} \big[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu}) - \frac{\gamma}{2}\|x^\nu - y^\nu\|^2\big] \\ &= \sum_{\nu=1}^{N} \big[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu})\big] - \frac{\gamma}{2}\|x-y\|^2 \\ &= \Psi(x,y) - \frac{\gamma}{2}\|x-y\|^2 \end{aligned} \tag{4}$$

that was first introduced in [15] in the context of standard NEPs, and then further exploited by the authors [17] in the context of GNEPs, see also the discussion in [25] for the class of equilibrium programming problems. We now define the corresponding value function by

$$V_\gamma(x) := \max_{y \in X} \Psi_\gamma(x,y) = \Psi_\gamma\big(x, y_\gamma(x)\big), \tag{5}$$

where $y_\gamma(x)$ denotes the unique solution of the uniformly concave maximization problem

$$\max_y \ \Psi_\gamma(x,y) \quad \text{s.t.} \quad y \in X. \tag{6}$$

As noted in [17], the function V_γ is continuously differentiable with gradient given by

$$\nabla V_\gamma(x) = \nabla_x \Psi_\gamma(x,y)\big|_{y=y_\gamma(x)}. \tag{7}$$

Moreover, it was noted that x^* is a normalized Nash equilibrium of the GNEP if and only if it solves the constrained optimization problem

$$\min \ V_\gamma(x) \quad \text{s.t.} \quad x \in X \tag{8}$$

with optimal function value $V_\gamma(x^*) = 0$. Taking two parameters $0 < \alpha < \beta$ and denoting by V_α, V_β the corresponding value functions, it was also shown in [17] that x^* is a normalized Nash equilibrium of the GNEP if and only if x^* solves the unconstrained optimization problem

$$\min \ V_{\alpha\beta}(x), \quad x \in \mathbb{R}^n, \tag{9}$$

with optimal function value $V_{\alpha\beta}(x^*) = 0$, where

$$V_{\alpha\beta}(x) := V_\alpha(x) - V_\beta(x) \tag{10}$$

denotes the difference of two regularized Nikaido-Isoda functions.

Unfortunately, V_γ is, in general, not twice continuously differentiable, hence neither the constrained optimization problem (8) nor the unconstrained optimization problem (9)

are twice continuously differentiable. However, in view of Assumption 1.1, we see that the regularized Nikaido-Isoda-function $\Psi_\gamma(x,y)$ is twice continuously differentiable. Using the fact that the composition of semismooth functions is again semismooth, see [11], it therefore follows immediately from the representation (7) of the gradient ∇V_γ that V_γ (hence also $V_{\alpha\beta}$) is an SC^1-function if the mapping $x \mapsto y_\gamma(x)$ is semismooth. Our aim in this section is therefore to prove the semismoothness of this mapping.

To this end, we first consider a more general (parameterized) optimization problem of the form

$$\min_y \; f(x,y) \quad \text{s.t.} \quad y \in X \tag{11}$$

where $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and uniformly convex with respect to the variable y (for every fixed x). The feasible set X is given by a number of inequalities as in (1) such that Assumption 1.1 (b) holds. Then the Lagrangian of the optimization problem (11) is given by

$$L(x,y,\lambda) = f(x,y) + \sum_{i=1}^{m} \lambda_i g_i(y),$$

where, again, $x \in \mathbb{R}^n$ is supposed to be fixed. Let $y = y(x)$ be the unique solution of the optimization problem (11). Then, under a suitable constraint qualification (like the Slater condition), it follows that there exists a Lagrange multiplier $\lambda = \lambda(x) \in \mathbb{R}^m$ such that (y,λ) (together with the fixed x) solves the KKT system

$$\nabla_y L(x,y,\lambda) = \nabla_y f(x,y) + \nabla g(y)\lambda = 0, \quad 0 \le \lambda \perp -g(y) \ge 0. \tag{12}$$

Using the minimum function $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \varphi(a,b) := \min\{a,b\}$, we can reformulate the KKT system (12) as a system of nonlinear equations $\Phi(x,y,\lambda) = 0$ via the function

$$\Phi(x,y,\lambda) := \begin{pmatrix} \nabla_y L(x,y,\lambda) \\ \phi\big(-g(y),\lambda\big) \end{pmatrix} \tag{13}$$

with

$$\phi\big(-g(y),\lambda\big) := \big(\varphi(-g_1(y),\lambda_1),\ldots,\varphi(-g_m(y),\lambda_m)\big)^T \in \mathbb{R}^m.$$

Our first result gives a representation of the B-subdifferential and the generalized Jacobian of the mapping Φ.

Lemma 3.2 *Suppose that f and g are C^2-functions. Let $w = (x,y,\lambda) \in \mathbb{R}^{n+n+m}$. Then, each element $H \in \partial\Phi(w)^T$ can be represented as follows:*

$$H = \begin{pmatrix} \nabla^2_{yx} L(x,y,\lambda)^T & 0 \\ \nabla^2_{yy} L(x,y,\lambda) & -\nabla g(y) D_a(y,\lambda) \\ \nabla g(y)^T & D_b(y,\lambda) \end{pmatrix}$$

where $D_a(y,\lambda) := \operatorname{diag}(a_1(y,\lambda),\ldots,a_m(y,\lambda))$, $D_b(y,\lambda) = \operatorname{diag}(b_1(y,\lambda),\ldots,b_m(y,\lambda)) \in \mathbb{R}^{m\times m}$ *are diagonal matrices whose ith diagonal elements are given by*

$$a_i(y,\lambda) = \begin{cases} 1, & \text{if } -g_i(y) < \lambda_i, \\ 0, & \text{if } -g_i(y) > \lambda_i, \\ \mu_i, & \text{if } -g_i(y) = \lambda_i, \end{cases} \quad \text{and} \quad b_i(y,\lambda) = \begin{cases} 0, & \text{if } -g_i(y) < \lambda_i, \\ 1, & \text{if } -g_i(y) > \lambda_i, \\ 1-\mu_i, & \text{if } -g_i(y) = \lambda_i, \end{cases}$$

for any $\mu_i \in [0,1]$. *The elements* $H \in \partial_B\Phi(w)^T$ *are obtained by choosing* $\mu_i \in \{0,1\}$.

