
Globally Convergent Algorithms for the Solution of Generalized Nash Equilibrium Problems

Dissertation zur Erlangung
des naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

AXEL DREVES

aus
Marburg

Würzburg, 2011

Eingereicht am: 15. Dezember 2011

1. Gutachter: Prof. Dr. Christian Kanzow, Universität Würzburg
 2. Gutachter: Prof. Dr. Oliver Stein, Karlsruhe Institute of Technology
-

Acknowledgments

First of all I thank my advisor Christian Kanzow. His guidance through the landscape of game theory and his admirable intuition to send me on paths that lead to unknown, surprising and gorgeous views on generalized Nash equilibrium problems made the past three years an exciting, challenging and variedly tour for me. Without his constant support and his ever open door I surely would still be searching for a way through the jungle of reformulations, algorithms or convergence properties. Special thanks for the endurance in reading my new results, also the technical ones, and the hints for improvement.

Further I would like to thank Oliver Stein and Francisco Facchinei for our joint research and, in particular, for the discussions during their visits in Würzburg. Moreover thanks to Anna von Heusinger, Simone Sagratella and Masao Fukushima for the joint work which lead to the papers and journal articles that are the basis of this thesis.

Finally I am grateful to the Deutsche Forschungsgemeinschaft (DFG) for financially supporting my work under grant KA 1296/17-1, thus giving me the time to concentrate fully on my research.

Contents

1. Introduction	1
1.1. Abbreviations and Notations	1
1.2. Generalized Nash Equilibrium Problems	3
1.3. Existing Literature	6
1.4. Summary	7
2. Analytical Solution of Small GNEPs	9
3. A Globalized Newton Method for Normalized Nash Equilibria	15
3.1. Preliminaries	15
3.2. A Globalization of a Newton Method via $V_{\alpha\beta}$	19
4. Optimization Reformulations of GNEPs Using the Nikaido-Isoda Function	25
4.1. Optimization Reformulations	25
4.2. Smoothness Properties of the Constrained Reformulation	37
4.3. Smoothness Properties of the Unconstrained Reformulation	46
4.4. Stationarity	51
5. Solving the KKT System of a GNEP	57
5.1. Merit Function Approach	58
5.1.1. Stationarity Conditions	59
5.1.2. Coercivity	63
5.2. Interior Point Method	67
5.2.1. Nonsingularity Conditions	73
5.2.2. Boundedness	78
5.2.3. A Finite Termination Result	79
6. Numerical Methods	103
6.1. Globalized Newton Method via $V_{\alpha\beta}$	105
6.2. Gradient Sampling Algorithm	107
6.3. Matlab's fminunc Solver	110
6.4. Semismooth-like Minimization Algorithm	110
6.5. Potential Reduction Method	113
6.6. Scaled Trust-Region Solver for Constrained Nonlinear Equations	115
6.7. Comparison of the Algorithms Solving the KKT System	116

7. Final Remarks	119
A. Tables for Numerical Results	127
A.1. Data of the Test Problems	129
A.2. The Globalized and Local Newton Methods	130
A.3. The Gradient Sampling Algorithm	131
A.4. Results for Solving the KKT System	134

1. Introduction

As the title reveals, this thesis deals with algorithms for the solution of generalized Nash equilibrium problems. In the introduction a definition of these problems and a discussion of the used solution concept is given. Further an overview on existing approaches for its solution is presented. Therefore it is necessary to introduce some notations first.

1.1. Abbreviations and Notations

The abbreviations used throughout this thesis are summarized in Table 1.1.

s.t.	subject to
NEP	Nash Equilibrium Problem
GNEP	Generalized Nash Equilibrium Problem
VI	Variational Inequality Problem
QVI	Quasi-Variational Inequality Problem
KKT	Karush-Kuhn-Tucker
CRCQ	Constant Rank Constraint Qualification
CE	Constrained System of Equations
GMRES	Generalized Minimal Residual Method
BFGS	Broyden-Fletcher-Goldfarb-Shanno
STRSCNE	Scaled Trust-Region Solver for Constrained Equations

Table 1.1.: Abbreviations

Some standard notations are listed in Table 1.2.

\mathbb{N}	the natural numbers
\mathbb{R}^n	the n -dimensional real vector space
\mathbb{R}_+^n	$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \geq 0 \text{ for all } i = 1, \dots, n\}$
\mathbb{R}_-^n	$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i \leq 0 \text{ for all } i = 1, \dots, n\}$
\mathbb{R}_{++}^n	$\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_i > 0 \text{ for all } i = 1, \dots, n\}$
$\mathbb{B}_r(x)$	the open ball with radius r and centre x
$\text{int}(X)$	the interior of the set X
$\text{bd}(X)$	the boundary of the set X
$\text{cl}(X)$	the closure of the set X
$\text{conv}(X)$	the convex hull of the set X

$\ \cdot\ $	the Euclidean vector norm or the 2-matrix norm
$\ x\ _1$	$\sum_{i=1}^n x_i $, the Manhattan vector norm of a vector $x \in \mathbb{R}^n$
$\lceil x \rceil$	the smallest integer larger than $x \in \mathbb{R}$
$x \circ y$	$(x_1y_1, \dots, x_ny_n)^T$, the vector containing the products of the components of $x, y \in \mathbb{R}^n$
0_n	the vector $(0, \dots, 0)^T \in \mathbb{R}^n$
1_n	the vector $(1, \dots, 1)^T \in \mathbb{R}^n$
I_n	the identity matrix in $\mathbb{R}^{n \times n}$
M^{-T}	the transposed of the inverse of a nonsingular matrix M
$M_{\alpha\beta}$	the submatrix of M that contains only those rows and columns whose indices are in α and β , respectively
$\text{eig}_{\min}(M)$	the smallest eigenvalue of a symmetric matrix M
$\text{diag}(x)$	the diagonal matrix with the vector x on its diagonal
$\text{dom}(f)$	$\{x \in \mathbb{R}^n \mid f(x) \neq \emptyset\}$, the domain of the point-to-set map $f : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$
$JH(x, y)$	the Jacobian of the differentiable function $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ at (x, y)
$J_xH(x, y)$	the Jacobian of the differentiable function $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ at (x, y) with respect to x
$\nabla H(x, y)$	the transposed of $JH(x, y)$
$\nabla H_J(x, y)$	the matrix with column vectors $\{\nabla H_j(x, y)\}_{j \in J}$
$\nabla_x H(x, y)$	the transposed of $J_xH(x, y)$
$\nabla^2 \Psi(x, y)$	the Hessian matrix of the twice differentiable function $\Psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ at (x, y)
$\nabla_{xy}^2 \Psi(x, y)$	the matrix of the second partial derivatives of $\Psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ at (x, y) , first differentiated with respect to x and then with respect to y
$\partial_B F(x, y)$	the B-subdifferential of the locally Lipschitz continuous function $F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ at (x, y)
$\partial F(x, y)$	the generalized Jacobian of the locally Lipschitz continuous function $F : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ at (x, y) in the sense of Clarke [12]
$\pi_y \partial F(x, y)$	the set of all matrices $M \in \mathbb{R}^{n \times n}$ such that there exists a matrix $N \in \mathbb{R}^{n \times m}$ with $[N, M] \in \partial F(x, y)$

Table 1.2.: Notations

Beside this some further notation should be explained:

- $P_X[x]$ stands for the Euclidean projection of a vector $x \in \mathbb{R}^n$ on a nonempty, closed and convex set $X \subseteq \mathbb{R}^n$, i.e., $P_X[x]$ is the unique solution of

$$\min_z \frac{1}{2} \|z - x\|^2 \quad \text{s.t.} \quad z \in X.$$

- A function $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called a PC^1 (*piecewise continuously differentiable*) function in a neighbourhood of a given point x^* , if g is continuous and there exist a neighbourhood U of x^* and a finite family of continuous differentiable functions

$\{G_1, G_2, \dots, G_k\}$ defined on U , such that $g(x) \in \{G_1(x), G_2(x), \dots, G_k(x)\}$ for all $x \in U$.

- A matrix $M \in \mathbb{R}^{n \times n}$ is called a P_0 -matrix if $\det(M_{\alpha\alpha}) \geq 0$ for all $\alpha \subseteq \{1, 2, \dots, n\}$, and a P -matrix if $\det(M_{\alpha\alpha}) > 0$ for all $\alpha \subseteq \{1, 2, \dots, n\}$. Note that the class of P_0 -matrices strictly includes the positive semidefinite matrices, see [13] for more details.

1.2. Generalized Nash Equilibrium Problems

The generalized Nash equilibrium problem, GNEP for short, is a game which is described through a finite set of N players $\nu = 1, \dots, N$, each having a cost function and a strategy set. All players ν control their variables $x^\nu \in \mathbb{R}^{n_\nu}$ and the vector $x = ((x^1)^T, \dots, (x^N)^T)^T \in \mathbb{R}^n$ with $n = n_1 + \dots + n_N$ describes the decision vector of all players. To emphasize the role of player ν the notation $x = (x^\nu, x^{-\nu})$ is often used, where $x^{-\nu}$ contains the variables of all players except the ν -th one, but the ordering of the vector components is kept. The cost function $\theta_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ of the ν -th player can depend on all player's decisions, and the strategy set, or feasible set, $X_\nu(x^{-\nu}) \subseteq \mathbb{R}^{n_\nu}$ of the ν -th player can depend on the variables $x^{-\nu}$ of the other players. In the game every player ν tries to minimize his cost function within his feasible set, that is, he solves the optimization problem

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{subject to} \quad x^\nu \in X_\nu(x^{-\nu}). \quad (1.1)$$

By defining the Cartesian product of the strategy spaces

$$\Omega(x) := X_1(x^{-1}) \times \dots \times X_N(x^{-N}),$$

it is possible to give a definition of a solution of a GNEP.

Definition 1.1 *A vector $\bar{x} \in \Omega(\bar{x})$ is called a generalized Nash equilibrium, or a solution of the GNEP, if*

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{for all} \quad x^\nu \in X_\nu(\bar{x}^{-\nu})$$

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

In order to find a solution of a GNEP, one has to solve problem (1.1) simultaneously for each player $\nu = 1, \dots, N$. To do so, it is quite standard to assume that the feasible sets are defined explicitly by constraint functions, i.e.,

$$X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R}^{n_\nu} \mid g^\nu(x^\nu, x^{-\nu}) \leq 0\} \quad (1.2)$$

for suitable functions $g^\nu : \mathbb{R}^n \rightarrow \mathbb{R}^{m_\nu}$, $\nu = 1, \dots, N$. The total number of constraints is $m := m_1 + \dots + m_N$. To guarantee solvability of (1.1) the following standard assumption is used.

Assumption 1.2 (i) For all $\nu = 1, \dots, N$ the cost functions $\theta_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ are continuous and, as a function of x^ν alone, convex.

(ii) For all $\nu = 1, \dots, N$ and all $i = 1, \dots, m_\nu$, the component functions $g_i^\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ are continuous and, as a function of x^ν alone, convex.

GNEPs satisfying Assumption 1.2 are called *player convex* and they are in the main focus of this thesis.

There are some important subclasses of the GNEPs that are obtained by requiring special structures for the strategy sets. In the prominent *Nash equilibrium problem*, NEP for short, the strategy spaces of each player are independent of the decisions of all the other players, that is there are fixed sets X_ν with $X_\nu(x^{-\nu}) = X_\nu$ for all strategies $x \in \mathbb{R}^n$ and all players $\nu = 1, \dots, N$. More general are the *jointly convex* GNEPs, where a common convex strategy space $X \subseteq \mathbb{R}^n$ exists such that the feasible set of player ν is given by

$$X_\nu(x^{-\nu}) = \{x^\nu \in \mathbb{R}^{n_\nu} \mid (x^\nu, x^{-\nu}) \in X\}.$$

In the setting of (1.2) this means $g^1 = \dots = g^N =: g$ and

$$X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R}^{n_\nu} \mid g(x^\nu, x^{-\nu}) \leq 0\}$$

for all $\nu = 1, \dots, N$, with a common function g that is convex (“jointly”) in all variables x .

Several approaches have been made in order to solve GNEPs. The possibility of formulating a GNEP as a quasi-variational inequality problem (QVI) was pointed out in [5, 42]. Having continuously differentiable cost functions the QVI is given by: Find a $\bar{x} \in \Omega(\bar{x})$ such that

$$\begin{pmatrix} \nabla_{x^1} \theta_1(\bar{x}^1, \bar{x}^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(\bar{x}^N, \bar{x}^{-N}) \end{pmatrix}^T (y - \bar{x}) \geq 0 \quad \text{for all } y \in \Omega(\bar{x}).$$

In the special case of a NEP the QVI reduces to a simpler variational inequality problem (VI), where the feasible set does no longer depend on x . There are theoretical results and algorithms for the solution of a VI, see [31]. It is well known, see for example [25, 27], that in the jointly convex case every solution of the variational inequality problem, find a $\bar{x} \in X$ such that

$$\begin{pmatrix} \nabla_{x^1} \theta_1(\bar{x}^1, \bar{x}^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(\bar{x}^N, \bar{x}^{-N}) \end{pmatrix}^T (y - \bar{x}) \geq 0 \quad \text{for all } y \in X,$$

is also a solution of the GNEP, while the converse is not true in general. The solutions of the variational inequality problem are called *variational equilibria* or *normalized Nash equilibria*. Note, however, that the name normalized Nash equilibrium will be used in a

different context here, see Definition 1.3 below, and that its original definition was given by Rosen [68] in the context of the solution of Karush-Kuhn-Tucker (KKT) conditions. For jointly convex GNEPs given in the form

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t.} \quad g(x^\nu, x^{-\nu}) \leq 0$$

for all $\nu = 1, \dots, N$ with continuously differentiable functions $\theta_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, these KKT conditions claim for each solution \bar{x} the existence of multipliers $\lambda^\nu \in \mathbb{R}^m$ such that

$$\begin{aligned} \nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) + \nabla_{x^\nu} g(\bar{x}^\nu, \bar{x}^{-\nu}) \lambda^\nu &= 0, \\ g(\bar{x}^\nu, \bar{x}^{-\nu}) &\leq 0, \quad \lambda^\nu \geq 0, \quad (\lambda^\nu)^T g(\bar{x}^\nu, \bar{x}^{-\nu}) = 0 \end{aligned}$$

holds for all $\nu = 1, \dots, N$. Following the definition from [68] a *normalized Nash equilibrium* is a solution of the KKT conditions of a GNEP where $\lambda^1 = \dots = \lambda^N$ holds.

In order to define normalized Nash equilibria for jointly convex GNEPs in the absence of differentiability, and to introduce the definition that will be used in this thesis, consider the *Nikaido-Isoda function* (also called *Ky-Fan function*) defined by

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

cf. [59]. Using this function we get the following definition that corresponds to the one given in, e.g., [37, 72].

Definition 1.3 *A vector $\bar{x} \in X$ is called a normalized Nash equilibrium of a jointly convex GNEP, if*

$$\sup_{y \in X} \Psi(\bar{x}, y) \leq 0$$

holds.

Note that many algorithms that are used to solve jointly convex GNEPs search for these normalized Nash equilibria, since they are of special interest in some applications, see [42]. Since θ_ν is convex in x^ν by Assumption 1.2, it is easy to see that $\Psi(x, \cdot)$ is concave for any fixed x . Hence the *regularized Nikaido-Isoda-function*, cf. [41],

$$\Psi_\alpha(x, y) := \sum_{\nu=1}^N \left[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu}) - \frac{\alpha}{2} \|x^\nu - y^\nu\|^2 \right], \quad (1.3)$$

is uniformly concave as a function of the second argument, where $\alpha > 0$ denotes a fixed parameter. This property implies that the maximization problem $\max_{y \in X} \Psi_\alpha(x, y)$ has a unique solution for any nonempty, closed, and convex set $X \subseteq \mathbb{R}^n$, which can be exploited to get optimization reformulations or fixed-point characterizations, as it was done in [45].

1.3. Existing Literature

The roots of equilibrium concepts are in oligopolistic economies considered by A.-A. Cournot in 1838, see [14]. John F. Nash formally introduced the Nash equilibrium problem in [57, 58] in 1950/51. Soon after Nash's work K.J. Arrow and G. Debreu introduced the generalized Nash equilibrium concept in [1], where the term social equilibrium was used. Some existence and uniqueness results were proven by J.B. Rosen in his paper [68] in 1965 and he introduced normalized solutions. With the beginning of the 1990s the interest on GNEPs outside the economic field grew and a number of equivalent reformulations along with algorithms for its solution were developed. A detailed description of the historical development is given in the survey paper [27], which is also a good starting point for research on GNEPs.

There are several ways to reformulate GNEPs, or some subclasses of them, as optimization problems, fixed-point problems or (quasi-) variational inequalities. Many algorithms that are used to solve the resulting problems only find a special solution, whereas there are often multiple (typically infinitely many) solutions of the GNEP. Therefore, beside algorithms that can find one special solution, also methods characterizing the entire solution set are of interest.

There are some approaches for the solution of the general (not jointly convex) GNEP designed for special instances, like generalized potential games considered in [34], or for particular application with special structure, like the decomposition method used in [62]. For the general not jointly convex problem the existing approaches may be divided into three groups:

1. Methods based on quasi-variational inequality formulations;
2. Penalty-type methods;
3. Methods using the concatenated KKT system of all players.

While variational inequality problems, which occur for NEPs, are quite well understood from both the theoretical and numerical point of view, see [31], there are no efficient methods for the quasi-variational inequality problem available yet. A good overview on existing results in the VI approach for GNEPs, including existence and uniqueness results and some iterative algorithms, can be found in [33]. Some recent ideas in the algorithmic area are the use of gap functions as in the case of variational inequalities, see [39, 50, 70], a parameterized variational inequality approach, see [56], some projection-like methods, see [73], or to solve a sequence of penalized variational inequalities, see [61]. The last approach can also be seen as a penalty-type method and therefore may be put into the second group.

This group, the penalty-type methods, add either all or at least the difficult (shared) constraints to the players' cost function in order to obtain a simpler standard NEP as a subproblem at each iteration, see [28, 29, 30, 32, 38]. These penalty methods are globally convergent under suitable assumptions. The subproblems may be difficult to solve in practice, and the rate of convergence is typically slow. Nevertheless, extensive numerical results are available for at least one of these methods, see [28], and indicate that the method is working quite well.

The third group writes down the KKT system for all players and solves the resulting system either by using a mixed complementarity/variational inequality approach, see [60], or by applying a Newton-type method to a suitable reformulation of the system as a nonsmooth system of equations, see [26]. Recently the use of Fritz-John points instead of the KKT system was suggested in [18], and a nonsmooth projection method for its solution was proposed. The main focus of this group of methods is on the local rate of convergence which, however, is difficult since there is an inherent singularity problem as soon as there is at least one joint constraint shared by two or more players.

All the methods designed for the general case can in particular be used to solve jointly convex GNEPs. But there are also some further methods that are developed only for jointly convex problems, in particular the methods from [44, 45, 46, 47, 51, 53, 72]. In the jointly convex context the (regularized) Nikaido-Isoda function has become a useful tool and it will be shown that this tool can also be used for the general case.

1.4. Summary

The intention of this thesis is to contribute new algorithms with the focus on their global convergence properties. Since there are many small examples for the illustration of the assumptions and theoretical results, the presentation of some useful theorems for the analytical or graphical solution of small GNEPs is at the beginning in Section 2, before we concentrate on the new algorithms in the following sections.

The first algorithm to be developed in Section 3 is the only considered algorithm that deals with the jointly convex case. It is a Newton method and a globalization of the existing method given in [47]. For this algorithm global convergence to a normalized solution and, as one might expect for a Newton method, fast local convergence can be shown. This makes the method special since to the author's knowledge none of the existing methods has been shown to enjoy both local fast and global convergence properties, in particular not the methods mentioned above. The new method was published in [20], together with A. von Heusinger, C. Kanzow and M. Fukushima.

The remaining theory sections consider the general (not jointly convex) case. In Section 4 constrained and unconstrained optimization reformulations of the GNEP using the regularized Nikaido-Isoda function are developed, and the methods and results from [45] are generalized and extended. This nonsmooth optimization approach does not belong to any of the three groups of approaches stated in the previous section. Together with an appropriate algorithm the presented reformulations can be used to compute different solutions spreading over the entire solution set. The results of this section were published first for the jointly convex case together with C. Kanzow in [21] and then for the general player convex case together with C. Kanzow and O. Stein in [22].

In Section 5 there are two approaches that belong to the third group of approaches from the previous section, and there are two rather distinct classes of algorithms for the solution of the GNEP, one using a merit function for the KKT system and the other using an interior point approach. For all algorithms developed here theoretical conditions

guaranteeing global convergence are stated. In particular for the interior point method for GNEPs good theoretical convergence results can be shown and it turns out to be very competitive to existing algorithms from the practical point of view. All the results of this section, except the finite termination property for the interior point algorithm, were published together with F. Facchinei, C. Kanzow, and S. Sagratella in [19].

In Section 6 all new methods are tested numerically on a large test set of GNEPs. The globalized Newton method from Section 3 is compared to the existing local variant with different parameters. An appropriate algorithm solving the unconstrained reformulation from Section 4 is proposed and the advantages and disadvantages of the reformulation are discussed. Moreover the algorithms from Section 5 that are solving the KKT system of the GNEP are compared to each other and to the existing penalty method from [28], showing their ability.

To conclude the thesis a short summary of the results together with some hints on future research topics is given in the Final Remarks section.

Finally, the detailed numerical results are given in the appendix.

2. Analytical Solution of Small GNEPs

It is often useful to give small examples to illustrate the content of assumptions or theorems. In many cases the solution sets of the GNEPs have to be computed for the discussion of the examples. This section presents two theorems dealing with small (each player has exactly one variable) player convex GNEPs, that can be used to calculate the solution sets of these examples analytically, or for two players graphically, by using Euclidean projections. We will use the domain of the point-to-set mapping $x^{-\nu} \mapsto X_\nu(x^{-\nu})$, given by

$$\text{dom}(X_\nu) = \{x^{-\nu} \in \mathbb{R}^{n-n_\nu} \mid X_\nu(x^{-\nu}) \neq \emptyset\}.$$

Theorem 2.1 *Consider a player convex GNEP, where each player has exactly one variable. Suppose that for all $\nu = 1, \dots, N$ the function $\theta_\nu(\cdot, x^{-\nu})$ has a unique unconstrained global minimum for all $x^{-\nu} \in \text{dom}(X_\nu)$. Then the solution set of the GNEP is given by*

$$\bigcap_{\nu=1}^N \left\{ (P_{X_\nu(x^{-\nu})}[z^\nu], x^{-\nu}) \mid x^{-\nu} \in \text{dom}(X_\nu), \theta_\nu(z^\nu, x^{-\nu}) \leq \theta_\nu(x^\nu, x^{-\nu}) \forall x^\nu \in \mathbb{R} \right\}.$$

Proof. For a given $\nu \in \{1, \dots, N\}$ and a given $\bar{x}^{-\nu} \in \text{dom}(X_\nu)$ let $\bar{z}^\nu \in \mathbb{R}$ be the unique unconstrained global minimum of the function $\theta_\nu(\cdot, \bar{x}^{-\nu})$, that is

$$\theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) < \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{for all } x^\nu \in \mathbb{R} \setminus \{\bar{z}^\nu\}.$$

Let us first show an inequality that will play a central role in the proof. The convexity of the function $\theta_\nu(\cdot, x^{-\nu})$ implies for all $x^\nu \neq \bar{z}^\nu$ and all $\lambda \in [0, 1]$

$$\begin{aligned} \theta_\nu(\lambda \bar{z}^\nu + (1 - \lambda)x^\nu, \bar{x}^{-\nu}) &\leq \lambda \theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) + (1 - \lambda) \theta_\nu(x^\nu, \bar{x}^{-\nu}), \\ \iff \theta_\nu(\lambda \bar{z}^\nu + (1 - \lambda)x^\nu, \bar{x}^{-\nu}) &+ \underbrace{\lambda (\theta_\nu(x^\nu, \bar{x}^{-\nu}) - \theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}))}_{>0} \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}), \end{aligned}$$

and thus we get the inequality

$$\theta_\nu(\lambda \bar{z}^\nu + (1 - \lambda)x^\nu, \bar{x}^{-\nu}) < \theta_\nu(x^\nu, \bar{x}^{-\nu}) \tag{2.1}$$

for all $\lambda \in (0, 1]$ and all $x^\nu \neq \bar{z}^\nu$.

Now suppose \bar{x} is a solution of the GNEP, i.e., \bar{x}^ν solves the problem

$$\min_{x^\nu} \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{s.t. } x^\nu \in X_\nu(\bar{x}^{-\nu}) \tag{2.2}$$

for all $\nu = 1, \dots, N$. We claim that this is equivalent to

$$[\bar{x}^\nu, \bar{z}^\nu] \cap X_\nu(\bar{x}^{-\nu}) = \{\bar{x}^\nu\} \quad \text{for all } \nu = 1, \dots, N, \quad (2.3)$$

which can be seen as follows: On the one hand, if \bar{x} is a solution, we get for all $\nu = 1, \dots, N$ that $\bar{x}^\nu \in X_\nu(\bar{x}^{-\nu})$, and no interior point of the line segment $[\bar{x}^\nu, \bar{z}^\nu]$ can be feasible by (2.1), hence (2.3) holds. On the other hand, let (2.3) hold, and consider a fixed $\nu = 1, \dots, N$ and an arbitrary feasible point $x^\nu \in X_\nu(\bar{x}^{-\nu}) \setminus \{\bar{x}^\nu\}$. By the convexity of $X_\nu(\bar{x}^{-\nu})$ we obtain $[x^\nu, \bar{x}^\nu] \in X_\nu(\bar{x}^{-\nu})$, and taking into account (2.3), we deduce $\bar{x}^\nu \in (x^\nu, \bar{z}^\nu]$. Now we can find a $\lambda \in (0, 1]$ such that $\bar{x}^\nu = \lambda \bar{z}^\nu + (1 - \lambda)x^\nu$, and, by (2.1), \bar{x}^ν is a solution of the problem (2.2). This shows that \bar{x} being a solution of the GNEP is equivalent to (2.3).

Since each player has only one variable and $X_\nu(\bar{x}^{-\nu})$ is convex, condition (2.3) is equivalent to

$$\begin{aligned} \bar{x}^\nu &= P_{X_\nu(\bar{x}^{-\nu})}[\bar{z}^\nu] && \text{for all } \nu = 1, \dots, N, \\ \iff (\bar{x}^\nu, \bar{x}^{-\nu}) &= (P_{X_\nu(\bar{x}^{-\nu})}[\bar{z}^\nu], \bar{x}^{-\nu}) && \text{for all } \nu = 1, \dots, N, \end{aligned}$$

which in turn is equivalent to \bar{x} being an element of

$$\bigcap_{\nu=1}^N \left\{ (P_{X_\nu(x^{-\nu})}[z^\nu], x^{-\nu}) \mid x^{-\nu} \in \text{dom}(X_\nu), \theta_\nu(z^\nu, x^{-\nu}) \leq \theta_\nu(x^\nu, x^{-\nu}) \forall x^\nu \in \mathbb{R} \right\}.$$

□

If the global minima of the cost functions can be computed, Theorem 2.1 can be used to solve 2-player games graphically. We give a simple example here.

Example 2.2 Consider the jointly convex 2-player game defined via

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1 + 2)^2 \quad \text{s.t. } 0 \leq x^1 \leq 2, x^2 \leq x^1, x^1 - x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2 + 2)^2 \quad \text{s.t. } 0 \leq x^1 \leq 2, x^2 \leq x^1, x^1 - x^2 \leq 1. \end{aligned}$$

Convexity and differentiability of the cost functions together with Theorem 2.1 imply that the solution set is given by

$$\begin{aligned} & \bigcap_{\nu=1}^N \left\{ (P_{X_\nu(x^{-\nu})}[z^\nu], x^{-\nu}) \mid x^{-\nu} \in \text{dom}(X_\nu), \theta_\nu(z^\nu, x^{-\nu}) \leq \theta_\nu(x^\nu, x^{-\nu}) \forall x^\nu \in \mathbb{R} \right\} \\ &= \bigcap_{\nu=1}^N \left\{ (P_{X_\nu(x^{-\nu})}[x^\nu], x^{-\nu}) \mid x^{-\nu} \in \text{dom}(X_\nu), \nabla_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) = 0 \right\}. \end{aligned}$$

Hence the only generalized Nash equilibrium is $(0, -1)$, see Figure 2.1. ◇

For the next result we need continuously differentiable functions. Define the set

$$W := \{x \in \mathbb{R}^N \mid g^\nu(x) \leq 0 \text{ for all } \nu = 1, \dots, N\},$$

and let the following assumption on the strategy sets hold:

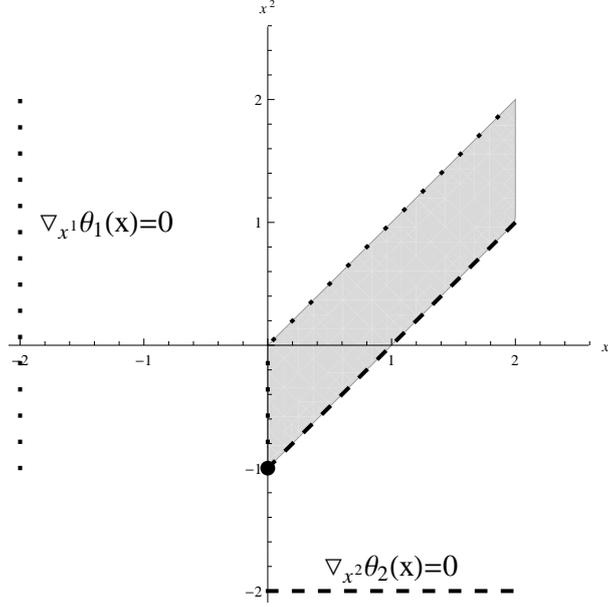


Figure 2.1.: Graphical solution of Example 2.2

Assumption 2.3 All constraint functions $g_i^\nu, i = 1, \dots, m_\nu, \nu = 1, \dots, N$ are continuously differentiable. For all $x \in W$ where $g_i^\nu(x) = 0$ and $\nabla_{x^\nu} g_i^\nu(x^\nu, x^{-\nu}) = 0$ for a player $\nu = 1, \dots, N$ and an index $i \in \{1, \dots, m_\nu\}$ the set $X_\nu(x^{-\nu})$ is single valued.

Remark 2.4 (a) Consider GNEPs with linear constraints only. Then the condition $\nabla_{x^\nu} g_i^\nu(x^\nu, x^{-\nu}) = 0$ implies that g_i^ν is independent of x^ν , hence this constraint can be dropped for the ν -th player. If this is done for all players and all constraints, the resulting GNEP is equivalent to the original one and it does satisfy Assumption 2.3 now.

(b) Consider a jointly convex 2-player GNEP with a single constraint

$$g(x) := (x^1)^2 + (x^2)^2 \leq 1.$$

Then we have $g(x) = 0$ and $\nabla_{x^\nu} g(x) = 2x^\nu = 0$, if and only if $x^\nu = 0$ and $x^{-\nu} = \pm 1$. Since there $X_\nu(x^{-\nu}) = \{0\}$ is single valued, Assumption 2.3 is satisfied also in this example.

(c) There are also GNEPs where Assumption 2.3 is not satisfied, for example consider a jointly convex 2-player GNEP, where the common strategy set is defined by

$$X = \{x \in [0, 1]^2 \mid g(x) := (x^2 - x^1)^2 - (x^1)^2 \leq 0\}.$$

Then Assumption 2.3 is violated for player 1 at the point $\bar{x} = (\bar{x}^1, 0)$, since $g(\bar{x}) = 0$, $\nabla_{x^1} g(\bar{x}) = 0$ independent of \bar{x}^1 , and

$$X_1(0) = \{x^1 \in [0, 1] \mid (0 - x^1)^2 - (x^1)^2 \leq 0\} = [0, 1]$$

2. Analytical Solution of Small GNEPs

is not single valued.

For all GNEPs, where each player has exactly one variable and that satisfy the above assumption, one can show the following theorem.

Theorem 2.5 *Consider a player convex GNEP, where each player has exactly one variable and Assumption 2.3 is satisfied. Suppose that for all $\nu = 1, \dots, N$ the function $\theta_\nu(\cdot, x^{-\nu})$ is continuously differentiable and a unique unconstrained global minimum of the function $\theta_\nu(\cdot, x^{-\nu})$ exists for all $x^{-\nu} \in \mathbb{R}^{N-1}$. Then the solution set is given by*

$$S := \bigcap_{\nu=1}^N \left(\left(\bigcup_{i=1}^{m_\nu} \{x \in W \mid g_i^\nu(x) = 0, \nabla_{x^\nu} \theta_\nu(x) \nabla_{x^\nu} g_i^\nu(x) \leq 0\} \right) \cup \{x \in W \mid \nabla_{x^\nu} \theta_\nu(x) = 0\} \right).$$

Proof. Suppose \bar{x} is a solution of the GNEP, that is \bar{x} solves the problem

$$\min_{x^\nu} \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{s.t.} \quad g^\nu(x^\nu, \bar{x}^{-\nu}) \leq 0 \quad (2.4)$$

for all $\nu = 1, \dots, N$. Then we have in particular $\bar{x} \in W$. Fix an arbitrary player $\nu \in \{1, \dots, N\}$ and denote by \bar{z}^ν the unconstrained minimum of the convex and differentiable function $\theta_\nu(\cdot, \bar{x}^{-\nu})$, which implies $\nabla_{x^\nu} \theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) = 0$. Then we either have $\bar{x}^\nu = \bar{z}^\nu$ and hence $\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0$, or $\bar{x}^\nu \neq \bar{z}^\nu$, in which case we can use the uniqueness of the unconstrained minimum \bar{z}^ν and the convexity of $\theta_\nu(\cdot, \bar{x}^{-\nu})$ to obtain equation (2.1) for \bar{x}^ν , i.e.,

$$\theta_\nu(\bar{x}^\nu + \lambda(\bar{z}^\nu - \bar{x}^\nu), \bar{x}^{-\nu}) < \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu})$$

for all $\lambda \in (0, 1]$. Then, since \bar{x} is a solution and the constraint functions are continuous, there exists an $i \in \{1, \dots, m_\nu\}$ such that

$$g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0 \quad \text{and} \quad g_i^\nu(\bar{x}^\nu + \lambda(\bar{z}^\nu - \bar{x}^\nu), \bar{x}^{-\nu}) > 0$$

for all $\lambda \in (0, 1]$. Thus we have

$$\nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu})(\bar{z}^\nu - \bar{x}^\nu) = \lim_{\lambda \downarrow 0} \frac{g_i^\nu(\bar{x}^\nu + \lambda(\bar{z}^\nu - \bar{x}^\nu), \bar{x}^{-\nu}) - g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu})}{\lambda} \geq 0.$$

From convexity and the minimum property of \bar{z}^ν we have

$$\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu})(\bar{z}^\nu - \bar{x}^\nu) \leq \theta_\nu(\bar{z}^\nu, \bar{x}^{-\nu}) - \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) < 0.$$

The above inequalities and the fact that each player has one variable imply

$$\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq 0.$$

Altogether we obtain

$$\bar{x} \in \bigcup_{i=1}^{m_\nu} \{x \in W \mid g_i^\nu(x) = 0, \nabla_{x^\nu} \theta_\nu(x) \nabla_{x^\nu} g_i^\nu(x) \leq 0\} \cup \{x \in W \mid \nabla_{x^\nu} \theta_\nu(x) = 0\}.$$

Since this holds for all $\nu \in \{1, \dots, N\}$ we have shown $\bar{x} \in S$.

Now suppose $\bar{x} \in S$ and consider a fixed $\nu \in \{1, \dots, N\}$. By the definition of S we have $\bar{x} \in W$ and either $\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0$, which implies that \bar{x}^ν is the unconstrained minimum of $\theta_\nu(\cdot, \bar{x}^{-\nu})$ and thus also the solution of problem (2.4), or we have an index $i \in \{1, \dots, m_\nu\}$ such that $g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0$ and $\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq 0$. Here we consider two cases:

First let $\nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0$. Then Assumption 2.3 implies that the set $X_\nu(\bar{x}^{-\nu})$ is single valued and thus \bar{x}^ν is the solution of the problem (2.4), since it is the only feasible point. The second case is $\nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \neq 0$. Then there exists a $\gamma \geq 0$ such that

$$\nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = -\gamma \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}).$$

Convexity of $\theta_\nu(\cdot, \bar{x}^{-\nu})$ and $g_i^\nu(\cdot, \bar{x}^{-\nu})$ imply for all $y^\nu \in \{x^\nu \in \mathbb{R} \mid g^\nu(x^\nu, \bar{x}^{-\nu}) \leq 0\}$

$$\begin{aligned} \theta_\nu(y^\nu, \bar{x}^{-\nu}) - \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) &\geq \nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu})(y^\nu - \bar{x}^\nu) \\ &= -\gamma \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu})(y^\nu - \bar{x}^\nu) \\ &\geq \gamma(g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) - g_i^\nu(y^\nu, \bar{x}^{-\nu})) \\ &= -\gamma g_i^\nu(y^\nu, \bar{x}^{-\nu}) \\ &\geq 0, \end{aligned}$$

and hence \bar{x}^ν is a solution of the problem (2.4).

In both cases \bar{x}^ν is a solution of the problem (2.4) and since this holds for all $\nu = 1, \dots, N$, \bar{x} is a solution of the GNEP, which completes the proof. \square

For some GNEPs it is possible to compute the single sets, in particular to determine the empty sets, in the solution formula from the previous theorem. In case that a constraint qualification is not satisfied for a GNEP, implying that the GNEP is not equivalent to the KKT conditions, Theorem 2.5 can still be used. One academic example is the following.

Example 2.6 Consider a 3-player game, where each player controls a single variable.

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1 - 1)^2 \quad \text{s.t. } (x^1)^2 - (x^2)^2 \leq 0, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^1 - x^2)^2 \quad \text{s.t. } (x^1)^2 + (x^2)^2 \leq 1, \\ \text{Player 3: } & \min_{x^3} \frac{1}{2}(x^3 - 3)^2 \quad \text{s.t. } x^3 \geq 0, x^1 + x^2 + x^3 \leq 2. \end{aligned}$$

We want to compute the solutions of this GNEP, using Theorem 2.5. First of all $x \in W$ has to hold and thus by the constraints of the first two players $(x^1)^2 \leq (x^2)^2 \leq 1 - (x^1)^2$, implying $x^1 \leq \frac{\sqrt{2}}{2}$. Therefore $\nabla_{x^1} \theta_1(x) = x^1 - 1 < 0$ and Theorem 2.5 implies $(x^1)^2 = (x^2)^2$ and $\nabla_{x^1} g^1(x) = 2x^1 \geq 0$, thus $x^1 \in \left[0, \frac{\sqrt{2}}{2}\right]$.

If $x^1 \neq x^2$ we can not have a solution, since $\nabla_{x^2} \theta_2(x) = x^2 - x^1 < 0$ and $\nabla_{x^2} g^2(x) = 2x^2 < 0$. Therefore $x^1 = x^2 \in \left[0, \frac{\sqrt{2}}{2}\right]$ and further $\nabla_{x^2} \theta_2(x) = 0$ must hold at a solution.

2. Analytical Solution of Small GNEPs

Using this we get for the third player $0 \leq x^3 \leq 2$ and thus $\nabla_{x^3} \theta_3(x) = x^3 - 3 < 0$. By Theorem 2.5 this means that the constraint $x^1 + x^2 + x^3 \leq 2$ has to be satisfied with equality. This yields the solution set

$$S = \left\{ (\lambda, \lambda, 2 - 2\lambda) \mid \lambda \in \left[0, \frac{\sqrt{2}}{2} \right] \right\}.$$

Note that in this example the solution $(0, 0, 2)$ is not a KKT point. ◇

3. A Globalized Newton Method for Normalized Nash Equilibria

The purpose of this section is to propose a globally convergent algorithm for the computation of a normalized Nash equilibrium for a jointly convex GNEP. The algorithm developed here is of particular interest since it also has fast local convergence properties. The method is based on the characterizations of normalized Nash equilibria given in [45, 47]. More precisely, the algorithm consists of a combination of the locally superlinearly convergent Newton method from [47] and a globally convergent gradient method with a continuously differentiable merit function. It draws on the features of the regularized Nikaido-Isoda function, which has earlier been used as a tool to derive globally convergent methods for the computation of normalized Nash equilibria, see for example [46]. In the following sections we first provide some known results before we state the algorithm and prove its convergence properties. The stated results have been published in [20].

3.1. Preliminaries

The algorithm to be developed is suited for jointly convex GNEPs in the following setting.

Assumption 3.1 (i) *The cost functions $\theta_\nu, \nu = 1, \dots, N$, are twice continuously differentiable and, as a function of x^ν alone, convex.*

(ii) *The joint strategy set $X \subseteq \mathbb{R}^n$ is nonempty and defined by*

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

where $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a twice continuously differentiable function with convex component functions.

In contrast to Assumption 1.2 we further assume smoothness properties for the cost functions and constraints which are necessary to design a locally superlinearly convergent method and we restrict to jointly convex problems.

Assumption 3.1 implies that the regularized Nikaido-Isoda function

$$\Psi_\gamma(x, y) := \sum_{\nu=1}^N \left[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu}) - \frac{\gamma}{2} \|x^\nu - y^\nu\|^2 \right],$$

3. A Globalized Newton Method for Normalized Nash Equilibria

where $\gamma > 0$ is a fixed parameter, is twice continuously differentiable and strongly concave with respect to the second argument, and further that the feasible set X is nonempty, closed and convex. Therefore the optimization problem

$$\max_{y \in X} \Psi_\gamma(x, y) \quad (3.1)$$

has a unique solution. However, the solution function defined by

$$y_\gamma(x) := \operatorname{argmax}_{y \in X} \Psi_\gamma(x, y) \quad (3.2)$$

is, in general, not differentiable everywhere, see [45, Example 2.4]. Further define the function

$$V_\gamma(x) := \Psi_\gamma(x, y_\gamma(x)), \quad (3.3)$$

and, in particular, for two different fixed parameters $\beta > \alpha > 0$ consider the function built from the difference of two functions V_α and V_β ,

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

The functions y_γ and $V_{\alpha\beta}$ have a number of interesting properties that are summarized in the following result, whose proof can be found in [45].