Proof. The first n components of the vector function Φ are continuously differentiable and Φ is continuously differentiable with respect to x, so the expression for the first n rows and columns of H readily follows. To compute the remaining entries of H, we use the fact that each element of the generalized Jacobian of ϕ can be represented by an element of the C-subdifferential of ϕ, that is

$$\partial\phi(-g(y),\lambda)^T \subseteq \partial\varphi(-g_1(y),\lambda_1)^T \times \cdots \times \partial\varphi(-g_m(y),\lambda_m)^T.$$

If i is such that $-g_i(y) \neq \lambda_i$, then φ is continuously differentiable at $(-g_i(y),\lambda_i)$ and the expression for the $(n+i)$th column of H follows. If instead $-g_i(y) = \lambda_i$, then, using the definition of the B-subdifferential, it follows that

$$\partial_B\varphi(-g_i(y),\lambda_i)^T = \big\{(-\nabla g_i(y)^T, 0), (0, e_i^T)\big\}.$$

Taking the convex hull, we therefore get

$$\partial\varphi(-g_i(y),\lambda_i)^T = \{(-\mu_i\nabla g_i(y)^T, (1-\mu_i)e_i^T) \mid \mu_i \in [0,1]\}.$$

(Note that this representation cannot be obtained by directly applying [3, Theorem 2.3.9 (iii)] since the min-function is not regular in the sense of [3, Definition 2.3.4].) This gives the representation of $H \in \partial\Phi(w)^T$. □

Our next aim is to establish conditions for the nonsingularity of all elements in $\pi_{(y,\lambda)}\partial\Phi(w)^T$ at a point $w = (x,y,\lambda)$ satisfying $\Phi(w) = 0$. By definition, taking the continuous differentiability of Φ with respect to x into account, the elements $V \in \pi_{(y,\lambda)}\partial\Phi(w)^T$ can be obtained by deleting the first n rows of the matrices H from Lemma 3.2. In order to get a more detailed description of the matrices $V \in \pi_{(y,\lambda)}\partial\Phi(w)^T$, let us partition the index set $\{1,\ldots,m\}$ into

$$I_0 := \{i \mid g_i(y) = 0\} \quad \text{and} \quad I_< := \{i \mid g_i(y) < 0\},$$

where both the set of active constraints I_0 and the set of inactive constraints $I_<$ depend on the current vector y. The set of active constraints can be further divided into

$$I_{00} := \{i \in I_0 \mid \lambda_i = 0\} \quad \text{and} \quad I_+ := \{i \in I_0 \mid \lambda_i > 0\},$$

with both sets depending on y and λ. The set I_{00} will further be partitioned into

$$I_{01} := \{i \in I_{00} \mid \mu_i = 1\}, \quad I_{02} := \{i \in I_{00} \mid \mu_i \in (0,1)\}, \quad I_{03} := \{i \in I_{00} \mid \mu_i = 0\}.$$

Note that these index sets also depend (via μ_i) on the particular element taken from the generalized Jacobian of $\Phi(w)$.

With these index sets, and using a suitable reordering of the constraints, every element $V \in \pi_{(y,\lambda)}\partial\Phi(x,y,\lambda)^T$ has the following structure (the dependence on $w = (x,y,\lambda)$ is suppressed for simplicity):

$$V = \begin{pmatrix} \nabla^2_{yy}L & -\nabla g_+ & -\nabla g_{01} & -\nabla g_{02}(D_a)_{02} & 0 & 0 \\ \nabla g_+^T & 0 & 0 & 0 & 0 & 0 \\ \nabla g_{01}^T & 0 & 0 & 0 & 0 & 0 \\ \nabla g_{02}^T & 0 & 0 & (D_b)_{02} & 0 & 0 \\ \nabla g_{03}^T & 0 & 0 & 0 & I & 0 \\ \nabla g_<^T & 0 & 0 & 0 & 0 & I \end{pmatrix}, \tag{14}$$

where $(D_a)_{02}$ and $(D_b)_{02}$ are positive definite diagonal matrices, and where we used the abbreviations $\nabla g_+, \nabla g_{01}$ etc. for the matrices $\nabla g_{I_+}, \nabla g_{I_{01}}$ etc.

In order to obtain a suitable nonsingularity result, let us introduce the matrices

$$M(J) := \begin{pmatrix} \nabla^2_{yy}L & -\nabla g_+ & -\nabla g_J \\ \nabla g_+^T & 0 & 0 \\ \nabla g_J^T & 0 & 0 \end{pmatrix},$$

where J is any subset of I_{00}. Using these matrices, we next define the concept of strong regularity for the parameterized optimization problem (11). This name comes from the fact that our condition corresponds to Robinson's strong regularity assumption (see [34]) in the context of ordinary nonlinear programs, cf. [24, 6].

Definition 3.3 *A triple $w^* = (x^*, y^*, \lambda^*)$ satisfying $\Phi(w^*) = 0$ is called* strongly regular *for the optimization problem* (11) *if the matrices $M(J)$ have the same nonzero orientation for all $J \subseteq I_{00}$.*

According to Robinson [34], strong regularity holds if the strong second order sufficiency condition and the linear independence constraint qualification (LICQ for short) hold, where LICQ means that the gradients $\nabla g_i(x^*)$ ($i : g_i(x^*) = 0$) of the active inequality constraints are linearly independent (note that LICQ is also a necessary condition for strong regularity). In particular, it therefore follows that all matrices $M(J)$ have the same nonzero orientation if $\nabla^2_{yy}L$ is positive definite and LICQ holds. This is the situation we are particularly interested in. In fact, in this case, there is an easy way to see that strong regularity holds at $w^* = (x^*, y^*, \lambda^*)$. To this end, write

$$M(J) = \begin{pmatrix} H & -A_J \\ A_J^T & 0 \end{pmatrix} \quad \text{with} \quad H := \nabla^2_{yy}L \quad \text{and} \quad A_J := \big(\nabla g_+, \nabla g_J\big).$$

Using block Gaussian elimination, it follows that

$$M(J) = \begin{pmatrix} I & 0 \\ A_J^T H^{-1} & I \end{pmatrix} \begin{pmatrix} H & -A_J \\ 0 & A_J^T H^{-1} A_J \end{pmatrix}.$$

Consequently, we get

$$\det\big(M(J)\big) = \det \begin{pmatrix} H & -A_J \\ 0 & A_J^T H^{-1} A_J \end{pmatrix} = \det(H) \det\big(A_J^T H^{-1} A_J\big) > 0 \quad \forall J \subseteq I_{00}$$

since H is positive definite and A_J has full column rank for all $J \subseteq I_{00}$.

We next state our main result on the nonsingularity of the elements of the projected generalized Jacobian $\pi_{(y,\lambda)}\partial\Phi(x^*, y^*, \lambda^*)$. Its proof is similar to one given in [6] which, however, uses a different reformulation of the KKT system arising from variational inequalities.