Theorem 3.2 *Suppose that Assumption 3.1 holds and let $\gamma > 0$, $\beta > \alpha > 0$ be fixed parameters. Then:*

- (i) *the functions y_γ , V_γ and $V_{\alpha\beta}$ are continuous;*
- (ii) *any vector $\bar{x} \in \mathbb{R}^n$ with $y_\gamma(\bar{x}) = \bar{x}$ is a normalized Nash equilibrium;*
- (iii) *$V_{\alpha\beta}(x) \geq 0$ for all $x \in \mathbb{R}^n$;*
- (iv) *any vector $\bar{x} \in \mathbb{R}^n$ with $V_{\alpha\beta}(\bar{x}) = 0$ is a normalized Nash equilibrium;*
- (v) *the function $V_{\alpha\beta}$ is continuously differentiable and its gradient is given by*

$$\begin{aligned} \nabla V_{\alpha\beta}(x) &= \sum_{\nu=1}^N [\nabla \theta_\nu(y_\beta^\nu(x), x^{-\nu}) - \nabla \theta_\nu(y_\alpha^\nu(x), x^{-\nu})] \\ &\quad + \begin{pmatrix} \nabla_{x^1} \theta_1(y_\alpha^1(x), x^{-1}) - \nabla_{x^1} \theta_1(y_\beta^1(x), x^{-1}) \\ \vdots \\ \nabla_{x^N} \theta_N(y_\alpha^N(x), x^{-N}) - \nabla_{x^N} \theta_N(y_\beta^N(x), x^{-N}) \end{pmatrix} \\ &\quad - \alpha(x - y_\alpha(x)) + \beta(x - y_\beta(x)). \end{aligned}$$

The statements (ii) and (iv) do not yield a characterization of the full set of generalized Nash equilibria, but a characterization of the set of normalized Nash equilibria. Let $\gamma = \beta > \alpha > 0$. Statement (ii) says that a normalized Nash equilibrium is a solution of the nonlinear equation

$$F_\beta(x) := y_\beta(x) - x = 0, \quad (3.4)$$

which is nonsmooth in general. Furthermore, statements (iii) and (iv) imply that \bar{x} is a normalized Nash equilibrium if and only if it is a global solution with vanishing function value of the optimization problem

$$\min_{x \in \mathbb{R}^n} V_{\alpha\beta}(x), \quad (3.5)$$

where the objective function $V_{\alpha\beta}$ is continuously differentiable from (v). We combine both characterizations to design a globally, and locally fast, convergent method for the computation of a normalized Nash equilibrium. Specifically, we try to use a nonsmooth Newton direction for the nonlinear equation (3.4) as long as it decreases the merit function $V_{\alpha\beta}$, but we switch to the anti-gradient direction of the function $V_{\alpha\beta}$ if it fails to be a descent direction. The former aims at fast local convergence, while the latter ensures global convergence of the algorithm.

To validate the proposed algorithm, we should find a condition that ensures that a stationary point of the function $V_{\alpha\beta}$ is a solution of problem (3.5), and hence, a normalized Nash equilibrium. Further, for the nonsmooth Newton method, we have to establish conditions which ensure that the function F_β is at least Lipschitz-continuous, and calculate a suitable approximation for the Jacobian of $F_\beta(x)$, in order to make a nonsmooth Newton method conveniently implementable. These questions have been answered in [45, 47]. In particular, the proof of the next lemma can be found in [45, Theorem 4.6].

Lemma 3.3 *Assume that, in addition to Assumption 3.1, for any $x \in \mathbb{R}^n$ with $y_\alpha(x) \neq y_\beta(x)$ the inequality*

$$\sum_{\nu=1}^N [\nabla\theta_\nu(y_\beta^\nu(x), x^{-\nu}) - \nabla\theta_\nu(y_\alpha^\nu(x), x^{-\nu})]^T (y_\beta(x) - y_\alpha(x)) > 0 \quad (3.6)$$

holds. Then any stationary point of the function $V_{\alpha\beta}$ is a normalized Nash equilibrium.

Sufficient conditions for the condition (3.6) in Lemma 3.3 and a further discussion may be found in [45, 46] and are somewhat related to the corresponding conditions given in [37, 72], for example. Moreover, there is an inequality relation among the values of $V_{\alpha\beta}(x)$, $\|y_\alpha(x) - x\|$, and $\|y_\beta(x) - x\|$, which can be found in [45, Lemma 4.1].

Lemma 3.4 *The inequalities*

$$\frac{\beta - \alpha}{2} \|x - y_\beta(x)\|^2 \leq V_{\alpha\beta}(x) \leq \frac{\beta - \alpha}{2} \|x - y_\alpha(x)\|^2 \quad (3.7)$$

hold for all $x \in \mathbb{R}^n$.

In our method, this growth condition on the merit function $V_{\alpha\beta}$ is a necessary tool for the transition from the globally convergent gradient method based on $V_{\alpha\beta}$ to the locally convergent Newton method based on $F_\beta(x) = y_\beta(x) - x$. Such a technique has been applied to the numerical solution of nonlinear complementarity problems, see for instance [36] and [16].

For the nonsmooth Newton method, we use the very same approach as in [47]. That is, we compute an element of the *computable generalized Jacobian* of $F_\beta(x)$. By definition (3.2), to calculate the value of $y_\beta(x)$, we have to solve the maximization problem (3.1). By Theorem 3.2 (i), $y_\beta(\cdot)$ is a continuous function. To get further insight into the analytic properties of the function $y_\beta(\cdot)$, we formulate the KKT conditions for problem (3.1). To this end, we need a constraint qualification. For our purposes, the constant rank constraint qualification (CRCQ), which is weaker than the linear independence constraint qualification (LICQ), see [49], suffices. For each $x \in X$, let

$$I(x) := \{i \in \{1, \dots, m\} \mid g_i(x) = 0\}$$

denote the index set of active constraints at x .

Assumption 3.5 *The constant rank constraint qualification (CRCQ) holds at $x^* \in \mathbb{R}^n$ if there exists a neighbourhood N of x^* such that, for every subset $J \subseteq I(x^*)$, the set of gradient vectors*

$$\{\nabla g_i(x) \mid i \in J\}$$

has the same rank (depending on J) for all $x \in N$.

Note, in particular, that linear constraints satisfy the CRCQ. For $x \in \mathbb{R}^n$ and the unique solution $y_\beta(x)$ of problem (3.1), Assumption 3.5 at $y_\beta(x)$ guarantees the existence of a Lagrange multiplier $\lambda \in \mathbb{R}^m$ such that the KKT conditions hold, see [49], i.e.,

$$\begin{aligned} -\nabla_y \Psi_\beta(x, y_\beta(x)) + \sum_{i=1}^m \lambda_i \nabla g_i(y_\beta(x)) &= 0, \\ \lambda_i \geq 0, \quad g_i(y_\beta(x)) \leq 0, \quad \lambda_i g_i(y_\beta(x)) &= 0 \quad \forall i = 1, \dots, m. \end{aligned} \tag{3.8}$$

Since we assume only the constant rank constraint qualification to hold, the Lagrange multiplier is not necessarily unique. Let

$$\mathcal{M}(x) := \{\lambda \in \mathbb{R}^m \mid (x, y_\beta(x), \lambda) \text{ satisfies (3.8)}\} \tag{3.9}$$

denote the set of Lagrange multipliers. We define a family of certain subsets of the active index set $I(x)$ by

$$\mathcal{B}(x) := \left\{ J \subseteq I(x) \mid \exists \lambda \in \mathcal{M}(x) \text{ such that } \lambda_i = 0 \text{ for all } i \in I(x) \setminus J, \right. \\ \left. \text{and } \{\nabla g_i(x)\}_{i \in J} \text{ are linearly independent} \right\}. \tag{3.10}$$

The next theorem is concerned with a formula for the computable generalized Jacobian of F_β . For details regarding its derivation see [47, Lemmata 3.4 and 3.5].

Lemma 3.6 *Assume that Assumption 3.1 is satisfied and that the CRCQ Assumption 3.5 holds at $y_\beta(x)$. Then the function $F_\beta(x) := y_\beta(x) - x$ is a PC^1 function near x , and the computable generalized Jacobian of F_β at x is defined by*

$$\partial_C F_\beta(x) := \{ \nabla y_\beta^J(x)^T - I_n \mid J \in \mathcal{B}(x) \},$$

where

$$\nabla y_\beta^J(x)^T = C^{-1}A - C^{-1}D(D^T C^{-1}D)^{-1}D^T C^{-1}A$$

with

$$\begin{aligned} A &:= A(x) &:= \nabla_{yx}^2 \Psi_\beta(x, y_\beta(x)), \\ C &:= C^J(x) &:= -\nabla_{yy}^2 \Psi_\beta(x, y_\beta(x)) + \sum_{i \in J} \lambda_i \nabla^2 g_i(y_\beta(x)), \\ D &:= D^J(x) &:= \nabla g_J(y_\beta(x)). \end{aligned}$$

The explicit formulas for $\nabla_{yx}^2 \Psi_\beta(x, y)$ and $\nabla_{yy}^2 \Psi_\beta(x, y)$ are given as follows (cf. [44]):

$$\begin{aligned} \nabla_{yx}^2 \Psi_\beta(x, y) &= - \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & \cdots & \nabla_{x^1 x^N}^2 \theta_1(y^1, x^{-1}) \\ \vdots & & \vdots \\ \nabla_{x^N x^1}^2 \theta_N(y^N, x^{-N}) & \cdots & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} \\ &\quad + \text{blockdiag} \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} + \beta I_n, \\ \nabla_{yy}^2 \Psi_\beta(x, y) &= - \text{blockdiag} \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & & \\ & \ddots & \\ & & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix} - \beta I_n, \end{aligned}$$

where blockdiag means that the matrix has only the given square matrices as blocks on its diagonal and all other elements are zero. Note that the Lagrange multiplier λ used in the definition of C does depend on x , though not explicitly stated.

3.2. A Globalization of a Newton Method via $V_{\alpha\beta}$

Using the notation of the previous section, we are now ready to state the algorithm.

Algorithm 3.7 (*Globalized Newton method via $V_{\alpha\beta}$*)

(S.0) Choose $x^0 \in \mathbb{R}^n$, $\varepsilon \geq 0$, $s > 1$, $\rho > 0$, $\tau \in (0, 1)$, $\sigma \in (0, 1)$, and set $k := 0$.

(S.1) If $\|F_\beta(x^k)\| = \|y_\beta(x^k) - x^k\| \leq \varepsilon$, STOP.

(S.2) Compute an element $H_k \in \partial_C F_\beta(x^k)$, and find a solution $d^k \in \mathbb{R}^n$ of the linear system

$$H_k d^k = -F_\beta(x^k), \quad (3.11)$$

if one exists.

(S.3) If the system (3.11) was solved and if

$$V_{\alpha\beta}(x^k + d^k) \leq \tau V_{\alpha\beta}(x^k), \quad (3.12)$$

set $x^{k+1} := x^k + d^k$, $k \leftarrow k + 1$, and go to (S.1).

(S.4) If the system (3.11) was not solved or if the condition

$$\nabla V_{\alpha\beta}(x^k)^T d^k \leq -\rho \|d^k\|^s \quad (3.13)$$

is not satisfied, set $d^k := -\nabla V_{\alpha\beta}(x^k)$.

(S.5) Compute $t_k := \max\{2^{-l} \mid l = 0, 1, 2, \dots\}$ such that

$$V_{\alpha\beta}(x^k + t_k d^k) \leq V_{\alpha\beta}(x^k) + \sigma t_k \nabla V_{\alpha\beta}(x^k)^T d^k. \quad (3.14)$$

Set $x^{k+1} := x^k + t_k d^k$, $k \leftarrow k + 1$, and go to (S.1).

This algorithm is reminiscent of the algorithm in [40] for solving the mixed complementarity problem, which combines a nonsmooth Newton method for the natural residual equation with a globalization strategy using the D-gap function.

Our analysis of Algorithm 3.7 starts with its accumulation points.

Theorem 3.8 *Suppose Assumption 3.1 holds, and that Assumption 3.5 is satisfied for all $x \in X$. Then Algorithm 3.7 with $\varepsilon = 0$ either stops at a normalized Nash equilibrium or every accumulation point \bar{x} of a sequence generated by the algorithm is either a stationary point of $V_{\alpha\beta}$ or a normalized Nash equilibrium.*

Proof. The proof is similar to those of the corresponding global convergence results in [15, 16].

If the algorithm stops at (S.1), then we have $F_\beta(x^k) = 0$, and x^k is a normalized Nash equilibrium in view of Theorem 3.2 (ii). Otherwise, consider a subsequence $\{x^k\}$ converging to \bar{x} . If for an infinite set of indices in this subsequence, we have $d^k := -\nabla V_{\alpha\beta}(x^k)$, then \bar{x} is a stationary point of $V_{\alpha\beta}$ by standard arguments. If (3.12) holds infinitely often, we get $V_{\alpha\beta}(x^k) \rightarrow 0$ since $V_{\alpha\beta}(x) \geq 0$ and $\tau \in (0, 1)$, implying that \bar{x} is a normalized Nash equilibrium by Theorem 3.2(i),(iv). Therefore we assume, without loss of generality, that the search direction is always obtained from the linear system $H_k d^k = -F_\beta(x^k)$ and that the condition (3.13) is always satisfied.

From (3.13), we have $\nabla V_{\alpha\beta}(x^k)^T d^k \leq -\rho \|d^k\|^s$, which together with $s > 1$ and continuity of $\nabla V_{\alpha\beta}$ implies boundedness of the sequence $\{\|d^k\|\}$. Subsequencing if necessary, we can assume $d^k \rightarrow \bar{d}$.

Assume $\bar{d} \neq 0$. From (3.14), we have $V_{\alpha\beta}(x^k + t_k d^k) \leq V_{\alpha\beta}(x^k)$. Since $V_{\alpha\beta}(x) \geq 0$ for all $x \in \mathbb{R}^n$, it follows that $V_{\alpha\beta}(x^k + t_k d^k) - V_{\alpha\beta}(x^k) \rightarrow 0$, which yields

$$t_k \nabla V_{\alpha\beta}(x^k)^T d^k \rightarrow 0. \quad (3.15)$$

Now suppose $t_k \rightarrow 0$. Notice that the line search rule implies

$$\frac{V_{\alpha\beta}(x^k + 2t_k d^k) - V_{\alpha\beta}(x^k)}{2t_k} > \sigma \nabla V_{\alpha\beta}(x^k)^T d^k$$

for all $k \in \mathbb{N}$ sufficiently large. Passing to the limit and exploiting the continuous differentiability of $V_{\alpha\beta}$, we have

$$\nabla V_{\alpha\beta}(\bar{x})^T \bar{d} \geq \sigma \nabla V_{\alpha\beta}(\bar{x})^T \bar{d}.$$

Since $\sigma \in (0, 1)$, this yields $\nabla V_{\alpha\beta}(\bar{x})^T \bar{d} \geq 0$, contradicting (3.13) since $\bar{d} \neq 0$. Therefore, $\{t_k\}$ must be bounded away from zero. However, (3.15) and (3.13) then imply $d^k \rightarrow 0$, contradicting $d^k \rightarrow \bar{d} \neq 0$.

Therefore we must have $\bar{d} = 0$. Lemma 3.6 implies that F_β is a piecewise continuously differentiable function, and hence, in a neighbourhood of \bar{x} , each H_k is the Jacobian of one of finitely many C^1 functions. Thus, the sequence $\{H_k\}$ is bounded. Since $\|F_\beta(x^k)\| = \|H_k d^k\| \leq \|H_k\| \|d^k\|$ and $d^k \rightarrow \bar{d} = 0$, it follows that $F_\beta(\bar{x}) = 0$. Hence, from Theorem 3.2 (ii), we conclude that \bar{x} is a normalized Nash equilibrium. \square

Remark 3.9 Note that Theorem 3.8 requires Assumption 3.5 to hold at every point $x \in X$. This requirement is unnecessarily strong and is used here only for a simple statement of Algorithm 3.7 (since the computable generalized Jacobian may not exist without the CRCQ). However, the result would remain true if the matrix H_k is alternatively chosen from any set $G(x^k)$, where G is a set-valued mapping which is upper semi-continuous and such that $G(x)$ is a nonempty and compact set for all $x \in \mathbb{R}^n$. A possible candidate for this set-valued mapping is, for example, Clarke's generalized Jacobian, cf. [12]. The reason for using the computable generalized Jacobian from the very beginning is essentially due to the observation that this Jacobian allows us to prove a very nice local convergence result.

Lemma 3.3 shows a condition under which any stationary point of the function $V_{\alpha\beta}$ is a normalized Nash equilibrium.

Next we examine the local convergence of Algorithm 3.7. To this end, we consider the nonsingularity of matrices H_k and acceptance of the full Newton step in (S.3) of Algorithm 3.7. The nonsingularity of matrices H_k in a neighbourhood of a normalized Nash equilibrium is guaranteed by [47, Lemma 4.2] in conjunction with the following assumption. The result in [47, Lemma 4.2] is stated for $x \in X$. However, taking a closer look at the proof of the lemma reveals that the assertion holds for all $x \in \mathbb{R}^n$.

Assumption 3.10 For each $J \in \mathcal{B}(x)$ and $\lambda \in \mathcal{M}(x)$, we have

$$d^T \left(M(x, y_\beta(x)) + \sum_{j \in J} \lambda_j \nabla^2 g_j(y_\beta(x)) \right) d \neq 0 \quad \forall d \in \mathcal{T}^J(x) \setminus \{0\},$$

3. A Globalized Newton Method for Normalized Nash Equilibria

with $\mathcal{T}^J(x) := \{d \in \mathbb{R}^n \mid \nabla g_j(y_\beta(x))^T d = 0 \forall j \in J\}$ and

$$M(x, y) := \begin{pmatrix} \nabla_{x^1 x^1}^2 \theta_1(y^1, x^{-1}) & \dots & \nabla_{x^1 x^N}^2 \theta_1(y^1, x^{-1}) \\ \vdots & & \vdots \\ \nabla_{x^N x^1}^2 \theta_N(y^N, x^{-N}) & \dots & \nabla_{x^N x^N}^2 \theta_N(y^N, x^{-N}) \end{pmatrix}.$$

In addition to the nonsingularity of the matrices H_k , we need a superlinear approximation property to obtain local superlinear convergence. From [47, Theorem 4.5], we immediately obtain the following lemma.

Lemma 3.11 *Let \bar{x} be a normalized Nash equilibrium, suppose Assumptions 3.1, 3.5 and 3.10 hold at \bar{x} , and let $\{x^k\}$ be any sequence converging to \bar{x} . If d^k is a solution of the equation $H_k d^k = -F_\beta(x^k)$, then we have*

$$\|x^k + d^k - \bar{x}\| = o(\|x^k - \bar{x}\|), \quad i.e., \quad \lim_{k \in \mathbb{N}} \frac{\|x^k + d^k - \bar{x}\|}{\|x^k - \bar{x}\|} = 0.$$

Furthermore, if all functions $\theta_\nu, \nu = 1, \dots, N$ and $g_i, i = 1, \dots, m$ have locally Lipschitz continuous second derivatives, then

$$\|x^k + d^k - \bar{x}\| = O(\|x^k - \bar{x}\|^2), \quad i.e., \quad \limsup_{k \in \mathbb{N}} \frac{\|x^k + d^k - \bar{x}\|}{\|x^k - \bar{x}\|^2} < \infty.$$

It remains to show that the full Newton step is eventually accepted in (S.3) of Algorithm 3.7. This is done in a similar way as in [36] and [16] for the nonlinear complementarity problem, using Lemma 3.11 and the growth condition from Lemma 3.4.

Theorem 3.12 *Let \bar{x} be a normalized Nash equilibrium and suppose that \bar{x} is an accumulation point of a sequence $\{x^k\}$ generated by Algorithm 3.7. If Assumptions 3.1, 3.5 and 3.10 hold at \bar{x} , then the entire sequence $\{x^k\}$ converges to \bar{x} . Moreover, eventually the linear system (3.11) is solvable and condition (3.12) is satisfied, and $\{x^k\}$ converges superlinearly to \bar{x} . If, in addition, all functions $\theta_\nu, \nu = 1, \dots, N$ and $g_i, i = 1, \dots, m$ have locally Lipschitz continuous second derivatives, the convergence rate is quadratic.*

Proof. We divide the proof into three steps.

Step 1: We begin with some preliminary observations. From [47, Lemma 4.2], under Assumption 3.10, all matrices $H(\bar{x}) \in \partial_C F_\beta(\bar{x})$ are nonsingular, and so are the matrices $H(x) \in \partial_C F_\beta(x)$ for all x sufficiently close to \bar{x} . Hence, the system (3.11) is solvable for all x near \bar{x} , and

$$\|H(x)(x - \bar{x})\| \geq \frac{\|x - \bar{x}\|}{\|H(x)^{-1}\|} \geq c\|x - \bar{x}\|$$

for some constant $c > 0$ and all x in a sufficiently small neighbourhood of \bar{x} . Moreover, we have

$$\|F_\beta(x) - H(x)(x - \bar{x})\| \leq \frac{c}{2}\|x - \bar{x}\|$$

for all x sufficiently close to \bar{x} , cf. [47, Lemma 4.4]. Therefore we obtain

$$\begin{aligned} c\|x - \bar{x}\| - \|F_\beta(x)\| &\leq \|H(x)(x - \bar{x})\| - \|F_\beta(x)\| \\ &\leq \|F_\beta(x) - H(x)(x - \bar{x})\| \\ &\leq \frac{c}{2}\|x - \bar{x}\|, \end{aligned}$$

that is,

$$\frac{c}{2}\|x - \bar{x}\| \leq \|F_\beta(x)\|, \quad (3.16)$$

provided x is sufficiently close to \bar{x} .

Step 2: Next, we show that the entire sequence $\{x^k\}$ converges to \bar{x} . By [55, Lemma 4.10], it suffices to show that \bar{x} is a locally unique solution of $F_\beta(x) = 0$ and that $\{\|x^{k+1} - x^k\|\}_K$ converges to 0 for any subsequence $\{x^k\}_K$ converging to \bar{x} .

Under our assumptions, the fact that \bar{x} is a locally unique solution follows immediately from (3.16). The updating rules in Algorithm 3.7 imply

$$\|x^{k+1} - x^k\| \leq \|d^k\| \quad \forall k \in K. \quad (3.17)$$

Furthermore, for all $k \in K$ satisfying the test (3.13), the Cauchy-Schwarz inequality gives

$$\rho\|d^k\|^s \leq -\nabla V_{\alpha\beta}(x^k)^T d^k \leq \|\nabla V_{\alpha\beta}(x^k)\| \|d^k\|,$$

which together with $s > 1$ implies

$$\rho\|d^k\|^{s-1} \leq \|\nabla V_{\alpha\beta}(x^k)\|.$$

Then it follows from Theorem 3.8 that $\{\|d^k\|\}_K$ tends to 0. (Recall that $d^k = -\nabla V_{\alpha\beta}(x^k)$ for all $k \in K$ violating (3.13).) Hence the desired result follows from (3.17).

Step 3: Finally, we prove that $\{x^k\}$ converges locally superlinearly/quadratically to \bar{x} . This is done by showing that the globalized Newton method from Algorithm 3.7 eventually coincides with the local Newton method and, therefore, inherits the convergence properties from the local method. To this end, we have to show that the linear system (3.11) is solvable and the corresponding Newton direction d^k satisfies the test (3.12) for all $k \in \mathbb{N}$ sufficiently large.

From Lemma 3.4, we know that

$$\frac{\beta - \alpha}{2}\|F_\beta(x)\|^2 \leq V_{\alpha\beta}(x) \leq \frac{\beta - \alpha}{2}\|F_\alpha(x)\|^2 \quad (3.18)$$

for all $x \in \mathbb{R}^n$. Let $L > 0$ be the local Lipschitz constant of the function F_α around \bar{x} (which exists since F_α is piecewise continuously differentiable due to Lemma 3.6). Then we have

$$\sqrt{V_{\alpha\beta}(x^k + d^k)} \stackrel{(3.18)}{\leq} \sqrt{\frac{\beta - \alpha}{2}} \|F_\alpha(x^k + d^k) - \underbrace{F_\alpha(\bar{x})}_{=0}\|$$

$$\begin{aligned} &\leq \sqrt{\frac{\beta - \alpha}{2}} L \|x^k + d^k - \bar{x}\| \\ \stackrel{\text{Lemma 3.11}}{=} & o(\|x^k - \bar{x}\|). \end{aligned}$$

Therefore, for $k \in \mathbb{N}$ sufficiently large, we have

$$\begin{aligned} \sqrt{V_{\alpha\beta}(x^k + d^k)} &\leq \sqrt{\tau} \sqrt{\frac{\beta - \alpha}{2}} \frac{c}{2} \|x^k - \bar{x}\| \\ &\stackrel{(3.16)}{\leq} \sqrt{\tau} \sqrt{\frac{\beta - \alpha}{2}} \|F_\beta(x^k)\| \\ &\stackrel{(3.18)}{\leq} \sqrt{\tau} \sqrt{V_{\alpha\beta}(x^k)}. \end{aligned}$$

Hence the test (3.12) is eventually successful. By Lemma 3.11, we then have superlinear or quadratic convergence of $\{x^k\}$ to \bar{x} . \square

The above proof shows that the globalized Newton method eventually coincides with the local Newton method. Consequently [47, Proposition 4.6] can be applied to Algorithm 3.7 and we obtain the following finite termination property:

If the assumptions of Theorem 3.12 hold and the generalized Nash equilibrium problem is a quadratic game, i.e., the cost functions θ_ν are quadratic for all players $\nu = 1, \dots, N$ and the strategy set X is polyhedral, and if x^k is sufficiently close to a normalized Nash equilibrium \bar{x} , then the next iterate x^{k+1} coincides with \bar{x} .

4. Optimization Reformulations of GNEPs Using the Nikaido-Isoda Function

In section 3 only jointly convex GNEPs were considered and the method described there is able to find normalized Nash equilibria. Now we will consider the much more general case of player convex GNEPs. The aim of this section is to reformulate the player convex GNEP as an optimization problem. The minima of the resulting optimization problem then coincide with the entire solution set of the underlying GNEP, which allows us to find also non-normalized Nash equilibria. Further we will see that we obtain a nonsmooth objective function in general, but we will establish conditions needed to apply a solver for nonsmooth optimization problems. Using the Nikaido-Isoda function it is shown in [27] that the minima of a constrained “quasi-optimization” problem are exactly the solutions of the GNEP. Further [45] shows the equivalence of the GNEP solutions to the minima of a real constrained optimization problem in the case of jointly convex GNEPs. The constrained reformulation for player convex GNEPs we are going to present in the next section is similar to the results in [45]. Note, however, that the class of jointly convex GNEPs is just a small, but important, subclass of all player convex GNEPs. Having a constrained reformulation we will develop a new unconstrained one, which has almost the same smoothness properties. Since there are efficient algorithms for nonsmooth unconstrained optimization problems the unconstrained reformulation has advantages from a practical point of view, however, we will see that we don’t get a reformulation for the whole class of player convex GNEPs. The results of this section were published first for the jointly convex case in [21] and then for the player convex case in [22].

4.1. Optimization Reformulations

Here we present two reformulations of the GNEP, one as a constrained optimization problem and the other one as an unconstrained optimization problem. Using the regularized Nikaido-Isoda function Ψ_α from (1.3), we define

$$\begin{aligned} V_\alpha(x) &:= \max_{y \in \Omega(x)} \Psi_\alpha(x, y) \\ &= \max_{y \in \Omega(x)} \sum_{\nu=1}^N \left[\theta_\nu(x^\nu, x^{-\nu}) - \theta_\nu(y^\nu, x^{-\nu}) - \frac{\alpha}{2} \|x^\nu - y^\nu\|^2 \right] \end{aligned} \quad (4.1)$$

$$= \sum_{\nu=1}^N \left[\theta_{\nu}(x^{\nu}, x^{-\nu}) - \min_{y^{\nu} \in X_{\nu}(x^{-\nu})} \left(\theta_{\nu}(y^{\nu}, x^{-\nu}) + \frac{\alpha}{2} \|x^{\nu} - y^{\nu}\|^2 \right) \right].$$

In view of Assumption 1.2, $X_{\nu}(x^{-\nu})$, $\nu = 1, \dots, N$ and $\Omega(x)$ are closed and convex sets, and all appearing objective functions are uniformly concave or convex, respectively. Hence the mapping $V_{\alpha}(x)$ is well-defined for all $x \in \mathbb{R}^n$ satisfying $\Omega(x) \neq \emptyset$. In [45] it was shown that finding a solution of the GNEP is equivalent to computing a global minimum of the constrained optimization problem

$$\min V_{\alpha}(x) \quad \text{s.t.} \quad x \in \Omega(x).$$

To handle the constraints it is very natural to define the following set

$$\begin{aligned} W &:= \{x \in \mathbb{R}^n \mid x^{\nu} \in X_{\nu}(x^{-\nu}) \text{ for all } \nu = 1, \dots, N\} \\ &= \{x \in \mathbb{R}^n \mid g^{\nu}(x) \leq 0 \text{ for all } \nu = 1, \dots, N\}, \end{aligned} \quad (4.2)$$

where the second equality follows from the representation of the sets $X_{\nu}(x^{-\nu})$, cf. Assumption 1.2. From the last representation it is clear that W contains all those points which satisfy all the constraints of all players. Some further simple observations are summarized in the following remark.

Remark 4.1 (a) Consider a generalized Nash game with two players having arbitrary cost functions. Player 1 controls the single variable x^1 , and player 2 controls the single variable x^2 . Let the strategy spaces $X_1(x^2)$ and $X_2(x^1)$ be defined by the mappings

$$g^1(x) := (x^1)^2 - (x^2)^2 \quad \text{and} \quad g^2(x) := (x^1)^2 + (x^2)^2 - 1,$$

respectively. Note that these functions satisfy the properties from Assumption 1.2, but that g^1 is not convex as a function of the whole vector x . The corresponding set W is shown in Figure 4.1. Obviously, this set is not convex. Note also that there is no (clear) connection to the sets $\Omega(x)$. For example, taking $x := (\frac{1}{2}, \frac{1}{2}) \in W$, a simple calculation shows that $\Omega(x) = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{\sqrt{3}}{2}, \frac{\sqrt{3}}{2}]$, and this set is neither a subset of W nor vice versa.

(b) Consider once again a Nash game with two players, each having a single decision variable, and arbitrary cost functions. Let the strategy spaces be defined by

$$X_1(x^2) = \begin{cases} \mathbb{R}, & \text{if } x^2 \neq 0, \\ \emptyset, & \text{if } x^2 = 0, \end{cases} \quad \text{and} \quad X_2(x^1) = \begin{cases} \mathbb{R}, & \text{if } x^1 \neq 0, \\ \emptyset, & \text{if } x^1 = 0. \end{cases}$$

Note that these sets are always closed and convex. In this case, however, the set W is equal to $\mathbb{R}^2 \setminus \{(x^1, x^2) \mid x^1 x^2 = 0\}$, hence W is neither closed nor convex.

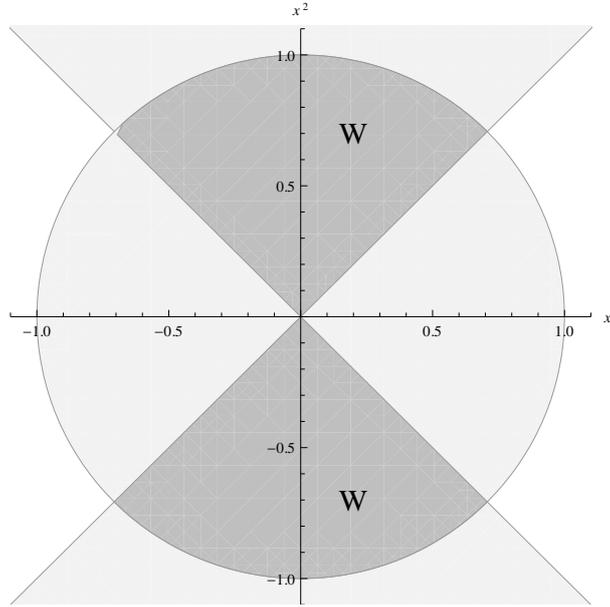


Figure 4.1.: The set W for the example from Remark 4.1 (a)

- (c) The previous counterexample is somewhat artificial since the sets $X_\nu(x^{-\nu})$ were not defined by some functions g^ν . Of course, if we have $X_\nu(x^{-\nu}) = \{x^\nu \mid g^\nu(x^\nu, x^{-\nu}) \leq 0\}$ for all $\nu = 1, \dots, N$ for some continuous functions g^ν (as required in Assumption 1.2), then the set W is obviously closed. Recall, however, that Figure 4.1 shows that W is nonconvex in general.
- (d) Let \bar{x} be a solution of the GNEP. Then $\bar{x} \in \Omega(\bar{x})$, and this implies $\bar{x} \in W$, see Theorem 4.2 (a). In particular, W is nonempty whenever the GNEP has at least one solution.

Using the set W and the mapping V_α it is possible to reformulate the GNEP as a constrained optimization problem not only for jointly convex GNEPs as it was done in [45] but also for player convex GNEPs. This is the content of the following theorem, whose proof is an adaption on the proofs of [45, Theorem 2.2 and Proposition 2.3].

Theorem 4.2 *Let Assumption 1.2 be satisfied. Then the following statements hold:*

- (a) $x \in W$ if and only if $x \in \Omega(x)$.
- (b) $V_\alpha(x) \geq 0$ for all $x \in W$.
- (c) \bar{x} is a generalized Nash equilibrium if and only if $\bar{x} \in W$ and $V_\alpha(\bar{x}) = 0$.
- (d) For all $x \in \mathbb{R}^n$ with $\Omega(x) \neq \emptyset$, there exists a unique vector

$$y_\alpha(x) := (y_\alpha^1(x), \dots, y_\alpha^N(x))$$

such that

$$\operatorname{argmin}_{y^\nu \in X_\nu(x^{-\nu})} \left[\theta_\nu(y^\nu, x^{-\nu}) + \frac{\alpha}{2} \|x^\nu - y^\nu\|^2 \right] = y_\alpha^\nu(x) \quad (4.3)$$

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

(e) \bar{x} is a generalized Nash equilibrium if and only if \bar{x} is a fixed point of the mapping $x \mapsto y_\alpha(x)$, i.e., if and only if $\bar{x} = y_\alpha(\bar{x})$.

Proof. (a) By definition, $x \in \Omega(x)$ means $x^\nu \in X_\nu(x^{-\nu})$ for all $\nu = 1, \dots, N$, which is equivalent to $x \in W$.

(b) For all $x \in W$ we have $x \in \Omega(x)$ by part (a). Therefore

$$V_\alpha(x) = \max_{y \in \Omega(x)} \Psi_\alpha(x, y) \geq \Psi_\alpha(x, x) = 0.$$

(c) Let \bar{x} be a generalized Nash equilibrium. Then we have $\bar{x} \in \Omega(\bar{x})$ (hence $\bar{x} \in W$ by part (a)) and for all $\nu = 1, \dots, N$

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \text{for all } x^\nu \in X_\nu(\bar{x}^{-\nu}).$$

This implies

$$\Psi_\alpha(\bar{x}, y) = \sum_{\nu=1}^N \underbrace{(\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) - \theta_\nu(y^\nu, \bar{x}^{-\nu}))}_{\leq 0} - \frac{\alpha}{2} \|\bar{x} - y\|^2 \leq 0$$

for all $y \in \Omega(\bar{x})$. Therefore, we get $V_\alpha(\bar{x}) = \max_{y \in \Omega(\bar{x})} \Psi_\alpha(\bar{x}, y) \leq 0$. Together with part (b), we obtain $V_\alpha(\bar{x}) = 0$.

Conversely, assume that $\bar{x} \in W$ (which is equivalent to $\bar{x} \in \Omega(\bar{x})$ by part (a)) and $V_\alpha(\bar{x}) = 0$. Then $\Psi_\alpha(\bar{x}, y) \leq 0$ holds for all $y \in \Omega(\bar{x})$. Let $\nu \in \{1, \dots, N\}$ be a fixed player, $x^\nu \in X_\nu(\bar{x}^{-\nu})$ and $\lambda \in (0, 1)$ arbitrary. Define $y = (y^1, \dots, y^N) \in \mathbb{R}^n$ by

$$y^\mu := \begin{cases} \bar{x}^\mu, & \text{if } \mu \neq \nu, \\ \lambda \bar{x}^\nu + (1 - \lambda)x^\nu, & \text{if } \mu = \nu \end{cases} \quad \forall \mu = 1, \dots, N.$$

The convexity of $X_\nu(\bar{x}^{-\nu})$ implies $y^\mu \in X_\mu(\bar{x}^{-\mu})$ for all $\mu = 1, \dots, N$ and therefore $y \in \Omega(\bar{x})$. Using this special y and exploiting the convexity of θ_ν with respect to x^ν , we get

$$\begin{aligned} 0 &\geq \Psi_\alpha(\bar{x}, y) \\ &= \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) - \theta_\nu(\lambda \bar{x}^\nu + (1 - \lambda)x^\nu, \bar{x}^{-\nu}) - \frac{\alpha}{2} (1 - \lambda)^2 \|\bar{x}^\nu - x^\nu\|^2 \\ &\geq (1 - \lambda)\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) - (1 - \lambda)\theta_\nu(x^\nu, \bar{x}^{-\nu}) - \frac{\alpha}{2} (1 - \lambda)^2 \|\bar{x}^\nu - x^\nu\|^2. \end{aligned}$$

Dividing both sides by $(1 - \lambda)$ and taking the limit $\lambda \uparrow 1$, we see

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq \theta_\nu(x^\nu, \bar{x}^{-\nu}).$$

Since this is true for all $x^\nu \in X_\nu(\bar{x}^{-\nu})$ and all players $\nu = 1, \dots, N$, \bar{x} is a generalized Nash equilibrium.

(d) For $x \in \mathbb{R}^n$ with $\Omega(x) \neq \emptyset$, the closed and convex sets $X_\nu(x^{-\nu})$ are nonempty for all $\nu = 1, \dots, N$. The statement therefore follows from the fact that the uniformly convex function (uniform with respect to y^ν)

$$\theta_\nu(y^\nu, x^{-\nu}) + \frac{\alpha}{2} \|x^\nu - y^\nu\|^2$$

attains a unique minimum on the nonempty, closed and convex set $X_\nu(x^{-\nu})$.

(e) First, let \bar{x} be a generalized Nash equilibrium. Then $\bar{x} \in \Omega(\bar{x})$ follows by definition, and further $V_\alpha(\bar{x}) = 0$ holds in view of part (c). Therefore, we obtain

$$\Psi_\alpha(\bar{x}, \bar{x}) = 0 = V_\alpha(\bar{x}) = \max_{y \in \Omega(\bar{x})} \Psi_\alpha(\bar{x}, y) = \Psi_\alpha(\bar{x}, y_\alpha(\bar{x})).$$

With $\bar{x} \in \Omega(\bar{x})$, we get $\bar{x} = y_\alpha(\bar{x})$ from part (d), taking into account the uniqueness of the maximizer $y_\alpha(\bar{x})$.

Conversely, let $\bar{x} \in \mathbb{R}^n$ be such that $\bar{x} = y_\alpha(\bar{x})$. Then $\bar{x} \in \Omega(\bar{x})$ and, therefore, $\bar{x} \in W$ in view of part (a). Moreover, we obtain

$$0 = \Psi_\alpha(\bar{x}, \bar{x}) = \Psi_\alpha(\bar{x}, y_\alpha(\bar{x})) = V_\alpha(\bar{x}),$$

and this means that \bar{x} is a generalized Nash equilibrium by part (c). □

In view of Theorem 4.2 finding a solution of the GNEP is equivalent to finding a minimum \bar{x} of the constrained optimization problem

$$\min V_\alpha(x) \quad \text{s.t.} \quad x \in W \tag{4.4}$$

satisfying $V_\alpha(\bar{x}) = 0$. Remark 4.1 shows that W is nonempty (at least if the GNEP has a solution) and closed under Assumption 1.2, but might be nonconvex. As for each $x \in W$ we have $x \in \Omega(x)$ by Theorem 4.2 (a), it follows that $\Omega(x)$ is nonempty and, thus, the objective function V_α is well-defined on W . However, V_α is, in general, nondifferentiable and might even be discontinuous. The following is an example for the latter effect.

Example 4.3 Consider a jointly convex 2-player game, where player 1 has one variable x^1 and player 2 has two variables (x_1^2, x_2^2) . As usual define $x = (x^1, x_1^2, x_2^2)$. Let the common strategy space be given by

$$X = \{x \in [0, 10] \times [-10, 10] \times [0, 20] \mid (x_1^2)^2 + (x_2^2 - x^1)^2 - (x^1)^2 \leq 0\}.$$

Further let the cost functions be defined by

$$\theta_1(x) := (x^1 + 10)^2 \quad \text{and} \quad \theta_2(x) := (x_1^2)^2 + (x_2^2)^2,$$

respectively. The corresponding regularized Nikaido-Isoda function is given by

$$\Psi_\alpha(x, y) := (x^1 + 10)^2 + (x_1^2)^2 + (x_2^2)^2 - (y^1 + 10)^2 - (y_1^2)^2 - (y_2^2)^2 - \frac{\alpha}{2} \|x - y\|^2.$$

The unconstrained maximum of $\Psi_\alpha(x, y)$ with respect to y is $\left(\frac{-20+\alpha x^1}{2+\alpha}, \frac{\alpha x_1^2}{2+\alpha}, \frac{\alpha x_2^2}{2+\alpha}\right)^T$. Now consider for $\delta > 0$ sufficiently small the feasible sequence

$$x(\delta) := (10, \sqrt{20\delta - \delta^2}, \delta)^T \rightarrow (10, 0, 0)^T := x^* \quad \text{for } \delta \downarrow 0.$$

Note that x^* belongs to X . Since the strategy set $X_1(x^2(\delta)) = \{10\}$ is single valued and the unconstrained maximum is feasible for the second player we have for all $\alpha > 0$ and all $\delta > 0$ sufficiently small

$$y_\alpha(x(\delta)) = \left(10, \frac{\alpha\sqrt{20\delta - \delta^2}}{2 + \alpha}, \frac{\alpha\delta}{2 + \alpha}\right)^T \rightarrow (10, 0, 0)^T \quad \text{for } \delta \downarrow 0.$$

On the other hand, for the parameter $\alpha = 2$ or, more generally, for an arbitrary parameter $\alpha \in (0, 2]$, it can be shown that $y_\alpha(x^*) = (0, 0, 0)^T$, hence the function y_α is not continuous in $(10, 0, 0)^T$. Furthermore, we have

$$\begin{aligned} V_\alpha(x(\delta)) &= \Psi_\alpha(x(\delta), y_\alpha(x(\delta))) \\ &= 20\delta \left(1 - \frac{\alpha^2}{(2 + \alpha)^2} - \frac{\alpha}{2} \left(1 - \frac{\alpha}{2 + \alpha}\right)^2\right) \rightarrow 0 \quad \text{for } \delta \downarrow 0, \end{aligned}$$

whereas $V_\alpha(x^*) = 20^2 - 10^2 - \frac{\alpha}{2}10^2 \neq 0$, which shows that V_α is not continuous in x^* . \diamond

In Section 4.2 we will give conditions guaranteeing that the function V_α is continuous and even a PC^1 mapping under fairly mild conditions. For the moment, however, we will leave the constrained problem (4.4) and present a new *unconstrained* optimization reformulation of the GNEP.