Theorem 3.4 *Consider the optimization problem* (11) *with* $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$ *being twice continuously differentiable. Let* $w^* = (x^*, y^*, \lambda^*) \in \mathbb{R}^{n+n+m}$ *be a solution of the system* $\Phi(w) = 0$*, and suppose that the strong regularity condition holds at* w^**. Then all elements* $V \in \pi_{(y,\lambda)}\partial\Phi(w^*)$ *are nonsingular.*

Proof. Consider an arbitrary but fixed element in $\pi_{(y,\lambda)}\partial\Phi(w^*)^T$. This element has the structure indicated in (14) and is obviously nonsingular if and only if the following matrix is nonsingular:

$$V = \begin{pmatrix} \nabla^2_{yy} L & -\nabla g_+ & -\nabla g_{01} & -\nabla g_{02} \\ \nabla g_+^T & 0 & 0 & 0 \\ \nabla g_{01}^T & 0 & 0 & 0 \\ \nabla g_{02}^T & 0 & 0 & (D_b)_{02}(D_a)_{02}^{-1} \end{pmatrix}. \tag{15}$$

The matrix (15) can be written as the sum of the matrix $M(J)$, with $J = I_{01} \cup I_{02}$, and the diagonal matrix

$$D := \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (D_b)_{02}(D_a)_{02}^{-1} \end{pmatrix}.$$

Given a square matrix $\bar{A}$ of dimension r and a diagonal matrix $\bar{D}$ of the same dimension, it follows from [5, p. 60] that

$$\det(\bar{D} + \bar{A}) = \sum_{\alpha} \det\bar{D}_{\alpha\alpha} \det\bar{A}_{\bar{\alpha}\bar{\alpha}},$$

where the summation ranges over all subsets α of $\{1, \ldots, r\}$ (with complements $\bar{\alpha} = \{1, \ldots, r\} \backslash \alpha$), and where it is assumed that the determinant of an "empty" matrix is equal to 1. Exploiting this formula, the determinant of (15) can be written as

$$\det M(J) + \sum_{\emptyset \neq \alpha \subseteq I_{02}} \det D_{\alpha\alpha} \det M(J)_{\bar{\alpha}\bar{\alpha}}, \tag{16}$$

where the first term corresponds to $\alpha = \emptyset$. Moreover, we have taken into account that if α contains an element which does not belong to I_{02}, then the determinant of $D_{\alpha\alpha}$ is 0. Since the nonzero diagonal elements of the matrix D are all positive, it follows that the determinants of $D_{\alpha\alpha}$ in (16) are all positive. Then to show that the determinant of (15) is nonzero and hence to conclude the proof, it will now be sufficient to show that the determinants of $M(J)$ and of all $M(J)_{\bar{\alpha}\bar{\alpha}}$ in (16) never have opposite signs, and that at least one of them is nonzero. But this is a direct consequence of Definition 3.3. □

Now we are able to apply Theorem 2.3 to the optimization problem (11).

Corollary 3.5 *Let the assumptions of Theorem 3.4 be satisfied. Then there exists a neighbourhood U of x^* and a semismooth function $G : U \to \mathbb{R}^{n+m}, x \mapsto \big(y(x), \lambda(x)\big)$ such that $\Phi(x, G(x)) = 0$ holds for all $x \in U$. In particular, the mapping $x \mapsto y(x)$ is semismooth.*

Proof. The existence and semismoothness of the implicit function $x \mapsto G(x) = \big(y(x), \lambda(x)\big)$ is an immediate consequence of Theorems 2.3 and 3.4. Using Lemma 2.2, this, in particular, implies the local semismoothness of the mapping $x \mapsto y(x)$. □

We now get back to our GNEP and the mapping V_γ defined in (5). The following is the main result of this section.

Theorem 3.6 *Let x^* be a solution of the generalized Nash equilibrium problem, and assume that LICQ holds at $y_\gamma(x^*)$. Then V_γ is an SC^1-function in a neighbourhood of x^*.*

Proof. In view of the introductory remarks of this section, we have to show that the mapping $x \mapsto y_\gamma(x)$ is semismooth in a neighbourhood of x^*. By definition, $y_\gamma(x)$ is the solution of the optimization problem

$$\max_y \ \Psi_\gamma(x, y) \quad \text{s.t.} \quad y \in X := \{y \in \mathbb{R}^n \mid g(y) \leq 0\}, \tag{17}$$

cf. (6). This is an optimization problem of the form (11) with $f(x, y) := -\Psi_\gamma(x, y)$. Here, the mapping f is uniformly convex with respect to y due to the regularization term in the definition of the regularized Nikaido-Isoda-function and the assumed convexity of the mappings θ_ν with respect to the variables x^ν. Corollary 3.5 therefore gives the semismoothness of the mapping $x \mapsto y_\gamma(x)$ provided that the strong regularity assumption holds at $\big(x^*, y_\gamma(x^*), \lambda_\gamma(x^*)\big)$, where $y_\gamma(x^*)$ denotes the solution of problem (17) with $x = x^*$ and $\lambda_\gamma(x^*)$ is the corresponding unique (due to LICQ) multiplier.

Since LICQ holds at $y_\gamma(x^*)$, it suffices to show that the Hessian (with respect to y) of the corresponding Lagrangian

$$L_\gamma(x, y, \lambda) = -\Psi_\gamma(x, y) + \sum_{i=1}^{m} \lambda_i g_i(y)$$

is positive definite at $(x, y, \lambda) = \big(x^*, y_\gamma(x^*), \lambda_\gamma(x^*)\big)$, see the comments after Definition 3.3. However, we already observed that $-\Psi_\gamma(x, y)$ is uniformly convex with respect to y, hence its Hessian is (uniformly) positive definite. Furthermore, $\nabla^2 g_i(y_\gamma(x^*))$ is positive semidefinite due to the assumed convexity of the functions g_i. Hence the assertion follows from the fact that $\lambda = \lambda_\gamma(x^*)$ is nonnegative (as a multiplier corresponding to an inequality constraint). □

Note that, if, in addition to the assumptions of Theorem 3.6, strict complementarity holds at $y_\gamma(x^*)$, then y_γ is continuously differentiable and V_γ is a C^2-function in a neighbourhood of x^*. This follows directly from the previous derivation by using the standard implicit function theorem in place of Theorem 2.3.

Furthermore, we would like to point out that the assertion of Theorem 3.6 holds at all $x \in \mathbb{R}^n$ such that LICQ is satisfied at $y_\gamma(x)$.