To this end, we have to find a way to define the function $V_\alpha(x) := \max_{y \in \Omega(x)} \Psi_\alpha(x, y)$ for those points $x \in \mathbb{R}^n$ where $\Omega(x)$ is empty. By Theorem 4.2 (a) $\Omega(x) \neq \emptyset$ for all $x \in W$. Hence the idea is first to project $x \in \mathbb{R}^n$ on W and then take the set Ω at the projected point. To get a well-defined Euclidean projection we need a nonempty closed and convex set. Since the set W is, in general, non convex, we define the set

$$X := \text{cl}(\text{conv}(W)), \tag{4.5}$$

and make the following central assumption for our subsequent analysis.

Assumption 4.4 *The set $W = \{x \in \mathbb{R}^n \mid x^\nu \in X_\nu(x^{-\nu}) \text{ for all } \nu = 1, \dots, N\}$ is nonempty and $\Omega(x)$ is nonempty for all $x \in X = \text{cl}(\text{conv}(W))$.*

By this Assumption 4.4 the set X is always nonempty, closed, and convex which implies that the Euclidean projection onto this set is well-defined and unique. This paves the way to our unconstrained optimization reformulation. The necessity of the closure in (4.5) is discussed in Remark 4.5 (d) below. Further the following remark shows that the class of GNEPs satisfying Assumption 4.4 contains the jointly convex GNEPs, and is, in fact, strictly larger, but that not all GNEPs satisfy Assumption 4.4.

Remark 4.5 (a) Consider a jointly convex GNEP. Then there is a common nonempty, closed, and convex set $Y \subseteq \mathbb{R}^n$ such that

$$X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R}^{n_\nu} \mid (x^\nu, x^{-\nu}) \in Y\}$$

for all players $\nu = 1, \dots, N$. In this case, we obviously have $W = Y$ which, in turn, implies $X = \text{cl}(\text{conv}(W)) = \text{cl}(\text{conv}(Y)) = Y$ since Y is already convex and closed. Moreover, we have $W \neq \emptyset$ and $\Omega(x) \neq \emptyset$ for all $x \in X = W = Y$ since x belongs to $\Omega(x)$ for all $x \in W$ by Theorem 4.2 (a). Hence Assumption 4.4 holds for jointly convex GNEPs.

(b) Remark (a) can be generalized on GNEPs with strategy sets

$$X_\nu(x^{-\nu}) := \{x^\nu \in \mathbb{R}^{n_\nu} \mid g^\nu(x^\nu, x^{-\nu}) \leq 0\}$$

for $\nu = 1, \dots, N$, where $g^\nu : \mathbb{R}^n \rightarrow \mathbb{R}^{m_\nu}$ are continuous and convex functions of *all* variables x . Then, again, $X = W$ holds due to the assumed continuity and convexity of all g^ν . Moreover $\Omega(x) \neq \emptyset$ for all $x \in W = X$. If a solution exists, we have $W \neq \emptyset$. Hence Assumption 4.4 also holds in this situation.

(c) An explicit example of a non-jointly convex GNEP which obviously satisfies Assumption 4.4 is the one from Remark 4.1 (a), see Figure 4.1. In particular, this shows that the class of GNEPs satisfying Assumption 4.4 strictly includes the class of jointly convex GNEPs.

(d) On the other hand, there exist GNEPs which do not satisfy Assumption 4.4. To see this, consider a GNEP with two players, each controlling a single variable, with strategy spaces defined by

$$X_1(x^2) := \{x^1 \in \mathbb{R} \mid g^1(x) := 1 - x^1 x^2 \leq 0\} = \begin{cases} (-\infty, 1/x^2], & \text{if } x^2 < 0, \\ \emptyset, & \text{if } x^2 = 0, \\ [1/x^2, \infty), & \text{if } x^2 > 0, \end{cases}$$

$$X_2(x^1) := \{x^2 \in \mathbb{R} \mid g^2(x) := x^2 - 1 \leq 0\} = (-\infty, 1].$$

Then the functions g^ν are convex in x^ν for fixed $x^{-\nu}$ and all the sets $X_\nu(x^{-\nu})$ are closed and convex. Moreover, we have $W = \{x \in \mathbb{R}^2 \mid x_1 x_2 \geq 1, x_2 \leq 1\}$ which is not connected and, in particular, not convex, see Figure 4.2. We further get $X = \text{cl}(\text{conv}(W)) = \mathbb{R} \times (-\infty, 1]$ and $\Omega((0,0)) = \emptyset$. Hence Assumption 4.4 is violated in this case.

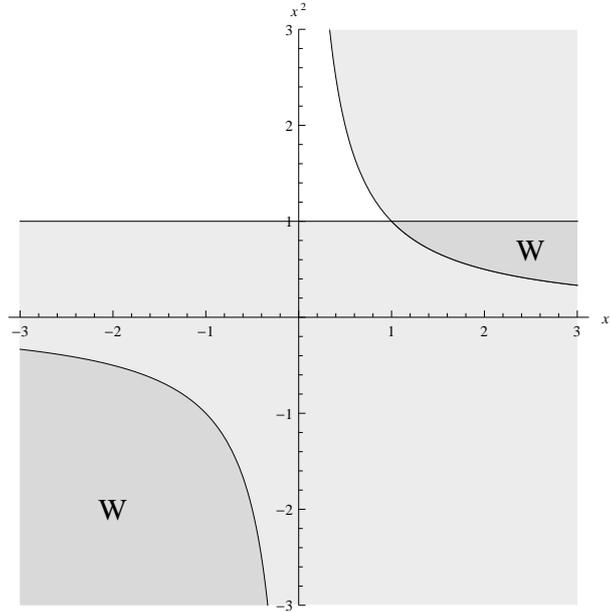


Figure 4.2.: The set W for the example from Remark 4.5 (d)

- (e) If we have a GNEP where Assumption 1.2 holds, but there exist $x \in X$ with $\Omega(x) = \emptyset$ then the set W is not convex, because otherwise, if $W = \text{conv}(W)$, Assumption 1.2 makes W a closed set, see Remark 4.1 (c), and hence $X = \text{cl}(\text{conv}(W)) = W$, so that $\Omega(x) \neq \emptyset$ for all $x \in X = W$ would follow from Theorem 4.2 (a). These GNEPs are the hard ones, since we do not get an unconstrained reformulation for them and even the constrained optimization problem is a non-convex and therefore difficult problem.

Next we define the function for our unconstrained optimization reformulation of the GNEPs satisfying Assumption 4.4.

Definition 4.6 Consider GNEPs where Assumption 4.4 holds. For those we define for all $x \in \mathbb{R}^n$ and $\alpha > 0$

$$\begin{aligned} \bar{y}_\alpha(x) &:= \operatorname{argmax}_{y \in \Omega(P_X[x])} \Psi_\alpha(x, y) \quad \text{and} \\ \bar{V}_\alpha(x) &:= \max_{y \in \Omega(P_X[x])} \Psi_\alpha(x, y) = \Psi_\alpha(x, \bar{y}_\alpha(x)). \end{aligned}$$

Given two parameters $0 < \alpha < \beta$ and a constant $c > 0$, we further define

$$\bar{V}_{\alpha\beta}^c(x) := \bar{V}_\alpha(x) - \bar{V}_\beta(x) + c\|x - P_X[x]\|^2$$

for all $x \in \mathbb{R}^n$, where $\bar{V}_\beta(x)$ is defined similarly to $\bar{V}_\alpha(x)$.

The difference between the definitions of V_α and \bar{V}_α is that we maximize over $\Omega(x)$ in the former case, whereas we maximize over the set $\Omega(P_X[x])$ in the latter case. This is important since $\Omega(x)$ might be empty for certain $x \in \mathbb{R}^n$, whereas the projection $P_X[x]$ always exists due to the fact that X is nonempty, closed, and convex as a consequence of Assumption 4.4 and, furthermore, the set $\Omega(P_X[x])$ is (closed, convex, and) nonempty again by Assumption 4.4. Consequently, $\bar{y}_\alpha(x)$ and therefore also $\bar{V}_\alpha(x)$ are well-defined for all $x \in \mathbb{R}^n$. This, in turn, implies that $\bar{V}_{\alpha\beta}^c$ is well-defined for all $x \in \mathbb{R}^n$. Therefore, Assumption 4.4 guarantees that our functions are well-defined. Note that we have

$$\bar{y}_\alpha(x) = y_\alpha(x) \quad \text{and therefore} \quad \bar{V}_\alpha(x) = V_\alpha(x) \quad \text{for all } x \in X, \quad (4.6)$$

hence these two functions coincide on the set X . This simple observation will be used fruitfully in the proof of the main theorem of this section, showing that we get an unconstrained optimization reformulation of the GNEP via $\bar{V}_{\alpha\beta}^c$. But first the following crucial lemma, which is similar to Lemma 3.4, has to be proven.

Lemma 4.7 *Let Assumption 4.4 hold. Then we have the following inequalities for all $x \in \mathbb{R}^n$:*

$$\begin{aligned} \frac{\beta - \alpha}{2} \|x - \bar{y}_\beta(x)\|^2 + c \|x - P_X[x]\|^2 &\leq \bar{V}_{\alpha\beta}^c(x), \\ \frac{\beta - \alpha}{2} \|x - \bar{y}_\alpha(x)\|^2 + c \|x - P_X[x]\|^2 &\geq \bar{V}_{\alpha\beta}^c(x). \end{aligned}$$

Proof. Assumption 4.4 guarantees that all involved functions are well-defined. We have $\bar{y}_\alpha(x) \in \Omega(P_X[x])$ and $\bar{y}_\beta(x) \in \Omega(P_X[x])$. Therefore

$$\bar{V}_\beta(x) = \Psi_\beta(x, \bar{y}_\beta(x)) = \max_{y \in \Omega(P_X[x])} \Psi_\beta(x, y) \geq \Psi_\beta(x, \bar{y}_\alpha(x)), \quad (4.7)$$

$$\bar{V}_\alpha(x) = \Psi_\alpha(x, \bar{y}_\alpha(x)) = \max_{y \in \Omega(P_X[x])} \Psi_\alpha(x, y) \geq \Psi_\alpha(x, \bar{y}_\beta(x)). \quad (4.8)$$

On the one hand, this implies

$$\begin{aligned} \bar{V}_{\alpha\beta}^c(x) &= \bar{V}_\alpha(x) - \bar{V}_\beta(x) + c \|x - P_X[x]\|^2 \\ &\stackrel{(4.7)}{\leq} \Psi_\alpha(x, \bar{y}_\alpha(x)) - \Psi_\beta(x, \bar{y}_\alpha(x)) + c \|x - P_X[x]\|^2 \\ &= \frac{\beta - \alpha}{2} \|x - \bar{y}_\alpha(x)\|^2 + c \|x - P_X[x]\|^2 \end{aligned}$$

and, on the other hand, we obtain

$$\begin{aligned} \bar{V}_{\alpha\beta}^c(x) &= \bar{V}_\alpha(x) - \bar{V}_\beta(x) + c \|x - P_X[x]\|^2 \\ &\stackrel{(4.8)}{\geq} \Psi_\alpha(x, \bar{y}_\beta(x)) - \Psi_\beta(x, \bar{y}_\beta(x)) + c \|x - P_X[x]\|^2 \\ &= \frac{\beta - \alpha}{2} \|x - \bar{y}_\beta(x)\|^2 + c \|x - P_X[x]\|^2 \end{aligned}$$

for all $x \in \mathbb{R}^n$. □

We are now in a position to show that the function $\bar{V}_{\alpha\beta}^c$ provides an unconstrained optimization reformulation of the GNEP.

Theorem 4.8 *Under Assumption 4.4 the following statements hold:*

- (a) $\bar{V}_{\alpha\beta}^c(x) \geq 0$ for all $x \in \mathbb{R}^n$.
- (b) \bar{x} is a generalized Nash equilibrium if and only if \bar{x} is a minimum of $\bar{V}_{\alpha\beta}^c$ satisfying $\bar{V}_{\alpha\beta}^c(\bar{x}) = 0$.

Proof. Again Assumption 4.4 is needed to guarantee that the functions are all well-defined. The first inequality in Lemma 4.7 immediately gives

$$\bar{V}_{\alpha\beta}^c(x) \geq \frac{\beta - \alpha}{2} \|x - \bar{y}_\beta(x)\|^2 + c \|x - P_X[x]\|^2 \geq 0$$

for all $x \in \mathbb{R}^n$, hence statement (a) holds.

In order to verify the second statement, first assume that \bar{x} is a generalized Nash equilibrium. Then $\bar{x} \in \Omega(\bar{x})$, and Theorem 4.2 (a) therefore implies $\bar{x} \in W \subseteq X$. This, in turn, gives $P_X[\bar{x}] = \bar{x}$, and together with the fixed point characterization of Theorem 4.2 (e), we get

$$\bar{x} = y_\alpha(\bar{x}) = \bar{y}_\alpha(\bar{x}),$$

where the second equality follows from (4.6). The second inequality of Lemma 4.7 then implies $\bar{V}_{\alpha\beta}^c(\bar{x}) \leq 0$. In view of part (a), we therefore have $\bar{V}_{\alpha\beta}^c(\bar{x}) = 0$.

Conversely, assume that $\bar{V}_{\alpha\beta}^c(\bar{x}) = 0$ for some $\bar{x} \in \mathbb{R}^n$. Then we obtain from the first inequality of Lemma 4.7 that

$$0 = \bar{V}_{\alpha\beta}^c(\bar{x}) \geq \frac{\beta - \alpha}{2} \|\bar{x} - \bar{y}_\beta(\bar{x})\|^2 + c \|\bar{x} - P_X[\bar{x}]\|^2 \geq 0. \quad (4.9)$$

Since $c > 0$, this means $P_X[\bar{x}] = \bar{x}$, i.e., $\bar{x} \in X$, and

$$\bar{x} = \bar{y}_\beta(\bar{x}) = y_\beta(\bar{x}),$$

where, once again, we used (4.6). Hence \bar{x} is a generalized Nash equilibrium by the fixed point characterization from Theorem 4.2 (e). □

Theorem 4.8 shows that the generalized Nash equilibria \bar{x} are exactly the minima of the function $\bar{V}_{\alpha\beta}^c$ satisfying $\bar{V}_{\alpha\beta}^c(\bar{x}) = 0$. We therefore have the unconstrained optimization reformulation

$$\min \bar{V}_{\alpha\beta}^c(x), \quad x \in \mathbb{R}^n, \quad (4.10)$$

in order to find solutions of a GNEP. Note, however, that we obtain this unconstrained reformulation only for the class of GNEPs which satisfy Assumption 4.4, whereas the

corresponding constrained reformulation (4.4) holds for an arbitrary player convex GNEP, not necessarily satisfying this condition.

In the jointly convex case, it is possible to show that the additional term $c\|x - P_X[x]\|^2$ is not necessary, so we can define $\bar{V}_{\alpha\beta}^c$ with $c := 0$ in this case. This is content of the following Corollary.

Corollary 4.9 *Assume Assumption 4.4 holds for a jointly convex GNEP and set $c := 0$ in the definition of $\bar{V}_{\alpha\beta}^0$. Then we obtain*

- (a) $\bar{V}_{\alpha\beta}^0(x) \geq 0$ for all $x \in \mathbb{R}^n$.
- (b) \bar{x} is a generalized Nash equilibrium if and only if \bar{x} is a minimum of $\bar{V}_{\alpha\beta}^0$ satisfying $\bar{V}_{\alpha\beta}^0(\bar{x}) = 0$.

Proof. Lemma 4.7 also holds for $c = 0$ and statement (a) is again an immediate consequence of it. We can follow the lines of the proof of Theorem 4.8 to get that a generalized Nash equilibrium \bar{x} is a minimum of $\bar{V}_{\alpha\beta}^0$ satisfying $\bar{V}_{\alpha\beta}^0(\bar{x}) = 0$, which is one part of (b). It remains to show the other direction of (b). Therefore assume that $\bar{V}_{\alpha\beta}^0(\bar{x}) = 0$ for some $\bar{x} \in \mathbb{R}^n$. Then we obtain from the first inequality in Lemma 4.7

$$0 = \bar{V}_{\alpha\beta}^0(\bar{x}) \geq \frac{\beta - \alpha}{2} \|\bar{x} - \bar{y}_\beta(\bar{x})\|^2 \geq 0.$$

Consequently, we have $\bar{x} = \bar{y}_\beta(\bar{x}) \in \Omega(P_X[\bar{x}])$, i.e.,

$$\bar{x}^\nu \in X_\nu((P_X[\bar{x}])^{-\nu}) = \{x^\nu \in \mathbb{R}^{n_\nu} \mid (x^\nu, (P_X[\bar{x}])^{-\nu}) \in X\}$$

for all $\nu = 1, \dots, N$. Let $\mu \in \{1, \dots, N\}$ be arbitrarily given. Then $(\bar{x}^\mu, (P_X[\bar{x}])^{-\mu}) \in X$ and we have

$$\begin{aligned} \|\bar{x} - (\bar{x}^\mu, (P_X[\bar{x}])^{-\mu})\|^2 &= \sum_{\nu=1, \nu \neq \mu}^N \|\bar{x}^\nu - (P_X[\bar{x}])^\nu\|^2 \\ &\leq \sum_{\nu=1}^N \|\bar{x}^\nu - (P_X[\bar{x}])^\nu\|^2 \\ &= \|\bar{x} - P_X[\bar{x}]\|^2. \end{aligned}$$

Since the projection $P_X[\bar{x}]$ onto the nonempty, closed and convex set X is the unique solution of the problem

$$\min \frac{1}{2} \|\bar{x} - z\|^2 \quad \text{s.t.} \quad z \in X,$$

we must have $\bar{x}^\mu = (P_X[\bar{x}])^\mu$. Since $\mu \in \{1, \dots, N\}$ was arbitrarily chosen, this is true for all components and hence $\bar{x} = P_X[\bar{x}]$, i.e., $\bar{x} \in X$. Thus we get $y_\beta(\bar{x}) = \bar{y}_\beta(\bar{x}) = \bar{x}$ by (4.6). Therefore, \bar{x} is a generalized Nash equilibrium by the fixed point characterization of Theorem 4.2 (e). \square

With the previous Corollary 4.9 the question rises, if the additional term is necessary in the general case. The next example gives the answer, it is strictly necessary.

Example 4.10 Consider the two player game defined via

$$\begin{aligned} \text{Player 1: } & \min_{x^1} (x^1)^2 & \text{s.t. } & (x^1)^2 + (x^2)^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} (x^2 + 3)^2 & \text{s.t. } & -2 \leq x^2 \leq -1. \end{aligned}$$

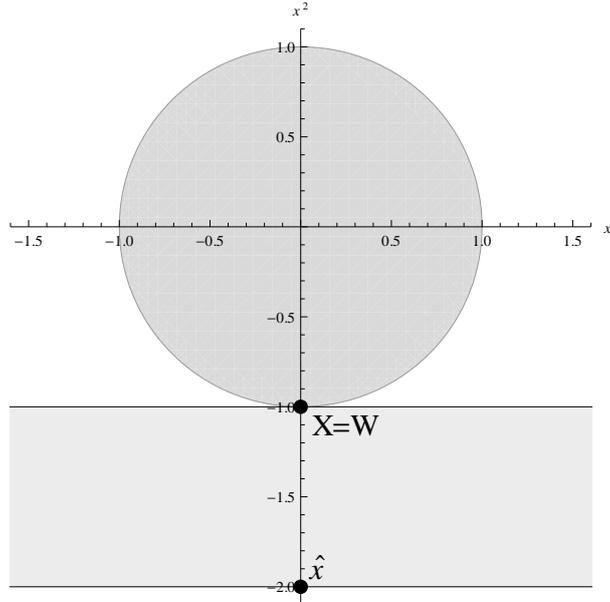


Figure 4.3.: The feasible set for Example 4.10

Here we have $W = X = \{(0, -1)\}$, see Figure 4.3. If we consider the point $\hat{x} = (0, -2)$, we have $P_X[\hat{x}] = (0, -1)$ and $\Omega((0, -1)) = \{0\} \times [-2, -1]$. Thus we get

$$\bar{y}_\gamma(\hat{x}) = (0, -2) = \hat{x}$$

for all $\gamma > 0$ and this implies for $0 < \alpha < \beta$

$$\bar{V}_\alpha(\hat{x}) - \bar{V}_\beta(\hat{x}) = \Psi_\alpha(\hat{x}, \bar{y}_\alpha(\hat{x})) - \Psi_\beta(\hat{x}, \bar{y}_\beta(\hat{x})) = 0.$$

But we have $\hat{x} \notin W$ and, therefore, \hat{x} is not a solution of the GNEP. This shows that we cannot skip the additional term $c\|x - P_X[x]\|^2$ in the definition of $\bar{V}_{\alpha\beta}^c$. \diamond

Similar to the constrained reformulation the objective function $\bar{V}_{\alpha\beta}^c$ is nondifferentiable in general and, even worse, might be discontinuous. The smoothness properties of $\bar{V}_{\alpha\beta}^c$ will be discussed in more detail in Section 4.3.

There exists an alternative unconstrained optimization formulation of the GNEP, described in the following remark. This formulation can be shown to have similar smoothness properties as those we show in the next two sections.

Remark 4.11 For an arbitrary parameter $\alpha > 0$, let us define the functions

$$\begin{aligned}\tilde{y}_\alpha(x) &:= \operatorname{argmax}_{y \in \Omega(P_X[x])} \Psi_\alpha(P_X[x], y) \quad \text{and} \\ \tilde{V}_\alpha(x) &:= \max_{y \in \Omega(P_X[x])} \Psi_\alpha(P_X[x], y) = \Psi_\alpha(P_X[x], \tilde{y}_\alpha(x))\end{aligned}$$

for all $x \in \mathbb{R}^n$. Then, given two parameters $0 < \alpha < \beta$ and a positive constant $c > 0$, let us define

$$\tilde{V}_{\alpha\beta}^c(x) := \tilde{V}_\alpha(x) - \tilde{V}_\beta(x) + c\|x - P_X[x]\|^2.$$

The difference to the previous reformulation is that the first argument of the function Ψ_α is the projection $P_X[x]$ instead of x . In a way similar to the above approach, one can show that finding a generalized Nash equilibrium is equivalent to finding solutions of the unconstrained optimization problem

$$\min \tilde{V}_{\alpha\beta}^c(x), \quad x \in \mathbb{R}^n$$

with vanishing function value. Note that the functions $\bar{V}_{\alpha\beta}^c$ and $\tilde{V}_{\alpha\beta}^c$ coincide on the set X and only differ outside of it. In this reformulation the additional term $c\|x - P_X[x]\|^2$ is needed to guarantee that the solutions of our optimization problem belong to X even in the jointly convex case.

4.2. Smoothness Properties of the Constrained Reformulation

In [45, Example 2.4] it was shown that the constrained reformulation (4.4) of a jointly convex GNEP does not result in a differentiable objective function V_α in general. This was the reason why the approach was not further investigated there. In Example 4.3 we have seen that the objective function might even be discontinuous. Nevertheless one can show that V_α has certain smoothness properties under mild conditions. We will show the following results:

- V_α is continuous provided that $X_\nu(x^{-\nu})$ either satisfies a Slater condition or consists of a single element;
- V_α is a PC^1 function provided that it is continuous, the functions g^ν and θ_ν are twice continuously differentiable and a constant rank constraint qualification holds.

To verify the continuity of V_α , we first recall some terminology and results from set-valued analysis.

Definition 4.12 Suppose $X \subseteq \mathbb{R}^n, Y \subseteq \mathbb{R}^m$, and $\Phi : X \rightrightarrows Y$ is a point-to-set mapping. Then Φ is called

- (a) lower semicontinuous in $x^* \in X$, if for all sequences $\{x^k\} \subseteq X$ with $x^k \rightarrow x^*$ and all $y^* \in \Phi(x^*)$, there exists a number $m \in \mathbb{N}$ and a sequence $\{y^k\} \subseteq Y$ with $y^k \rightarrow y^*$ and $y^k \in \Phi(x^k)$ for all $k \geq m$;
- (b) closed in $x^* \in X$, if for all sequences $\{x^k\} \subseteq X$ with $x^k \rightarrow x^*$ and all sequences $y^k \rightarrow y^*$ with $y^k \in \Phi(x^k)$ for all $k \in \mathbb{N}$ sufficiently large, we have $y^* \in \Phi(x^*)$;
- (c) lower semicontinuous or closed on X if it is lower semicontinuous or closed in every $x \in X$.

The definition of a lower semicontinuous set-valued mapping is in the sense of Berge [6]. Alternative names used in the literature are “open mapping” (see [48]) and “inner semicontinuous mapping” (see [67]). These references are also useful sources to find further results on set-valued analysis. For our subsequent analysis a useful result is stated in the next lemma, which is immediately obtained from [48, Corollaries 8.1 and 9.1].

Lemma 4.13 *Let $X \subseteq \mathbb{R}^n$ arbitrary, $Y \subseteq \mathbb{R}^m$ convex, and $f : X \times Y \rightarrow \mathbb{R}$ be concave in y for fixed x and continuous on $X \times Y$. Let $\Phi : X \rightrightarrows Y$ be a point-to-set map which is closed in a neighbourhood of \bar{x} and lower semicontinuous in \bar{x} , and let $\Phi(x)$ be convex in a neighbourhood of \bar{x} . Define*

$$Y(x) := \{z \in \Phi(x) \mid \sup_{y \in \Phi(x)} f(x, y) = f(x, z)\}$$

and assume that $Y(\bar{x})$ has exactly one element. Then the point-to-set mapping $x \mapsto Y(x)$ is lower semicontinuous and closed in \bar{x} .

The equivalence of closedness of a point-to-set mapping and closedness of its graph was shown in [48, Theorem 2], and has become a useful tool. We can use this equivalence and Lemma 4.13 to prove a sufficient condition for continuity of V_α .

Theorem 4.14 *Suppose that Assumption 1.2 holds and that the point-to-set mapping $x \mapsto \Omega(x)$ is lower semicontinuous in $x^* \in W$. Then the functions y_α and V_α are continuous at $x^* \in W$.*

Proof. First observe that Assumption 1.2 implies that the function $\Psi_\alpha(x, \cdot)$ is concave for fixed x and continuous on $\mathbb{R}^n \times \mathbb{R}^n$.

By the product structure $\Omega(x) = X_1(x^{-1}) \times \dots \times X_N(x^{-N})$ it is clear that $\Omega(x)$ is closed if and only if $X_\nu(x^{-\nu})$ is closed for all $\nu = 1, \dots, N$. The point-to-set mappings $x^{-\nu} \mapsto X_\nu(x^{-\nu})$, $\nu = 1, \dots, N$, are closed for all $x \in W$ since their graphs

$$\{(y^\nu, x^{-\nu}) \in \mathbb{R}^{n_\nu} \times \mathbb{R}^{n-n_\nu} \mid g^\nu(y^\nu, x^{-\nu}) \leq 0\}$$

are closed sets due to the assumed continuity of g^ν , cf. [48, Theorem 2].

Theorem 4.2 (a) implies that $\Omega(x)$ is nonempty for all $x \in W$; moreover, these sets are also convex as a consequence of Assumption 1.2. Theorem 4.2 (d) shows that the sets

$Y_\alpha(x) := \{z \in \Omega(x) \mid \sup_{y \in \Omega(x)} \Psi_\alpha(x, y) = \Psi_\alpha(x, z)\}$ consist of exactly one element for all $x \in W$, namely $y_\alpha(x)$. Lemma 4.13 therefore implies that $x \mapsto \{y_\alpha(x)\}$, viewed as a point-to-set mapping, is lower semicontinuous and closed in $x^* \in W$. This implies that the single-valued function $x \mapsto y_\alpha(x)$ is continuous at $x^* \in W$. Hence $V_\alpha(x) = \Psi_\alpha(x, y_\alpha(x))$, being a composition of continuous maps, is also continuous at $x^* \in W$. \square

In view of Theorem 4.14, our next aim is to find a condition guaranteeing that the mapping $x^{-\nu} \mapsto X_\nu(x^{-\nu}) = \{y^\nu \in \mathbb{R}^{n_\nu} \mid g^\nu(y^\nu, x^{-\nu}) \leq 0\}$ is lower semicontinuous for all $\nu = 1, \dots, N$. By [48, Theorem 12] lower semicontinuity can be obtained by continuity of g^ν , convexity of the components of g^ν in y^ν for fixed $x^{-\nu}$, and the *Slater condition*, saying that for a given $x^{-\nu} \in \mathbb{R}^{n-n_\nu}$ there exists a $y^\nu \in \mathbb{R}^{n_\nu}$ such that $g^\nu(y^\nu, x^{-\nu}) < 0$ for all $\nu = 1, \dots, N$. The proof of the following lemma is based on this fact.

Lemma 4.15 *Suppose that Assumption 1.2 holds. Then the functions $y_\alpha^\nu, \nu = 1, \dots, N$, and V_α are continuous in $x^* \in W$ provided the Slater condition holds at $X_\nu(x^{*, -\nu})$ for all $\nu = 1, \dots, N$.*

Proof. Let $x^* \in W$ be given such that $X_\nu(x^{*, -\nu})$ satisfies the Slater condition for all $\nu = 1, \dots, N$. By Assumption 1.2 we can apply [48, Theorem 12] and get lower semicontinuity of the point-to-set map $x^{-\nu} \mapsto X_\nu(x^{-\nu})$ at $x^{*, -\nu}$ for all $\nu = 1, \dots, N$. Therefore, also the point-to-set map $x \mapsto \Omega(x)$ is lower semicontinuous at x^* . Hence continuity of y_α , in particular of all components $y_\alpha^\nu, \nu = 1, \dots, N$, and of V_α at x^* follow from Theorem 4.14. \square

Unfortunately, it seems natural that many GNEPs possess points $x^* \in W$ at which the Slater condition is violated for some $\nu = 1, \dots, N$. In the example from Remark 4.1 (a), e.g., for $x^* = (0, 0)$ the set $X_1(0)$ violates the Slater condition.

Whereas the latter example is degenerate, at least in the jointly convex case with a bounded common strategy space $Y = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ the failure of the Slater condition at certain (boundary) points of $W = Y$ cannot be avoided: for any $\nu = 1, \dots, N$ the domain of X_ν ,

$$\text{dom}(X_\nu) = \{x^{-\nu} \in \mathbb{R}^{n-n_\nu} \mid X_\nu(x^{-\nu}) \neq \emptyset\},$$

is closed and bounded as the orthogonal projection of Y to \mathbb{R}^{n-n_ν} and, by the continuity of g , at all boundary points $\bar{x}^{-\nu}$ of $\text{dom}(X_\nu)$ the Slater condition has to be violated in $X_\nu(\bar{x}^{-\nu}) = \{y^\nu \in \mathbb{R}^{n_\nu} \mid g(y^\nu, \bar{x}^{-\nu}) \leq 0\}$. In view of $\bar{x}^{-\nu} \in \text{dom}(X_\nu)$, on the other hand, there exists some $\bar{y}^\nu \in \mathbb{R}^{n_\nu}$ with $(\bar{y}^\nu, \bar{x}^{-\nu}) \in Y$ or, equivalently, $g(\bar{y}^\nu, \bar{x}^{-\nu}) \leq 0$. As $X_\nu(\bar{x}^{-\nu})$ violates the Slater condition, the latter inequality has to be satisfied with equality and, thus, under mild assumptions $(\bar{y}^\nu, \bar{x}^{-\nu})$ is a boundary point of $W = Y$ (e.g., if Y itself satisfies the Slater condition). Note that simple examples show that in general not all boundary points of W correspond to the violation of the Slater condition in some player's strategy space.

In the following we will prove continuity of y_α and V_α at points $x \in W$ also in the case that the Slater condition is violated in one or more strategy spaces $X_\nu(x^{-\nu}), \nu = 1, \dots, N$,

as long as the strategy spaces then collapse to singletons. In view of Theorem 4.2 (a), they then have to coincide with the corresponding set $\{x^\nu\}$.

Note that a violation of the Slater condition does not automatically mean a degenerated strategy space, as one can see in the feasible set from Example 4.3. There we have for $x^2 = (0, 0)$ and arbitrary x^1 that the joint constraint $(x_1^2)^2 + (x_2^2 - x^1)^2 - (x^1)^2 \leq 0$ is satisfied with equality, thus violating the Slater condition. But the set $X_1((0, 0))$ does not collapse to a singleton here, since

$$X_1((0, 0)) = \{x^1 \in [0, 10] \mid (0)^2 + (0 - x^1)^2 - (x^1)^2 \leq 0\} = [0, 10].$$

Theorem 4.16 *Suppose Assumption 1.2 holds, and assume that for each $x^* \in W$ and all $\nu = 1, \dots, N$ the set $X_\nu(x^{*, -\nu})$ either satisfies the Slater condition or coincides with the singleton $\{x^{*, \nu}\}$. Then the functions y_α and V_α are continuous on W .*

Proof. Let $x^* \in W$ be given. In view of Theorem 4.14 we have to show lower semicontinuity of $x \mapsto \Omega(x)$ at x^* , that is, for all sequences $\{x^k\} \subseteq W$ with $\lim_{k \rightarrow \infty} x^k = x^*$ and all $y^* \in \Omega(x^*)$ we have to find a sequence $\{y^k\}$ converging to y^* with $y^k \in \Omega(x^k)$ for all $k \in \mathbb{N}$ sufficiently large. We will define the elements of y^k componentwise for each player $\nu = 1, \dots, N$. For those $\nu \in \{1, \dots, N\}$, where $X_\nu(x^{*, -\nu})$ satisfies the Slater condition, the mapping $x^{-\nu} \mapsto X_\nu(x^{-\nu})$ is lower semicontinuous at $x^{*, -\nu}$ by Lemma 4.15, and hence a sequence $\{y^{k, \nu}\}$ converging to $y^{*, \nu}$ with $y^{k, \nu} \in X_\nu(x^{k, -\nu})$ for all k sufficiently large exists. For all the other $\nu \in \{1, \dots, N\}$ we have $X_\nu(x^{*, -\nu}) = \{x^{*, \nu}\} = \{y^{*, \nu}\}$ by assumption. Defining $y^{k, \nu} := x^{k, \nu}$ we get a sequence $\{y^{k, \nu}\}$ converging to $y^{*, \nu}$ with $y^{k, \nu} = x^{k, \nu} \in X_\nu(x^{k, -\nu})$ by Theorem 4.2 (a), since $x^k \in W$. Therefore $x \mapsto \Omega(x)$ is lower semicontinuous at x^* , since we have a sequence $\{y^k\}$ with $\lim_{k \rightarrow \infty} y^k = y^*$ and $y^{k, \nu} \in X_\nu(x^{k, -\nu})$ for all $\nu = 1, \dots, N$, that is, $y^k \in \Omega(x^k)$ for all k sufficiently large. \square

Hence the optimization reformulation (4.4) of the GNEP is at least a continuous problem under the assumptions of Theorem 4.16. This observation immediately gives the existence result from part (a) of the following note.

Remark 4.17 (a) If the set W is nonempty and bounded and for each $x^* \in W$ and all $\nu = 1, \dots, N$, the set $X_\nu(x^{*, -\nu})$ either satisfies the Slater condition or coincides with the singleton $\{x^{*, \nu}\}$, it is an immediate consequence of the Weierstraß theorem that the optimization problem (4.4) possesses a globally minimal point.

- (b) In general, there are three possible situations which fully describe the relationship between the GNEP and the optimization problem (4.4):
- The GNEP has a solution, and therefore the optimization problem (4.4) also has a solution in view of Theorem 4.2 (with zero as optimal function value).
 - The GNEP has no solution, but the optimization problem (4.4) has a solution (then, necessarily, with a positive optimal function value).
 - Neither the GNEP nor the optimization problem (4.4) have a solution.

Under the assumption from part (a), the last case cannot occur. In this situation, the optimization problem (4.4) therefore characterizes the solvability of a GNEP: The existing minimum of (4.4) is a solution of the GNEP if and only if the optimal function value is zero.

- (c) Here we give an instance for the second case mentioned in (b). In Example 4.10 we have a nonempty and bounded set $W = \{(0, -1)\}$, the single valued set $X_1(-1) = \{0\}$ and the set $X_2(0) = [-2, -1]$, which satisfies the Slater condition, but we do not have a solution, since for the only possible point $(0, -1) \in W$, we get $y_\alpha((0, -1)) = (0, -2) \neq (0, -1)$ for all $\alpha \leq 2$. A short calculation shows that $V_\alpha((0, -1)) = 3 - \alpha/2$ holds for all $\alpha \leq 2$, hence the optimal value of the optimization problem (4.4) is strictly positive.

In our subsequent analysis we will show that (4.4) has, in fact, a piecewise continuously differentiable objective function under some stronger assumptions. This additional smoothness property is highly important from a practical point of view since it implies that several algorithms for nonsmooth optimization problems can be applied to the problem (4.4).

To this end it will be useful to define the function

$$h : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m \quad \text{by} \quad h(x, y) := \begin{pmatrix} g^1(y^1, x^{-1}) \\ \vdots \\ g^N(y^N, x^{-N}) \end{pmatrix},$$

where

$$m := m_1 + \dots + m_N$$

with m_ν being given by Assumption 1.2. This assumption also implies that all component functions h_i are convex as a function of y alone. Furthermore, the function h has the nice property that

$$y \in \Omega(x) \iff h(x, y) \leq 0 \tag{4.11}$$

for any given x .

Now we require some stronger smoothness properties of the defining functions θ_ν and g^ν .

Assumption 4.18 *The functions $\theta_\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g^\nu : \mathbb{R}^n \rightarrow \mathbb{R}^{m_\nu}$ are twice continuously differentiable for all $\nu = 1, \dots, N$.*

Note that Assumption 4.18 implies that the function h is also twice continuously differentiable. Hence $y_\alpha(x)$ is the unique solution of the twice continuously differentiable optimization problem

$$\max_y \Psi_\alpha(x, y) \quad \text{s.t.} \quad h(x, y) \leq 0. \tag{4.12}$$

Let

$$I(x) := \{i \in \{1, \dots, m\} \mid h_i(x, y_\alpha(x)) = 0\}$$

be the set of active constraints. Consider, for a fixed subset $I \subseteq I(x)$, the problem (which has equality constraints only)

$$\max_y \Psi_\alpha(x, y) \quad \text{s.t.} \quad h_i(x, y) = 0 \quad (i \in I). \quad (4.13)$$

Let

$$L_\alpha^I(x, y, \lambda) := -\Psi_\alpha(x, y) + \sum_{i \in I} \lambda_i h_i(x, y)$$

be the Lagrangian of the optimization problem (4.13). Then the KKT system of this problem reads

$$\begin{aligned} \nabla_y L_\alpha^I(x, y, \lambda) = -\nabla_y \Psi_\alpha(x, y) + \sum_{i \in I} \lambda_i \nabla_y h_i(x, y) = 0, \\ h_i(x, y) = 0 \quad \forall i \in I. \end{aligned} \quad (4.14)$$

This can be written as a nonlinear system of equations

$$\Phi_\alpha^I(x, y, \lambda) = 0 \quad \text{with} \quad \Phi_\alpha^I(x, y, \lambda) := \begin{pmatrix} \nabla_y L_\alpha^I(x, y, \lambda) \\ h_I(x, y) \end{pmatrix}, \quad (4.15)$$

where h_I consists of all components h_i of h with $i \in I$. The function Φ_α^I is continuously differentiable since Ψ_α and g are twice continuously differentiable, and its Jacobian is given by

$$J\Phi_\alpha^I(x, y, \lambda) = \begin{pmatrix} \nabla_{yx}^2 L_\alpha^I(x, y, \lambda)^T & \nabla_{yy}^2 L_\alpha^I(x, y, \lambda)^T & \nabla_y h_I(x, y) \\ J_x h_I(x, y) & J_y h_I(x, y) & 0 \end{pmatrix}.$$

Therefore, we obtain

$$\nabla_{(y,\lambda)} \Phi_\alpha^I(x, y, \lambda) = \begin{pmatrix} \nabla_{yy}^2 L_\alpha^I(x, y, \lambda) & \nabla_y h_I(x, y) \\ J_y h_I(x, y) & 0 \end{pmatrix}.$$

Then we have the following result whose proof is standard and therefore omitted.

Lemma 4.19 *Suppose that Assumption 4.18 holds, that $\nabla_{yy}^2 L_\alpha^I(x, y, \lambda)$ is positive definite and that the gradients $\nabla_y h_i(x, y)$ ($i \in I$) are linearly independent. Then $\nabla_{(y,\lambda)} \Phi_\alpha^I(x, y, \lambda)$ is nonsingular.*

Note that the positive definiteness assumption of the Hessian $\nabla_{yy}^2 L_\alpha^I(x, y, \lambda)$ can be relaxed in Lemma 4.19, but that this condition automatically holds in our situation, so we do not really need a weaker assumption here. Furthermore, we stress that the assumed linear independence of the gradients $\nabla_y h_i(x, y)$ ($i \in I$) is a very strong condition for certain index sets I , however, in our subsequent application of Lemma 4.19, we will only consider index sets I where this assumption holds automatically, so this condition is not crucial in our context.

We next introduce another assumption that will be used in order to show that our objective function V_α is a PC^1 mapping.

Assumption 4.20 *The (feasible) constant rank constraint qualification (CRCQ) holds at $x^* \in W$ if there exists a neighbourhood N of x^* such that for every subset $I \subseteq I(x^*)$, the set of gradient vectors*

$$\{\nabla_y h_i(x, y_\alpha(x)) \mid i \in I\}$$

has the same rank (depending on I) for all $x \in N \cap W$.

Note that the previous CRCQ definition requires the same rank only for those $x \in N$ which also belong to the common feasible set W ; this is why we call this assumption the feasible CRCQ, although, in our subsequent discussion, we will simply speak of the CRCQ condition when we refer to Assumption 4.20. This feasible CRCQ has also been used before in [31], for example, where the authors simply call this condition the CRCQ.

The following result is motivated by [65] (see also [47]) and states that both y_α and V_α are piecewise continuously differentiable functions.

Theorem 4.21 *Suppose that Assumptions 1.2 and 4.18 hold, let $x^* \in W$ be given, and suppose that the solution mapping $y_\alpha : W \rightarrow \mathbb{R}^n$ of (4.12) is continuous in a neighbourhood of x^* (see Theorem 4.16 for sufficient conditions). Then there exists a neighbourhood \hat{N} of $x^* \in W$ such that y_α is a PC^1 function on $\hat{N} \cap W$ provided that the (feasible) CRCQ condition from Assumption 4.20 holds at x^* .*

Proof. We divide the proof into four steps.

Step 1: Here we introduce some notation and summarize some preliminary statements that will be useful later on.