4 Newton-type Methods

In view of Theorem 3.6, both the constrained optimization reformulation (8) and the unconstrained reformulation (9) of the GNEP are SC^1 optimization problems. Hence it is reasonable to believe that locally superlinearly convergent Newton-type methods can be derived for the solution of GNEPs via the solution of these optimization problems. Here we focus on the unconstrained reformulation (9) and show that one can indeed expect local fast convergence of a nonsmooth Newton-type method under suitable assumptions. The constrained reformulation (8) will be considered in a separate paper.

The nonsmooth Newton-type method from [32, 31] for the minimization of the unconstrained function $V_{\alpha\beta}$ from (10) is an iterative procedure of the form

$$x^{k+1} := x^k + d^k, \quad k = 0, 1, 2, \ldots, \tag{18}$$

where $x^0 \in \mathbb{R}^n$ is a starting point and d^k is a solution of the linear system

$$H_k d = -\nabla V_{\alpha\beta}(x^k) \quad \text{for some} \quad H_k \in \partial^2 V_{\alpha\beta}(x^k), \tag{19}$$

where $\partial^2 V_{\alpha\beta}(x^k)$ denotes the *generalized Hessian* of $V_{\alpha\beta}$ at x^k in the sense of [18], i.e., $\partial^2 V_{\alpha\beta}(x^k)$ is the generalized Jacobian in the sense of Clarke [3] of the locally Lipschitz mapping $F := \nabla V_{\alpha\beta}$.

In order to compute the gradient and (generalized) Hessian matrix of the mapping $V_{\alpha\beta}$, we need several (partial) derivatives of the mapping Ψ_γ from (4). These derivatives are summarized in the following result whose proof is omitted since it follows from standard calculus rules.

Lemma 4.1 *The mapping Ψ_γ from (4) is twice continuously differentiable with (partial)*

derivatives

$$
\begin{aligned}
\nabla_x \Psi_\gamma(x,y) &= \sum_{\nu=1}^{N} \left[\nabla\theta_\nu(x^\nu, x^{-\nu}) - \nabla\theta_\nu(y^\nu, x^{-\nu})\right] + \begin{pmatrix} \nabla_{x^1}\theta_1(y^1, x^{-1}) \\ \vdots \\ \nabla_{x^N}\theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma(x-y), \\
\nabla_y \Psi_\gamma(x,y) &= -\begin{pmatrix} \nabla_{x^1}\theta_1(y^1, x^{-1}) \\ \vdots \\ \nabla_{x^N}\theta_N(y^N, x^{-N}) \end{pmatrix} + \gamma(x-y), \\
\nabla^2_{xx} \Psi_\gamma(x,y) &= \sum_{\nu=1}^{N} \left[\nabla^2\theta_\nu(x^\nu, x^{-\nu}) - \nabla^2\theta_\nu(y^\nu, x^{-\nu})\right] \\
&\quad + \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^1x^N}\theta_N(y^N, x^{-N}) \\ \vdots & \ddots & \vdots \\ \nabla^2_{x^Nx^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} \\
&\quad + \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^1x^N}\theta_1(y^1, x^{-1}) \\ \vdots & \ddots & \vdots \\ \nabla^2_{x^Nx^1}\theta_N(y^N, x^{-N}) & \cdots & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} \\
&\quad - diag \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma I, \\
\nabla^2_{xy} \Psi_\gamma(x,y) &= -\begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^1x^N}\theta_N(y^N, x^{-N}) \\ \vdots & \ddots & \vdots \\ \nabla^2_{x^Nx^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} \\
&\quad + diag \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} + \gamma I, \\
\nabla^2_{yx} \Psi_\gamma(x,y) &= \nabla^2_{xy} \Psi_\gamma(x,y)^T \\
&= -\begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & \cdots & \nabla^2_{x^1x^N}\theta_1(y^1, x^{-1}) \\ \vdots & \ddots & \vdots \\ \nabla^2_{x^Nx^1}\theta_N(y^N, x^{-N}) & \cdots & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} \\
&\quad + diag \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} + \gamma I, \\
\nabla^2_{yy} \Psi_\gamma(x,y) &= -diag \begin{pmatrix} \nabla^2_{x^1x^1}\theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla^2_{x^Nx^N}\theta_N(y^N, x^{-N}) \end{pmatrix} - \gamma I.
\end{aligned}
$$

We next consider the problem of how to implement the nonsmooth Newton-type method. To this end, we have to compute, at each iterate x^k, an element $H_k \in \partial^2 V_{\alpha\beta}(x^k)$. Since this is not an easy task, we first assume that V_α and V_β are both twice continuously differentiable at x^k, hence $V_{\alpha\beta}$ is twice continuously differentiable at $x := x^k$ with Hessian

$$\nabla^2 V_{\alpha\beta}(x) = \nabla^2 V_\alpha(x) - \nabla^2 V_\beta(x). \tag{20}$$

Hence we need to calculate the Hessians $\nabla^2 V_\gamma(x)$ for $\gamma \in \{\alpha, \beta\}$. Therefore, let $\gamma \in \{\alpha, \beta\}$ be fixed and recall that V_γ is given by (5) with gradient $\nabla V_\gamma(x) = \nabla_x \Psi_\gamma\big(x, y_\gamma(x)\big)$, cf. (7), where $y_\gamma(x)$ denotes the solution of the optimization problem (6). Using the chain rule, we therefore get

$$\nabla^2 V_\gamma(x) = \nabla^2_{xx} \Psi_\gamma\big(x, y_\gamma(x)\big) + \nabla^2_{xy} \Psi_\gamma\big(x, y_\gamma(x)\big) Dy_\gamma(x), \tag{21}$$

where $Dy_\gamma(x) \in \mathbb{R}^{n\times n}$ denotes the usual Jacobian (with respect to x) of the mapping y_γ. Expressions for the matrices $\nabla^2_{xx} \Psi_\gamma\big(x, y_\gamma(x)\big)$ and $\nabla^2_{xy} \Psi_\gamma\big(x, y_\gamma(x)\big)$ are given in Lemma 4.1. At a nondifferentiable point, we have the following result.