First let $x^* \in W$ be fixed such that Assumption 4.20 holds in a neighbourhood N of x^* . Recall that

$$I(x) := \{i \mid h_i(x, y_\alpha(x)) = 0\}$$

for all $x \in N \cap W$. Furthermore, for any such $x \in N \cap W$, let us denote by

$$\mathcal{M}(x) := \{\lambda \in \mathbb{R}^m \mid (y_\alpha(x), \lambda) \text{ is a KKT point of (4.12)}\}$$

the set of all Lagrange multipliers of the optimization problem (4.12). Since CRCQ holds at x^* , it is easy to see that CRCQ also holds for all $x \in W$ sufficiently close to x^* . Without loss of generality, let us say that CRCQ holds for all $x \in N \cap W$ with the same neighbourhood N as before. Then it follows from a result in [49] that the set $\mathcal{M}(x)$ is nonempty for all $x \in N \cap W$. This, in turn, implies that the set

$$\mathcal{B}(x) := \left\{ I \subseteq I(x) \mid \begin{array}{l} \nabla_y h_i(x, y_\alpha(x)) \ (i \in I) \text{ are linearly independent} \\ \text{and } \text{supp}(\lambda) \subseteq I \text{ for some } \lambda \in \mathcal{M}(x) \end{array} \right\}$$

is also nonempty for all x in a sufficiently small neighbourhood of x^* , say, again, for all $x \in N \cap W$ (see [47] for a formal proof), where $\text{supp}(\lambda)$ denotes the support of the nonnegative vector λ , i.e.,

$$\text{supp}(\lambda) := \{i \mid \lambda_i > 0\}.$$

Furthermore, it can be shown that, in a suitable neighbourhood of x^* (which we assume to be N once again), we have $\mathcal{B}(x) \subseteq \mathcal{B}(x^*)$, see, e.g., [65, 47].

Step 2: Here we show that, for every $x \in N \cap W$ and every $I \in \mathcal{B}(x)$, there is a unique multiplier $\lambda_\alpha^I(x) \in \mathcal{M}(x)$ such that $\Phi_\alpha^I(x, y_\alpha(x), \lambda_\alpha^I(x)) = 0$, where N , $\mathcal{M}(x)$, and $\mathcal{B}(x)$ are defined as in Step 1.

To this end, let $x \in N \cap W$ and $I \in \mathcal{B}(x)$ be arbitrarily given. The definition of $\mathcal{B}(x)$ implies that there is a Lagrange multiplier $\lambda_\alpha^I(x) \in \mathcal{M}(x)$ with $\text{supp}(\lambda_\alpha^I(x)) \subseteq I$. Since $(x, y_\alpha(x), \lambda_\alpha^I(x))$ satisfies the KKT conditions of the optimization problem (4.12), $[\lambda_\alpha^I(x)]_i = 0$ for all $i \notin I$, and $h_i(x, y_\alpha(x)) = 0$ for all $i \in I$ (since $I \subseteq I(x)$), it follows that $\Phi_\alpha^I(x, y_\alpha(x), \lambda_\alpha^I(x)) = 0$. Moreover, the linear independence of the gradients $\nabla_y h_i(x, y_\alpha(x))$ for $i \in I$ shows that the multiplier $\lambda_\alpha^I(x)$ is unique.

Step 3: Here we claim that, for any given $x^* \in W$ satisfying Assumption 4.20 and an arbitrary $I \in \mathcal{B}(x^*)$ with corresponding multiplier λ^* , there exist open neighbourhoods $N^I(x^*)$ and $N^I(y_\alpha(x^*), \lambda^*)$ as well as a C^1 -diffeomorphism $(y^I(\cdot), \lambda^I(\cdot)) : N^I(x^*) \rightarrow N^I(y_\alpha(x^*), \lambda^*)$ such that $y^I(x^*) = y_\alpha(x^*)$, $\lambda^I(x^*) = \lambda^*$ and $\Phi_\alpha^I(x, y^I(x), \lambda^I(x)) = 0$ for all $x \in N^I(x^*)$.

To verify this statement, let $x^* \in W$ be given such that the CRCQ holds, choose $I \in \mathcal{B}(x^*)$ arbitrarily, and let $\lambda^* \in \mathcal{M}(x^*)$ with $\text{supp}(\lambda^*) \subseteq I$ be a corresponding multiplier coming from the definition of the set $\mathcal{B}(x^*)$. Now, consider once again the nonlinear system of equations $\Phi_\alpha^I(x, y, \lambda) = 0$ with Φ_α^I being defined in (4.15). The function Φ_α^I is continuously differentiable, and the triple $(x^*, y_\alpha(x^*), \lambda^*)$ satisfies this system. The convexity of θ_ν with respect to x^ν implies that $-\Psi_\alpha^I(x^*, \cdot)$ is strongly convex with respect to the second argument and, therefore, $\nabla_{yy}^2(-\Psi_\alpha^I(x^*, y_\alpha(x^*)))$ is positive definite. Moreover, the convexity of $h_i(x^*, \cdot)$ in the second argument implies the positive semidefiniteness of $\nabla_{yy}^2 h_i(x^*, y_\alpha(x^*))$. Since $\lambda^* \geq 0$, it follows that the Hessian of the Lagrangian L_α^I evaluated in $(x^*, y_\alpha(x^*), \lambda^*)$, i.e., the matrix

$$\nabla_{yy}^2 L_\alpha^I(x^*, y_\alpha(x^*), \lambda^*) = -\nabla_{yy}^2 \Psi_\alpha(x^*, y_\alpha(x^*)) + \sum_{i \in I} \lambda_i^* \nabla_{yy}^2 h_i(x^*, y_\alpha(x^*)),$$

is positive definite. Since, in addition, $\nabla_y h_i(x^*, y_\alpha(x^*))$ ($i \in I$) are linearly independent in view of our choice of $I \in \mathcal{B}(x^*)$, the matrix $\nabla_{(y,\lambda)} \Phi_\alpha^I(x^*, y_\alpha(x^*), \lambda^*)$ is nonsingular by Lemma 4.19. The statement therefore follows from the standard implicit function theorem, where, without loss of generality, we can assume that $N^I(x^*) \subseteq N$.

Step 4: Here we verify the statement of our theorem.

Let $x^* \in W$ satisfying the CRCQ be given. Define $\hat{N} := \bigcap_{I \in \mathcal{B}(x^*)} N^I(x^*)$ with the neighbourhoods $N^I(x^*)$ from Step 3. Since $\mathcal{B}(x^*)$ is a finite set, \hat{N} is a neighbourhood of x^* . Choose $x \in \hat{N} \cap W$ arbitrarily. Step 2 shows that for each $I \subseteq \mathcal{B}(x)$ ($\subseteq \mathcal{B}(x^*)$) there exists a unique multiplier $\lambda_\alpha^I(x) \in \mathcal{M}(x)$ satisfying $\Phi_\alpha^I(x, y_\alpha(x), \lambda_\alpha^I(x)) = 0$. Further, Step 3 guarantees that there exist neighbourhoods $N^I(x^*)$ and $N^I(y_\alpha(x^*), \lambda^*)$ and a C^1 -diffeomorphism $(y^I(\cdot), \lambda^I(\cdot)) : N^I(x^*) \rightarrow N^I(y_\alpha(x^*), \lambda^*)$ such that $\Phi_\alpha^I(x, y^I(x), \lambda^I(x)) = 0$ for all $x \in N^I(x^*)$. In particular, $(y^I(x), \lambda^I(x))$ is the locally unique solution of the system of equations $\Phi_\alpha^I(x, y, \lambda) = 0$. Hence, as soon as we can show that $(y_\alpha(x), \lambda_\alpha^I(x))$ belongs

to the neighbourhood $N^I(y_\alpha(x^*), \lambda^*)$ for all $x \in \hat{N} \cap W$ sufficiently close to x^* , the local uniqueness implies $y_\alpha(x) = y^I(x)$ (for all $I \in \mathcal{B}(x) \subseteq \mathcal{B}(x^*)$).

Suppose this is not true in a sufficiently small neighbourhood. Then there is a sequence $\{x^k\} \subseteq W$ with $\{x^k\} \rightarrow x^*$ and a corresponding sequence of index sets $I^k \in \mathcal{B}(x^k)$ such that

$$(y_\alpha(x^k), \lambda_\alpha^{I^k}(x^k)) \notin N^{I^k}(y_\alpha(x^*), \lambda^*) \quad \text{for all } k \in \mathbb{N}.$$

Since $\mathcal{B}(x^k) \subseteq \mathcal{B}(x^*)$ contains only finitely many index sets, we may assume that I^k is the same index set for all k which we denote by I .

By the continuity of y_α in x^* , we have $y_\alpha(x^k) \rightarrow y_\alpha(x^*)$. On the other hand, for every x^k with associated $y_\alpha(x^k)$ and $\lambda_\alpha^I(x^k)$ from Step 2, we have

$$-\nabla_y \Psi_\alpha(x^k, y_\alpha(x^k)) + \sum_{i \in I} [\lambda_\alpha^I(x^k)]_i \nabla_y h_i(x^k, y_\alpha(x^k)) = 0 \quad (4.16)$$

for all k . The continuity of all functions involved, together with the linear independence of the vectors $\nabla_y h_i(x^*, y_\alpha(x^*))$ (which is a consequence of $I \in \mathcal{B}(x^k) \subseteq \mathcal{B}(x^*)$ and the assumed CRCQ condition) implies that the sequence $\{\lambda_\alpha^I(x^k)\}$ is convergent, say $\{\lambda_\alpha^I(x^k)\} \rightarrow \bar{\lambda}^I$ for some limiting vector $\bar{\lambda}^I$. Taking the limit in (4.16) and using once again the continuity of the solution mapping $y_\alpha(\cdot)$ in x^* then gives

$$-\nabla_y \Psi_\alpha(x^*, y_\alpha(x^*)) + \sum_{i \in I} \bar{\lambda}_i^I \nabla_y h_i(x^*, y_\alpha(x^*)) = 0.$$

Note that the CRCQ condition implies that $\bar{\lambda}^I$ is uniquely defined by this equation and the fact that $\bar{\lambda}_i^I = 0$ for all $i \notin I$. However, by definition, the vector λ^* also satisfies this equation, hence we have $\lambda_\alpha^I(x^k) \rightarrow \lambda^*$. But then it follows that $(y_\alpha(x^k), \lambda_\alpha^I(x^k)) \in N^I(y_\alpha(x^*), \lambda^*)$, and this implies the desired statement. \square

Thus we get the following corollary.

Corollary 4.22 *Suppose that Assumptions 1.2 and 4.18 hold. Moreover, suppose that for each $x^* \in W$ and all $\nu = 1, \dots, N$ the set $X_\nu(x^{*, -\nu})$ either satisfies the Slater condition or coincides with the singleton $\{x^{*, \nu}\}$, and that Assumption 4.20 holds in $x^* \in W$. Then y_α and V_α are PC^1 functions in a neighbourhood of x^* in W .*

Proof. From Corollary 4.16, we obtain the continuity of y_α , whereas Theorem 4.21 implies the PC^1 property of y_α near x^* . Hence the composite mapping $V_\alpha(x) = \Psi_\alpha(x, y_\alpha(x))$ is also continuous and a PC^1 mapping in a neighbourhood of x^* . \square

To close this section, we want to give an example showing that the (feasible) CRCQ from Assumption 4.20, without the Slater condition, does not even imply continuity.

Example 4.23 Consider the following 2-player game where each player controls a single variable.

$$\begin{aligned} \text{Player 1: } & \min_{x^1} (x^1 - 1)^2 \quad \text{s.t.} \quad (x^1)^2(x^2)^2 \leq 0, \\ \text{Player 2: } & \min_{x^2} (x^2 - 1)^2 \quad \text{s.t.} \quad (x^1)^2(x^2)^2 \leq 0. \end{aligned}$$

The common constraint function $g^1(x) \equiv g^2(x) := (x^1)^2(x^2)^2$ satisfies the requirements from Assumptions 1.2 and 4.18 (whereas the Slater condition is violated). We have $W = \{x \in \mathbb{R}^2 \mid x^1 = 0 \text{ or } x^2 = 0\}$. An easy calculation shows that

$$y_\alpha(x) = \frac{1}{1 + \alpha} \begin{cases} (1, 1), & \text{if } x = (0, 0), \\ (1 + \alpha x^1, 0), & \text{if } x^1 \neq 0, x^2 = 0, \\ (0, 1 + \alpha x^2), & \text{if } x^1 = 0, x^2 \neq 0, \\ (0, 0), & \text{if } x^1 \neq 0, x^2 \neq 0. \end{cases}$$

Using this expression for $y_\alpha(x)$ and Theorem 4.2 (e), we deduce that the GNEP has two solutions given by $(0, 1)$ and $(1, 0)$. With the function $h : \mathbb{R}^4 \rightarrow \mathbb{R}^2$, $h(x, y) = \begin{pmatrix} (y^1)^2(x^2)^2 \\ (y^2)^2(x^1)^2 \end{pmatrix}$, we have

$$\nabla_y h(x, y_\alpha(x)) = \begin{pmatrix} 2y_\alpha^1(x)(x^2)^2 & 0 \\ 0 & 2y_\alpha^2(x)(x^1)^2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

for all $x \in \mathbb{R}^2$ which shows, in particular, that the (feasible) CRCQ condition from Assumption 4.20 holds everywhere. But, obviously, the function y_α is not continuous in any point of W except $(-\frac{1}{\alpha}, 0)$ and $(0, -\frac{1}{\alpha})$, in particular, it is discontinuous in the two solutions. Moreover, this function is discontinuous in $(0, 0)$ even if we view it as a mapping on W only. \diamond

4.3. Smoothness Properties of the Unconstrained Reformulation

In contrast to the previous section where a general player convex GNEP was considered, this section deals with GNEPs satisfying Assumption 4.4. Recall, however, that this is still a rather large class of GNEPs including, in particular, the jointly convex ones, cf. the observations from Remark 4.5. Under Assumption 4.4, we have the unconstrained optimization reformulation (4.10) with the objective function $\bar{V}_{\alpha\beta}^c$ from Definition 4.6 and $X = \text{cl}(\text{conv}(W))$ from (4.5).

We will show that the PC^1 property, that was shown for the constrained reformulation in the previous section, also holds for the unconstrained reformulation and, with a further assumption, also the continuity property transfers to the unconstrained reformulation. The proofs of these smoothness properties are similar (though not identical) to the proofs given in the previous section, so that we concentrate on the differences in the proofs without recapitulating all the details.

Our first aim is to obtain a continuity result for $\bar{V}_{\alpha\beta}^c$. Since the projection mapping is continuous the additional term $c\|x - P_X[x]\|^2$ is continuous, hence we only need continuity

of \bar{y}_α for arbitrary $\alpha > 0$ to get this property for \bar{V}_α and $\bar{V}_{\alpha\beta}^c$. In Theorem 4.16 we used the property $x \in \Omega(x)$ for all $x \in W$. The problem occurring here is that the corresponding property $P_X[x] \in \Omega(P_X[x])$ is only valid for $x \in W$ but not necessarily for $x \in X$. To prove a continuity result for the unconstrained reformulation we need the additional assumption of uniform continuity of the functions $g_i^\nu(y^\nu, \cdot) : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}, \nu = 1, \dots, N, i = 1, \dots, m_\nu$ for all $y^\nu \in \mathbb{R}^{n_\nu}$.

Theorem 4.24 *Suppose Assumptions 1.2 and 4.4 hold and further assume that the functions $g_i^\nu(y^\nu, \cdot) : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}, \nu = 1, \dots, N, i = 1, \dots, m_\nu$ are uniformly continuous for all $y^\nu \in \mathbb{R}^{n_\nu}$. Then $\bar{V}_{\alpha\beta}^c$ is continuous in $x^* \in \mathbb{R}^n$ provided the sets $X_\nu(P_X[x^*]^{-\nu})$ are either single-valued or satisfy the Slater condition.*

Proof. As in Theorem 4.14 we obtain continuity of $\bar{V}_{\alpha\beta}^c$ in x^* if the point-to-set mappings $x \mapsto X_\nu(P_X[x]^{-\nu}), \nu = 1, \dots, N$ are closed on \mathbb{R}^n and lower semicontinuous in x^* . The proof of closedness is analogous to the constrained formulation, see Theorem 4.14, and so is the proof of lower semicontinuity in the case where $X_\nu(P_X[x^*]^{-\nu})$ satisfies the Slater condition, see Lemma 4.15.

Hence it remains to show lower semicontinuity when $X_\nu(P_X[x^*]^{-\nu})$ is single valued. Therefore let an $x^* \in \mathbb{R}^n$ and an arbitrary but fixed $\nu \in \{1, \dots, N\}$ be given such that we have a single valued set $X_\nu(P_X[x^*]^{-\nu}) = \{y^{*,\nu}\}$. For a given sequence $\{x^k\} \subseteq \mathbb{R}^n$ with $x^k \rightarrow x^*$ we have to show the existence of a sequence $\{y^{k,\nu}\} \subseteq \mathbb{R}^{n_\nu}$ with $y^{k,\nu} \rightarrow y^{*,\nu}$ and $y^{k,\nu} \in X_\nu(P_X[x^k]^{-\nu})$ for all $k \in \mathbb{N}$ sufficiently large.

Define the function $g_{\max}^\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$g_{\max}^\nu(y^\nu, x^{-\nu}) := \max_{i=1, \dots, m_\nu} g_i^\nu(y^\nu, x^{-\nu}).$$

With the functions g_i^ν also the function g_{\max}^ν is uniformly continuous and convex in y^ν . Further define the set

$$K := \{k \in \mathbb{N} \mid y^{*,\nu} \notin X_\nu(P_X[x^k]^{-\nu})\}.$$

For $k \notin K$ we simply set $y^{k,\nu} := y^{*,\nu}$. If K is finite the proof is already complete. Otherwise consider only the subsequence K . We have $g_{\max}^\nu(y^{*,\nu}, P_X[x^k]^{-\nu}) > 0$ and, since $X_\nu(P_X[x^k]^{-\nu})$ is nonempty by Assumption 4.4, there exists a $w^{k,\nu} \in \mathbb{R}^{n_\nu}$ such that $g_{\max}^\nu(w^{k,\nu}, P_X[x^k]^{-\nu}) \leq 0$. Continuity of g_{\max}^ν implies the existence of an $y^{k,\nu} \in X_\nu(P_X[x^k]^{-\nu})$ on the line segment from $w^{k,\nu}$ to $y^{*,\nu}$ with $g_{\max}^\nu(y^{k,\nu}, P_X[x^k]^{-\nu}) = 0$. It remains to show that $y^{k,\nu}$ converges to $y^{*,\nu}$.

First of all we have $\lim_{k \in K} g_{\max}^\nu(y^{k,\nu}, P_X[x^k]^{-\nu}) = 0$. The uniform continuity of g_{\max}^ν together with the continuity of the projection map and $x^k \rightarrow x^*$ imply

$$\lim_{k \in K} |g_{\max}^\nu(y^{k,\nu}, P_X[x^k]^{-\nu}) - g_{\max}^\nu(y^{k,\nu}, P_X[x^*]^{-\nu})| = 0,$$

and thus we obtain

$$\lim_{k \in K} g_{\max}^\nu(y^{k,\nu}, P_X[x^*]^{-\nu}) = 0. \quad (4.17)$$

4. Optimization Reformulations of GNEPs Using the Nikaido-Isoda Function

We have $g_{\max}^{\nu}(z^{\nu}, P_X[x^*]^{-\nu}) > 0$ for all $z^{\nu} \neq y^{*,\nu}$, because $X_{\nu}(P_X[x^*]^{-\nu}) = \{y^{*,\nu}\}$. Therefore $y^{*,\nu}$ is a strict global minimum of the convex function $g_{\max}^{\nu}(\cdot, P_X[x^*]^{-\nu})$ which implies

$$\varepsilon := \min_{z^{\nu} \in \text{bd}(\mathbb{B}_1(y^{*,\nu}))} g_{\max}^{\nu}(z^{\nu}, P_X[x^*]^{-\nu}) > 0,$$

where $\text{bd}(\mathbb{B}_1(y^{*,\nu})) := \{z^{\nu} \mid \|z^{\nu} - y^{*,\nu}\| = 1\}$ is the boundary of the ball $B_1(y^{*,\nu})$ with centre $y^{*,\nu}$ and radius 1. With the convexity of $g_{\max}^{\nu}(\cdot, P_X[x^*]^{-\nu})$ we get

$$g_{\max}^{\nu}(y^{\nu}, P_X[x^*]^{-\nu}) \geq \varepsilon \quad \text{for all } y^{\nu} \notin B_1(y^{*,\nu}).$$

This together with (4.17) shows that $\{y^{k,\nu}\} \in B_1(y^{*,\nu})$ for all $k \in K$ sufficiently large. But this implies boundedness of the entire sequence $\{y^{k,\nu}\}$ and thus the existence of an accumulation point \hat{y}^{ν} . Closedness of the point-to-set mapping $x \mapsto X_{\nu}(P_X[x]^{-\nu})$ therefore shows $\hat{y}^{\nu} \in X_{\nu}(P_X[x^*]^{-\nu}) = \{y^{*,\nu}\}$. Since this is true for all accumulation points, we have convergence of the sequence $\{y^{k,\nu}\}$ to $y^{*,\nu}$, which completes the proof. \square

Our next aim is to show that the function $\bar{V}_{\alpha\beta}^c$ is a PC^1 mapping under suitably adopted assumptions. To this end, we first define the function

$$\bar{h} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^m \quad \text{by} \quad \bar{h}(x, y) := \begin{pmatrix} g^1(y^1, (P_X[x])^{-1}) \\ \vdots \\ g^N(y^N, (P_X[x])^{-N}) \end{pmatrix}.$$

This function will play the role of the mapping h from Section 4.2. In particular, it has the corresponding property that, for any given x ,

$$y \in \Omega(P_X[x]) \iff \bar{h}(x, y) \leq 0. \quad (4.18)$$

This implies that $\bar{y}_{\alpha}(x)$ is the unique solution of

$$\max_y \Psi_{\alpha}(x, y) \quad \text{s.t.} \quad \bar{h}(x, y) \leq 0. \quad (4.19)$$

Note, however, that (in contrast to the function h) the function \bar{h} is not differentiable in general (even if all g^{ν} are differentiable) due to the projection term inside the definition of \bar{h} . This causes some technical difficulties in generalizing the PC^1 property to the unconstrained reformulation. However \bar{h} is a PC^1 mapping if all g^{ν} are smooth and the projection mapping is PC^1 . The latter holds in view of [63] under the smoothness conditions of Assumption 4.18 and a constant rank constraint qualification in a version that we define next.