Lemma 4.2 *The following inclusion holds at an arbitrary point $x \in \mathbb{R}^n$:*

$$\nabla^2_{xx} \Psi_\gamma(x, y_\gamma(x)) + \nabla^2_{xy} \Psi_\gamma(x, y_\gamma(x)) \partial_B y_\gamma(x) \subseteq \partial^2_B V_\gamma(x).$$

Proof. Let $x \in \mathbb{R}^n$ be arbitrarily given, and let $Y \in \partial_B y_\gamma(x)$. Then there is a sequence $\{\xi^k\} \to x$ such that y_γ is differentiable at each ξ^k and $Dy_\gamma(\xi^k) \to Y$ for $k \to \infty$. Then the representation (7) of ∇V_γ shows that V_γ is twice differentiable at each ξ^k, and we therefore obtain, taking the continuity of y_γ and the twice continuous differentiability of Ψ_γ into account:

$$\begin{aligned} \nabla^2 V_\gamma(\xi^k) &= \nabla^2_{xx} \Psi_\gamma(\xi^k, y_\gamma(\xi^k)) + \nabla^2_{xy} \Psi_\gamma(\xi^k, y_\gamma(\xi^k)) Dy_\gamma(\xi^k) \\ &\to \nabla^2_{xx} \Psi_\gamma(x, y_\gamma(x)) + \nabla^2_{xy} \Psi_\gamma(x, y(x)) Y. \end{aligned}$$

This shows that the right-hand side belongs to $\partial^2_B V_\gamma(x)$. □

Hence we only need to consider the computation of $\partial_B y_\gamma(x)$. By definition, $y_\gamma(x)$ is the unique solution of the optimization problem

$$\min_y -\Psi_\gamma(x, y) \quad \text{s.t.} \quad y \in X := \big\{y \in \mathbb{R}^n \mid g(y) \le 0\big\}.$$

Assume that LICQ holds at $y_\gamma(x)$, and let

$$L_\gamma(x, y, \lambda) := -\Psi_\gamma(x, y) + \sum_{i=1}^m \lambda_i g_i(y)$$

be the Lagrangian of this optimization problem. Since LICQ holds at $y_\gamma(x)$, it follows that there exist unique multipliers $\lambda_\gamma(x)$ such that the following KKT conditions hold at $(x, y, \lambda) = \big(x, y_\gamma(x), \lambda_\gamma(x)\big)$:

$$\nabla_y L_\gamma(x, y, \lambda) = 0, \quad \lambda \geq 0, \;\; g(y) \leq 0, \;\; \lambda^T g(y) = 0.$$

Here we have

$$\nabla_y L_\gamma(x, y, \lambda) = -\nabla_y \Psi_\gamma(x, y) + \sum_{i=1}^{m} \lambda_i \nabla g_i(y).$$

Therefore, assuming, for the moment, that strict complementarity holds, we then obtain from the standard implicit function theorem that the implicit function $G(x) := \big(y_\gamma(x), \lambda_\gamma(x)\big)$ satisfies (locally) the system of equations

$$\Phi_\gamma\big(x, G(x)\big) = 0, \quad \text{where} \quad \Phi_\gamma(x, y, \lambda) := \begin{pmatrix} \nabla_y L_\gamma(x, y, \lambda) \\ \min\{-g(y), \lambda\} \end{pmatrix}. \tag{22}$$

Differentiating this system therefore gives

$$0 = D_x \Phi_\gamma(x, G(x)) = D_x \Phi_\gamma(x, G(x)) + D_{(y,\lambda)} \Phi_\gamma(x, G(x)) D_x G(x),$$

from which we obtain $D_x G(x) = \big(D_x y_\gamma(x), D_x \lambda_\gamma(x)\big)$ by solving the linear system

$$D_{(y,\lambda)} \Phi_\gamma(x, G(x)) D_x G(x) = -D_x \Phi_\gamma(x, G(x)). \tag{23}$$

Note, however, that we are only interested in the first part of $D_x G(x)$ (namely $D_x y_\gamma(x)$)). This expression for $D_x G(x)$ is helpful for the computation of an element from $\partial_B y_\gamma(x)$.

Lemma 4.3 *Let $x \in \mathbb{R}^n$ and $\Theta \in \pi_{(y,\lambda)} \partial_B \Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$ be arbitrarily given, where $\pi_{(y,\lambda)} \partial_B \Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$ denotes the projection of the B-subdifferential $\partial_B \Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$ onto the (y, λ)-space. Suppose that LICQ holds at $y_\gamma(x)$, and let*

$$Y := \big\{\Theta^{-1}\big\}_{\{1,\ldots,n\} \times \{1,\ldots,n\}} \nabla^2_{yx} \Psi_\gamma(x, y_\gamma(x)),$$

where $\big\{\Theta^{-1}\big\}_{\{1,\ldots,n\} \times \{1,\ldots,n\}}$ denotes the upper left $n \times n$ submatrix of Θ^{-1}. Then $Y \in \partial_B y_\gamma(x)$.

Proof. First consider an arbitrary point ξ such that Φ_γ is differentiable at ξ and LICQ holds at $y_\gamma(\xi)$. Then, locally, we have $\Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi)) = 0$. As before, differentiation therefore yields

$$0 = D_x \Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi)) + D_{(y,\lambda)} \Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi)) \begin{pmatrix} D_x y_\gamma(x) \\ D_x \lambda_\gamma(x) \end{pmatrix}.$$

Taking a closer look at the function Φ_γ, we see that

$$D_x \Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi)) = \begin{pmatrix} -\nabla^2_{yx} \Psi_\gamma(\xi, y_\gamma(\xi)) \\ 0 \end{pmatrix}.$$

Noting that $D_{(y,\lambda)}\Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi))$ is nonsingular in view of LICQ (cf. the proof of Theorems 3.4, 3.6), we therefore obtain

$$\begin{aligned}\begin{pmatrix} Dy_\gamma(x) \\ D\lambda_\gamma(x) \end{pmatrix} &= -\big(D_{(y,\lambda)}\Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi))\big)^{-1} \begin{pmatrix} -\nabla^2_{yx}\Psi_\gamma(\xi, y_\gamma(\xi)) \\ 0 \end{pmatrix} \\ &= \big\{\big(D_{(y,\lambda)}\Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi))\big)^{-1}\big\}_{\{1,\ldots,n+m\}\times\{1,\ldots,n\}} \nabla^2_{yx}\Psi_\gamma(\xi, y_\gamma(\xi)).\end{aligned}$$

This implies

$$Dy_\gamma(x) = \big\{\big(D_{(y,\lambda)}\Phi_\gamma(\xi, y_\gamma(\xi), \lambda_\gamma(\xi))\big)^{-1}\big\}_{\{1,\ldots,n\}\times\{1,\ldots,n\}} \nabla^2_{yx}\Psi_\gamma(\xi, y_\gamma(\xi)). \tag{24}$$

Now let $x \in \mathbb{R}^n$ be arbitrarily given and $\Theta \in \pi_{(y,\lambda)}\partial_B\Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$. Then there is a sequence $\{\xi^k\} \to x$ such that Φ_γ is differentiable at each ξ^k with $D_{(y,\lambda)}\Phi_\gamma(\xi^k, y_\gamma(\xi^k), \lambda_\gamma(\xi^k)) \to \Theta$. Since $\nabla^2_{yx}\Psi_\gamma, y_\gamma$, and λ_γ are continuous and all elements in $\pi_{(y,\lambda)}\partial_B\Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$ are nonsingular in view of LICQ (see, once again, the proof of Theorems 3.4, 3.6), it follows from the first part of the proof that

$$\begin{aligned}Dy_\gamma(\xi^k) &= \big\{\big(D_{(y,\lambda)}\Phi_\gamma(\xi^k, y_\gamma(\xi^k), \lambda_\gamma(\xi^k))\big)^{-1}\big\}_{\{1,\ldots,n\}\times\{1,\ldots,n\}} \nabla^2_{yx}\Psi_\gamma(\xi^k, y_\gamma(\xi^k)) \\ &\to \big\{\Theta^{-1}\big\}_{\{1,\ldots,n\}\times\{1,\ldots,n\}} \nabla^2_{yx}\Psi_\gamma(x, y_\gamma(x)).\end{aligned}$$

Consequently, the right-hand side belongs to $\partial_B y_\gamma(x)$. □

Note that the element $\Theta \in \pi_{(y,\lambda)}\partial_B\Phi_\gamma(x, y_\gamma(x), \lambda_\gamma(x))$, that is required for the computation of an element from $\partial_B y_\gamma(x)$ in Lemma 4.3, can be computed via Lemma 3.2.

The following is the central local convergence result for our nonsmooth Newton-type method for the solution of the GNEP.

Theorem 4.4 *Let x^* be a normalized Nash equilibrium of the GNEP such that all elements $V \in \partial^2 V_{\alpha\beta}(x^*)$ are nonsingular. Then the nonsmooth Newton-type method from* (18), (19) *is locally superlinearly convergent to x^*.*

Proof. Since $V_{\alpha\beta}$ is an SC^1-function in view of Theorem 3.6, the result follows immediately from [32]. □

There are a number of comments that we would like to add in the following remark.

Remark 4.5 (a) Theorem 4.4 remains true if we replace the assumption that all elements of $\partial^2 V_{\alpha\beta}(x^*)$ are nonsingular by the weaker condition that all elements from the smaller set $\partial^2_B V_{\alpha\beta}(x^*)$ are nonsingular, where $\partial^2_B V_{\alpha\beta}(x^*)$ denotes the B-subdifferential of the locally Lipschitz continuous mapping $F(x) := \nabla V_{\alpha\beta}(x)$. This follows immediately from a result in [31].

(b) Theorem 4.4 is a local convergence result only. However, since $V_{\alpha\beta}$ is continuously differentiable, it is easy to globalize this method by either a line search or a trust-region strategy. These globalized methods typically find a stationary point of $V_{\alpha\beta}$ only, and a sufficient condition for such a stationary point to be a normalized Nash equilibrium of the GNEP is given in [17].

(c) Theorem 4.4 gives a local superlinear rate of convergence. It is also possible to get a locally quadratically convergent method by our approach. To this end, we have to strengthen Assumption 1.1 to some extend and assume that, in addition, the Hessian matrices $\nabla^2\theta_\nu$ and $\nabla^2 g_i$ are locally Lipschitz around a solution x^* of the GNEP. Moreover, one has to use another implicit function theorem like the one from Sun [37] in order to guarantee a local quadratic approximation property (as defined in [37]) or to modify Theorem 2.3 in a suitable way.

(d) A simple sufficient condition for the nonsingularity of all elements from $\partial^2 V_{\alpha\beta}(x^*)$ (or $\partial_B^2 V_{\alpha\beta}(x^*)$) is as follows: Suppose that the solution x^* from Theorem 4.4 is locally unique (for which there are simple sufficient conditions like LICQ and a suitable second-order condition) and satisfies strict complementarity, i.e., the corresponding vectors $y_\gamma(x^*)$ and $\lambda_\gamma(x^*)$ satisfy the strict complementarity condition. Then the standard implicit function theorem guarantees that $V_{\alpha\beta}$ is twice continuously differentiable around x^*. Therefore, the local uniqueness of the normalized solution x^* implies that the Hessian $\nabla^2 V_{\alpha\beta}(x^*)$ is positive definite. Note, however, that this technique of proof cannot be used in the case where strict complementarity does not hold, cf. [18] for a counterexample (in a different context).

5 Numerical Results

We implemented a semismooth Newton-type method for the solution of generalized Nash equilibrium problems via the unconstrained optimization reformulation (9). To this end, we compute the elements from $\partial_B^2 V_{\alpha\beta}(x)$ in a way outlined in Lemmas 4.2 and 4.3, and we use a simple Armijo-type line search in order to globalize this method. Moreover, we switch to the steepest descent direction whenever the generalized Newton direction is not computable or does not satisfy a sufficient decrease condition. In our experiments, however, we were always able to take the generalized Newton direction. The method is terminated whenever $V_{\alpha\beta}(x^k) \leq \varepsilon$ with $\varepsilon := 10^{-8}$ and uses the two parameters $\alpha = 0.01, \beta = 0.05$ for the definition of $V_{\alpha\beta}$. The implementation is done in MATLAB, using the build-in function `fmincon` from the Optimization Toolbox in order to calculate $y_\gamma(x^k)$ at each iteration k.

Example 5.1 This test problem is the river basin pollution game taken from [20]. Remarkably, the problem is solved in one iteration only, see Table 1. This seems to indicate that our objective function $V_{\alpha\beta}$ is quadratic, though we have not calculated it by hand.

k	x_1^k	x_2^k	x_3^k	$V_{\alpha\beta}(x^k)$	$\|\nabla V_{\alpha\beta}(x^k)\|$
0	0.0000000000	0.0000000000	0.0000000000	10.297117	0.8614643071
1	21.1448955742	16.0278402662	2.7258866516	0.0000000010	0.0000044546

Table 1: Numerical results for Example 5.1

Example 5.2 This test problem is an internet switching model introduced by Kesselman et al. [19] and also analysed by Facchinei et al. [8]. We modify this example slightly and add the additional constraint $x^\nu \geq 0.01, \nu = 1, \ldots, N$ in order to avoid possible domain violations of the utility functions (which contain logarithmic terms). This does not alter the solution. We set $N = 10$ (and $B = 1$ in the description of the model in [8]) and use the starting point $x^0 = (0.1, \ldots, 0.1)^T$. The exact solution of this problem is $x^* = (0.9, \ldots, 0.9)^T$. We only state the first two components of the iteration vector in Table 2.

k	x_1^k	x_2^k	$V_{\alpha\beta}(x^k)$	$\|\nabla V_{\alpha\beta}(x)\|$
0	0.100000	0.100000	0.000512031434	0.0331501603
1	0.092101	0.092101	0.000021749587	0.0068259609
2	0.090046	0.090046	0.000000010343	0.0001459690

Table 2: Numerical results for Example 5.2

Example 5.3 This test problem is the Cournot oligopoly problem with shared constraints and nonlinear payoff functions as described in Outrata et al. [28, p. 233]. Our results (using different values for the parameter P from [28]) are given in Table 3. Note that our solutions are different from those presented in [28]. We therefore checked some additional termination criteria which showed that our computed solutions seem to be correct.

Example 5.4 Here we solve the electricity market problem suggested by Contreras et al. [4]. This model involves three power generating companies with one, two, and three power plants, respectively. We consider the game where restriction is only imposed on the power production of each power plant, which corresponds to Case Study 1 in [4]. Note that this is a standard Nash equilibrium problem. The corresponding numerical results are given in Table 4.

Example 5.5 Here we consider a simple two-player game originally suggested by Rosen [35]. The solution violates strict complementarity. More precisely, our example has the two payoff functions

$$\theta_1(x_1, x_2) = 0.5x_1^2 - x_1x_2 \quad \text{and} \quad \theta_2(x_1, x_2) = x_2^2 + x_1x_2$$

k	x_1^k	x_2^k	x_3^k	x_4^k	x_5^k	$V_{\alpha\beta}(x^k)$	$\|\nabla V_{\alpha\beta}(x^k)\|$
$P = 75$							
0	10.00000	10.00000	10.00000	10.00000	10.00000	2.6786204345	0.43990983
1	11.02142	13.09565	15.14096	16.94890	18.38387	0.0065309982	0.01476791
2	10.40614	13.03456	15.40627	17.38103	18.77081	-0.0000000644	0.00018618
profits	359.0655	472.6096	584.9635	688.6138	773.1191	—	—
$P = 100$							
0	10.00000	10.00000	10.00000	10.00000	10.00000	40.1459577318	1.78414125
1	26.21177	28.81211	30.81873	31.84833	31.46531	0.1143197642	0.06811610
2	23.59919	28.67806	32.02794	33.28246	32.41732	0.0000028000	0.00029325
3	23.58896	28.68095	32.02596	33.28957	32.41520	0.0000005150	0.00015646
4	23.58873	28.68424	32.02141	33.28740	32.41821	-0.0000000195	0.00001219
profits	288.9038	392.1367	481.3594	543.9701	571.7381	—	—
$P = 150$							
0	10.00000	10.00000	10.00000	10.00000	10.00000	10.1939346932	0.88576221
1	15.43209	17.97838	20.32408	22.22361	23.33434	0.0330976856	0.03454114
2	14.05544	17.79522	20.90496	23.11128	24.13275	0.0000003336	0.00012263
3	14.05014	17.79917	20.90723	23.11018	24.13270	0.0000000225	0.00003063
4	14.05012	17.79836	20.90715	23.11144	24.13290	-0.0000000000	0.00000048
profits	333.6345	451.6328	563.3555	657.3081	720.5298	—	—
$P = 200$							
0	10.00000	10.00000	10.00000	10.00000	10.00000	90.0826591342	2.68177086
1	38.41325	40.11679	41.03357	40.81589	39.08101	0.1307246076	0.07614674
2	35.78794	40.74931	42.80285	41.96721	38.69809	-0.0000002203	0.00012314
profits	209.1723	290.5175	358.5510	403.7686	422.5701	—	—

Table 3: Numerical results for Example 5.3

and the joint constraints given by

$$X := \{x \in \mathbb{R}^2 \mid x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \geq 1\}.$$

The unique normalized Nash equilibrium of this GNEP is $x^* = (1, 0)^T$ and does not satisfy strict complementarity since an easy calculation shows that $y_\gamma(x^*) = (1, 0)^T$ and $\lambda_\gamma(x^*) = (0, 0, 1)^T$, hence strict complementarity does not hold in the second component. Table 5 shows our results using the starting point $x^0 := (1, 1)^T$. We also tried a variety of other starting points, each of which gives convergence in just one iteration. Hence we still observe very fast convergence although strict complementarity is violated.

6 Final Remarks and Discussion

We further investigated the properties of suitable optimization reformulations of the generalized Nash equilibrium problem and showed that these optimization problems are SC^1-

k	x_1^k	x_2^k	x_3^k	$V_{\alpha\beta}(x^k)$	$\|\nabla V_{\alpha\beta}(x^k)\|$
0	0.000000	0.000000	0.000000	285.800638776273	7.2816852177
1	79.999386	15.420192	10.119499	9.860455473615	1.3148442024
2	46.659109	22.110930	12.341396	0.000000646648	0.0004393538
3	46.654084	22.110549	12.342651	-0.000008339256	0.0104832874

Table 4: Numerical results for Example 5.4

k	x_1^k	x_2^k	$V_{\alpha\beta}(x^k)$	$\|\nabla V_{\alpha\beta}(x^k)\|$
0	1.000000	1.000000	0.020000000000	0.0400000000
1	1.000000	0.000000	-0.000000000000	0.0000000000

Table 5: Numerical results for Example 5.5 using $x^0 = (1, 1)^T$

problems. This paves the way to locally superlinearly convergent methods for the solution of GNEPs, and one such (Newton-type) method for an unconstrained optimization reformulation was analyzed in some more detail.

We believe that, in the moment, the two most promising approaches for the solution of GNEPs are

- the Newton-type methods based on the (regularized) Nikaido-Isoda-function as discussed in this paper and its predecessor [17] (NI-approach for short)
- the Newton-type methods based on a VIP-reformulation as outlined, for example, in [7, 8] (VIP-approach for short).

Both methods have its advantages and disadvantages. They both allow application of suitable Newton-type methods in order to get local fast convergence. The NI-approach is more expensive than the VIP-approach since it requires the solution of a constrained optimization problem in order to evaluate the mapping V_γ. Hence it is very tempting to say that the VIP-approach is more promising. On the other hand, the VIP-approach gives local fast convergence in (x, λ) (with λ being a suitable Lagrange multiplier), whereas the NI-approach gives local fast convergence in x alone which is a stronger property. Moreover, the level sets of $V_{\alpha\beta}$ are automatically compact for, e.g., compact strategy sets X, cf. [17], thus guaranteeing global convergence results, whereas such a nice boundedness property is not known for Newton-type methods applied to the VIP-approach. Moreover, a globalization of the VIP-approach from [8] has not been investigated in the literature so far, neither theoretically nor numerically. This has certainly to be done in the near future in order to see which approach is really more promising from a practical point of view.

Finally, though this is less important for this paper, we would like to point out that the NI-approach can also be used for the case of continuous utility functions, whereas the VI-approach needs these mappings to be continuously differentiable. In such a case, however, we also have to minimize the (regularized) Nikaido-Isoda function by derivative-free methods.

References

[1] T. Başar: *Relaxation techniques and asynchronous algorithms for on-line computation of non-cooperative equilibria.* Journal of Economic Dynamics and Control 11, 1987, pp. 531–549.

[2] A. Bensoussan: *Points de Nash dans le cas de fonctionelles quadratiques et jeux differentiels lineaires a N personnes.* SIAM Journal on Control 12, 1974, pp. 460–499.

[3] F.H. Clarke: *Optimization and Nonsmooth Analysis.* John Wiley, New York, 1983 (reprinted by SIAM, Philadelphia, 1990).

[4] J. Contreras, M. Klusch, and J.B. Krawczyk: *Numerical solutions to Nash-Cournot equilibria in coupled constraint electricity markets.* IEEE Transactions on Power Systems 19, 2004, pp. 195–206.

[5] R.W. Cottle, J.-S. Pang, and R.E. Stone: *The Linear Complementarity Problem.* Academic Press, Boston, 1992.

[6] F. Facchinei, A. Fischer, and C. Kanzow: *Regularity properties of a semismooth reformulation of variational inequalities.* SIAM Journal on Optimization 8, 1998, pp. 850–869.

[7] F. Facchinei, A. Fischer, and V. Piccialli: *On generalized Nash games and VIs.* Operations Research Letters, to appear.

[8] F. Facchinei, A. Fischer, and V. Piccialli: *Generalized Nash equilibrium problems and Newton methods.* Technical Report, Department of Computer and System Sciences "A. Ruberti", Università di Roma "La Sapienza", Rome, Italy, February 2006.

[9] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I.* Springer, New York, NY, 2003.

[10] F. Facchinei and J.-S. Pang: *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume II.* Springer, New York, NY, 2003.

[11] A. Fischer: *Solution of monotone complementarity problems with locally Lipschitzian mappings.* Mathematical Programming 76, 1997, pp. 513–532.

[12] S.D. Flåm and A.S. Antipin: *Equilibrium programming using proximal-like algorithms.* Mathematical Programming 78, 1997, pp. 29–41.

[13] S.D. Flåm and A. Ruszczyński: *Noncooperative convex games: Computing equilibria by partial regularization.* IIASA Working Paper 94-42, Laxenburg, Austria, May 1994.

[14] M.S. Gowda: *Inverse and implicit function theorems for H-differentiable and semismooth functions.* Optimization Methods and Software 19, 2004, pp. 443–461.

[15] G. Gürkan and J.-S. Pang: *Approximations of Nash equilibria.* Technical Report, Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY, January 2006.

[16] P.T. Harker: *Generalized Nash games and quasivariational inequalities.* European Journal of Operations Research 54, 1991, pp. 81–94.

[17] A. von Heusinger and C. Kanzow: *Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions.* Technical Report, Institute of Mathematics, University of Würzburg, Germany, July 2006.

[18] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V.H. Nguyen: *Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data.* Applied Mathematics and Optimization 11, 1984, pp. 43–56.

[19] A. Kesselman, S. Leonardi, and V. Bonifaci: *Game-theoretic analysis of internet switching with selfish users.* Lecture Notes in Computer Science 3828, 2005, pp. 236–245.

[20] J.B. Krawczyk and S. Uryasev: *Relaxation algorithms to find Nash equilibria with economic applications.* Environmental Modeling and Assessment 5, 2000, pp. 63–73.

[21] B. Kummer: *Newton's method for non-differentiable functions.* In J. Guddat et al. (eds.): *Advances in Mathematical Optimization.* Akademie-Verlag, Berlin, 1988, pp. 114–125.

[22] B. Kummer: *Newton's method based on generalized derivatives for nonsmooth functions: Convergence analysis.* In W. Oettli and D. Pallaschke (eds.): *Advances in Optimization.* Lecture Notes in Economics and Mathematical Systems 382, Springer-Verlag, Berlin, 1992, pp. 171–194.

[23] S. Li and T. Başar: *Distributed algorithms for the computation of noncooperative equilibria.* Automatica 23, 1987, pp. 523–533.

[24] J. Liu: *Strong stability in variational inequalities.* SIAM Journal on Control and Optimization 33, 1995, pp. 725–749.

[25] G. Mastroeni: *Gap functions for equilibrium problems.* Journal of Global Optimization 27, 2003, pp. 411–426.

[26] R. Mifflin: *Semismooth and semiconvex functions in constrained optimization.* SIAM Journal on Control and Optimization 15, 1977, pp. 957–972.

[27] H. Nikaido and K. Isoda: *Note on noncooperative convex games.* Pacific Journal of Mathematics 5, 1955, pp. 807–815.

[28] J. Outrata, M. Kocvara and J. Zowe *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer, 1998.

[29] J.-S. Pang and M. Fukushima: *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games.* Computational Management Science 2, 2005, pp. 21–56.

[30] J.-S. Pang, D. Sun, and J. Sun: *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems.* Mathematics of Operations Research 28, 2003, pp. 39–63.

[31] L. Qi: *Convergence analysis of some algorithms for solving nonsmooth equations.* Mathematics of Operations Research 18, 1993, pp. 227–244.

[32] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–367.

[33] H. Rademacher: *Über partielle und totale Differenzierbarkeit I.* Mathematical Annals 89, 1919, pp. 340–359.

[34] S.M. Robinson: *Strongly regular generalized equations.* Mathematics of Operations Research 5, 1980, pp. 43–62.

[35] J.B. Rosen: *Existence and uniqueness of equilibrium points for concave N-person games.* Econometrica 33, 1965, pp. 520–534.

[36] A. Shapiro: *On concepts of directional differentiability.* Journal of Optimization Theory and Applications 66, 1990, pp. 477–487.

[37] D. Sun: *A further result on an implicit function theorem for locally Lipschitz functions.* Operations Research Letters 28, 2001, pp. 193–198.

[38] S. Uryasev and R.Y. Rubinstein: *On relaxation algorithms in computation of noncooperative equilibria.* IEEE Transactions on Automatic Control 39, 1994, pp. 1263–1267.

[39] L. Zhang and J. Han: *Unconstrained optimization reformulation of equilibrium problems.* Technical Report, Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China, 2006.