Assumption 4.25 *The constant rank constraint qualification (CRCQ) holds at $x^* \in \mathbb{R}^n$ if there exists a neighbourhood N of x^* such that, for every subset $I \subseteq \bar{I}(x^*) := \{i \mid \bar{h}_i(x^*, \bar{y}_{\alpha}(x^*)) = 0\}$, the set of gradient vectors*

$$\{\nabla_y \bar{h}_i(x, \bar{y}_{\alpha}(x)) \mid i \in I\}$$

has the same rank (depending on I) for all $x \in N$.

Assumption 4.25 is slightly stronger than the feasible CRCQ from Assumption 4.20 since a full-dimensional neighbourhood N of x^* is used, whereas in Assumption 4.20 only a feasible neighbourhood of x^* is used.

Consider the optimization problem (4.19) once again. Let

$$\bar{I}(x) := \{i \in \{1, \dots, m\} \mid \bar{h}_i(x, \bar{y}_\alpha(x)) = 0\}$$

be the set of active inequality constraints. Consider, for a fixed subset $I \subseteq \bar{I}(x)$, the equality constrained problem

$$\max_y \Psi(x, y) \quad \text{s.t.} \quad \bar{h}_i(x, y) = 0 \quad (i \in I). \quad (4.20)$$

Let

$$\bar{L}_\alpha^I(x, y, \lambda) := -\Psi_\alpha(x, y) + \sum_{i \in I} \lambda_i \bar{h}_i(x, y)$$

be the corresponding Lagrangian. Then the KKT conditions of (4.20) are equivalent to the nonlinear system of equations

$$\bar{\Phi}_\alpha^I(x, y, \lambda) = 0 \quad \text{with} \quad \bar{\Phi}_\alpha^I(x, y, \lambda) := \begin{pmatrix} \nabla_y \bar{L}_\alpha^I(x, y, \lambda) \\ \bar{h}_I(x, y) \end{pmatrix}. \quad (4.21)$$

In the proof of Theorem 4.21 (Step 3), we applied the implicit function theorem to the mapping Φ_α^I from the previous section. In contrast to Φ_α^I , however, $\bar{\Phi}_\alpha^I$ is not differentiable everywhere, hence the standard implicit function theorem cannot be used in the current situation. But under suitable assumptions, including the CRCQ condition, the projection map and therefore also $\bar{\Phi}_\alpha^I$ is a PC^1 mapping. Hence we need an implicit function theorem for PC^1 equations.

Theorem 4.26 *Assume $H : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a PC^1 function in a neighbourhood of (\bar{x}, \bar{y}) with $H(\bar{x}, \bar{y}) = 0$ and all matrices in $\pi_y \partial H(\bar{x}, \bar{y})$ have the same nonzero determinantal sign. Then there exists an open neighbourhood U of \bar{x} and a function $g : U \rightarrow \mathbb{R}^n$ which is a PC^1 function on U such that $g(\bar{x}) = \bar{y}$ and $H(x, g(x)) = 0$ for all $x \in U$.*

Proof. Define the mapping $F : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^m \times \mathbb{R}^n$ by

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

Then we have

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

and all elements $A \in \partial F(\bar{x}, \bar{y})$ have the same nonzero determinantal sign, because the matrices in $\pi_y \partial H(\bar{x}, \bar{y})$ have. Therefore they are all nonsingular. With H also the function F is a PC^1 function in a neighbourhood of (\bar{x}, \bar{y}) and hence in particular locally Lipschitz continuous. Clarke's inverse function theorem [12, Theorem 7.1.1] implies the existence

of open neighbourhoods V of (\bar{x}, \bar{y}) and W of $(0, 0) = F(\bar{x}, \bar{y})$ such that $F : V \rightarrow W$ is a homeomorphism and the local inverse $G : W \rightarrow V$ is a locally Lipschitz continuous function. Moreover, by [31, Theorem 4.6.5], G is a PC^1 function, because F is a locally Lipschitz continuous homeomorphism and a PC^1 function. Define the set

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

U is nonempty and open (in \mathbb{R}^n) since $(0, 0) \in W$ and W is open. Let $x \in U$ be arbitrarily given. Then we have $(x - \bar{x}, 0) \in W$ and hence, by the definition of a homeomorphism, we obtain the existence of a unique y with $(x, y) \in V$ and $F(x, y) = (x - \bar{x}, 0)$. Thus we have $H(x, y) = 0$. Since y depends on x , we write $y =: g(x)$ which defines a function $g : U \rightarrow \mathbb{R}^n$ such that $H(x, g(x)) = 0$ for each $x \in U$. Therefore we have

$$F(x, g(x)) = \begin{pmatrix} x - \bar{x} \\ H(x, g(x)) \end{pmatrix} = \begin{pmatrix} x - \bar{x} \\ 0 \end{pmatrix}$$

for all $x \in U$. Applying the inverse function G on both sides, we obtain

$$(x, g(x)) = G(x - \bar{x}, 0)$$

for all $x \in U$. Since g coincides with some component functions of the PC^1 function G , it is a PC^1 function itself which completes the proof. \square

Now we are in a position to generalize Theorem 4.21 to the unconstrained optimization reformulation.

Theorem 4.27 *Suppose that Assumptions 1.2, 4.4, and 4.18 hold. Furthermore, suppose that $x^* \in \mathbb{R}^n$ is such that the CRCQ condition from Assumption 4.25 is satisfied at x^* and the solution mapping $\bar{y}_\alpha : \mathbb{R}^n \rightarrow \mathbb{R}^n$ of (4.19) is continuous in a neighbourhood of x^* (see Theorem 4.24 for a sufficient condition). Then \bar{y}_α is a PC^1 function in a neighbourhood of x^* .*

Proof. We follow the proof of Theorem 4.21 by dividing the proof into four steps. Rather than giving all the details we more or less only mention the differences.

Step 1: Similar to the discussion in Section 4.2, let us introduce the sets

$$\begin{aligned} \bar{\mathcal{M}}(x) &:= \{\lambda \in \mathbb{R}^m \mid (\bar{y}_\alpha(x), \lambda) \text{ is a KKT point of (4.19)}\} && \text{and} \\ \bar{\mathcal{B}}(x) &:= \{I \subseteq \bar{I}(x) \mid \nabla_y \bar{h}_i(x, \bar{y}_\alpha(x)) (i \in I) \text{ are linearly independent} \\ &&& \text{and } \text{supp}(\lambda) \subseteq I \text{ for some } \lambda \in \bar{\mathcal{M}}(x)\}. \end{aligned}$$

Then Assumption 4.25 implies that there is a neighbourhood N of x^* such that $\bar{\mathcal{M}}(x) \neq \emptyset$, $\bar{\mathcal{B}}(x) \neq \emptyset$ and $\bar{\mathcal{B}}(x) \subseteq \bar{\mathcal{B}}(x^*)$ for all $x \in N$.

Step 2: Using the notation of (4.19), \bar{L}_α^I and $\bar{\Phi}_\alpha^I$, it follows as in the proof of Theorem 4.21 that, for every $x \in N$ and every $I \in \bar{\mathcal{B}}(x)$, there is a unique multiplier $\lambda_\alpha^I(x) \in \bar{\mathcal{M}}(x)$ such

that $\bar{\Phi}_\alpha^I(x, \bar{y}_\alpha(x), \lambda_\alpha^I(x)) = 0$, where N , $\bar{\mathcal{M}}(x)$, and $\bar{\mathcal{B}}(x)$ are the sets defined in Step 1.

Step 3: Here we have the main difference to the proof of Theorem 4.21 since the mapping $\bar{\Phi}_\alpha^I$ defined in Step 2 is only a PC^1 function, but not continuously differentiable (as the mapping Φ_α^I). Therefore, we have to use an implicit function theorem for PC^1 functions instead of the standard implicit function theorem. Let any $x^* \in \mathbb{R}^n$ satisfying Assumption 4.25 and an arbitrary $I \in \bar{\mathcal{B}}(x^*)$ with corresponding multiplier λ^* be given. Since $\bar{\Phi}_\alpha^I(x, y, \lambda)$ is continuously differentiable with respect to y and λ , it follows that $\pi_{(y, \lambda)} \partial \bar{\Phi}_\alpha^I(x^*, \bar{y}_\alpha(x^*), \lambda^*)$ has only one element, whose nonsingularity can be shown as in the proof of Theorem 4.21. In particular, the same nonzero orientation of all the elements is guaranteed. Using the PC^1 implicit function Theorem 4.26 we get the existence of open neighbourhoods $N^I(x^*)$ and $N^I(\bar{y}_\alpha(x^*), \lambda^*)$ as well as a PC^1 function $(y^I(\cdot), \lambda^I(\cdot)) : N^I(x^*) \rightarrow N^I(\bar{y}_\alpha(x^*), \lambda^*)$ such that $y^I(x^*) = \bar{y}_\alpha(x^*)$, $\lambda^I(x^*) = \lambda^*$ and $\bar{\Phi}_\alpha^I(x, y^I(x), \lambda^I(x)) = 0$ for all $x \in N^I(x^*)$.

Step 4: Repeating the arguments from Step 4 of the proof of Theorem 4.21, we obtain $\bar{y}_\alpha(x) \in \{y^I(x) \mid I \in \bar{\mathcal{B}}(x^*)\}$ for all x in a sufficiently small neighbourhood of x^* . Since all y^I are PC^1 functions, it follows that also \bar{y}_α is a PC^1 mapping in a neighbourhood of any x^* satisfying the CRCQ condition from Assumption 4.25. \square

Altogether, we get the following corollary.

Corollary 4.28 *Suppose that Assumptions 1.2, 4.4 and 4.18 hold. Moreover, suppose that the functions $g_i^\nu(y^\nu, \cdot) : \mathbb{R}^{n-n_\nu} \rightarrow \mathbb{R}$ are uniformly continuous for all $y^\nu \in \mathbb{R}^{n_\nu}$, the CRCQ Assumption 4.25 holds in $x^* \in \mathbb{R}^n$ and that the sets $X_\nu(P_X[x]^{-\nu})$, $\nu = 1, \dots, N$ either satisfy the Slater condition or coincide with a singleton for all x sufficiently close to x^* . Then $\bar{V}_{\alpha\beta}^c$ is a PC^1 function in a neighbourhood of x^* .*

Proof. Since the projection mapping has PC^1 property, the additional term $c\|x - P_X[x]\|^2$ also has. From Theorem 4.24 we obtain the continuity of \bar{y}_α . Theorem 4.27 therefore implies the PC^1 property of \bar{y}_α near x^* satisfying the CRCQ condition from Assumption 4.25. Hence the composite mapping $\bar{V}_\alpha(x) = \Psi_\alpha(x, \bar{y}_\alpha(x))$ and therefore also $\bar{V}_{\alpha\beta}^c = \bar{V}_\alpha(x) - \bar{V}_\beta(x) + c\|x - P_X[x]\|^2$ are PC^1 mappings in a neighbourhood of x^* . \square

Being a PC^1 mapping, it follows that $\bar{V}_{\alpha\beta}^c$ is, in particular, directionally differentiable, locally Lipschitz continuous and semismooth, cf. [10].

4.4. Stationarity

In the previous sections a constrained reformulation of all solutions of a player convex GNEP and an unconstrained reformulation for a class of general player convex GNEPs which, in particular, includes the jointly convex case was given. Both reformulations characterize all solutions of general player convex GNEPs as solutions of optimization problems. These problems have PC^1 objective functions, which allows the application of

nonsmooth optimization software for finding a solution of it. However, most algorithms can only find Clarke stationary points which may correspond to local minima and therefore are not necessarily solutions of the GNEP. The following example illustrates this situation.

Example 4.29 Consider once again Example 2.2, a jointly convex 2-player game defined via

$$\theta_1(x) := \frac{1}{2}(x^1 + 2)^2, \quad \theta_2(x) := \frac{1}{2}(x^2 + 2)^2, \quad \text{and}$$

$$X := \{x \in \mathbb{R}^2 \mid 0 \leq x^1 \leq 2, x^2 - x^1 \leq 0, x^1 - x^2 - 1 \leq 0\}.$$

A simple calculation shows that, for all $\alpha \in (0, 1]$ and all $x \in X$, we have

$$\bar{y}_\alpha(x) = (\max\{0, x^2\}, x^1 - 1).$$

It was shown in Example 2.2 that the only solution of this GNEP is given by $(0, -1)$. Taking $0 < \alpha < \beta \leq 1$, we obtain from the previous observation that $\bar{y}_\alpha(x) = \bar{y}_\beta(x)$ and, therefore, using $c = 0$ since we have a jointly convex GNEP,

$$\bar{V}_{\alpha\beta}^0(x) = \frac{\beta - \alpha}{2} \|x - \bar{y}_\alpha(x)\|^2.$$

Thus we have for all $x^2 \geq 0$

$$\bar{V}_{\alpha\beta}^0(x) = \frac{\beta - \alpha}{2} ((x^1 - x^2)^2 + (x^2 - x^1 + 1)^2) = (\beta - \alpha) \left(\left(x^2 - x^1 + \frac{1}{2} \right)^2 + \frac{1}{4} \right).$$

Hence we see, that for all $x \in X$ with $x^2 \geq 0$ and $x^2 - x^1 = -\frac{1}{2}$, the function $\bar{V}_{\alpha\beta}^0$ has local minima with function value $\frac{\beta - \alpha}{4} > 0$ and hence stationary points, which are no solutions of the GNEP. \diamond

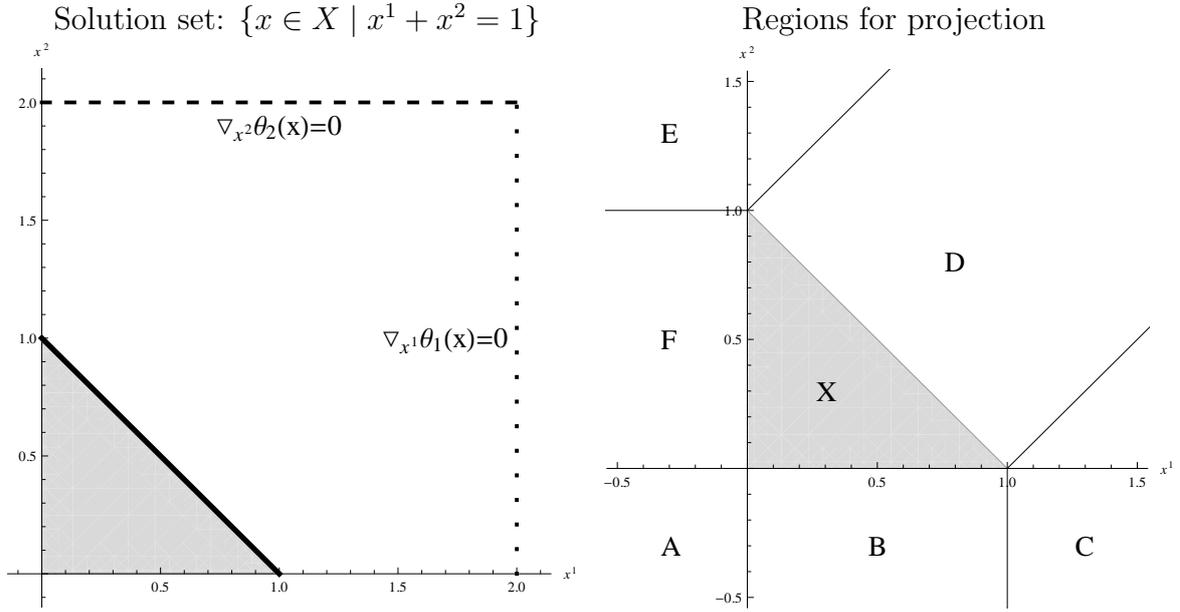
In view of this example and the application of an algorithm searching for Clarke stationary points, a result giving suitable conditions for a stationary point to be a solution of the GNEP is important. Unfortunately such a result has not been found yet, and this topic is nontrivial since standard techniques provide overestimates for Clarke's generalized Jacobian of $\bar{V}_{\alpha\beta}^c$ which are, in general even in simple examples, by far too large. This is due to the fact that the objective function involves the difference of two closely related nondifferentiable terms. A lengthy example illustrating this is the following.

Example 4.30 Consider the jointly convex 2-player game, where each player controls a single variable, the cost functions are

$$\theta_1(x) = \frac{1}{2}(x^1 - 2)^2 \quad \text{and} \quad \theta_2(x) = \frac{1}{2}(x^2 - 2)^2,$$

and the common strategy set is

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


 Figure 4.4.: Solution set and regions for calculations of $P_X[x]$

Theorem 2.1 yields the solution set

$$\{x \in X \mid x^1 + x^2 = 1\},$$

see Figure 4.4 on the left side. Here we have

$$\Psi_\gamma(x, y) = \frac{1}{2}(x^1 - 2)^2 + \frac{1}{2}(x^2 - 2)^2 - \frac{1}{2}(y^1 - 2)^2 - \frac{1}{2}(y^2 - 2)^2 - \frac{\gamma}{2}\|x - y\|^2,$$

and for $\nu = 1, 2$

$$\bar{y}_\gamma^\nu(x) = \begin{cases} 0, & \text{for } x^\nu \leq -\frac{2}{\gamma}, \\ \frac{2+\gamma x^\nu}{1+\gamma}, & \text{for } -\frac{2}{\gamma} < x^\nu < \frac{1}{\gamma}(\gamma - 1 - (1 + \gamma)(P_X[x])^{-\nu}), \\ 1 - (P_X[x])^{-\nu}, & \text{for } x^\nu \geq \frac{1}{\gamma}(\gamma - 1 - (1 + \gamma)(P_X[x])^{-\nu}). \end{cases}$$

We only discuss the main case, where we have

$$\bar{y}_\gamma(x) = (1 - (P_X[x])^2, 1 - (P_X[x])^1) \quad \text{for } \gamma = \alpha, \beta,$$

because this case includes the whole set X for $\gamma < 1$. Therein we obtain

$$\bar{V}_{\alpha\beta}^0(x) = \frac{1}{2}(\beta - \alpha) (1 - x^2 - (P_X[x])^1)^2 + \frac{1}{2}(\beta - \alpha) (1 - x^1 - (P_X[x])^2)^2.$$

Depending on the projection $P_X[x]$ we computed $\bar{V}_{\alpha\beta}^0$ and $\nabla \bar{V}_{\alpha\beta}^0$, see Table 4.1 and Figure 4.4 on the right hand side.

4. Optimization Reformulations of GNEPs Using the Nikaido-Isoda Function

Region	$P_X[x]$	$\frac{1}{\beta-\alpha}\bar{V}_{\alpha\beta}^0(x)$	$\frac{1}{\beta-\alpha}\nabla\bar{V}_{\alpha\beta}^0(x)$
int(X)	x	$(1-x^1-x^2)^2$	$(2(x^1+x^2-1), 2(x^1+x^2-1))$
int(A)	$(0, 0)$	$\frac{1}{2}(1-x^2)^2 + \frac{1}{2}(1-x^1)^2$	(x^1-1, x^2-1)
int(B)	$(x^1, 0)$	$\frac{1}{2}(1-x^1-x^2)^2 + \frac{1}{2}(1-x^1)^2$	$(2x^1+x^2-2, x^1+x^2-1)$
int(C)	$(1, 0)$	$\frac{1}{2}(x^2)^2 + \frac{1}{2}(1-x^1)^2$	(x^1-1, x^2)
int(D)	$(\frac{1+x^1-x^2}{2}, \frac{1-x^1+x^2}{2})$	$\frac{1}{4}(1-x^1-x^2)^2$	$(\frac{1}{2}(x^1+x^2-1), \frac{1}{2}(x^1+x^2-1))$
int(E)	$(0, 1)$	$\frac{1}{2}(x^1)^2 + \frac{1}{2}(1-x^2)^2$	(x^1, x^2-1)
int(F)	$(0, x^2)$	$\frac{1}{2}(1-x^2)^2 + \frac{1}{2}(1-x^1-x^2)^2$	(x^1+x^2-1, x^1+2x^2-2)

Table 4.1.: Calculation of $\bar{V}_{\alpha\beta}^0$ and $\nabla\bar{V}_{\alpha\beta}^0$

Using the results of Table 4.1 we see, that $\bar{V}_{\alpha\beta}^0$ is continuous differentiable at the interior of each region, and Clarke's generalized Jacobian, which is then equal to the gradient, is unequal to zero there. On the border of different regions the generalized Jacobian can be computed as the convex hull of the limits of the gradients from the adjacent regions, see Table 4.2.

Region	$\frac{1}{\beta-\alpha}\partial\bar{V}_{\alpha\beta}^0(x)$
$(0,0)$	$\text{conv}\{(-2, -2), (-1, -1), (-2, -1), (-1, -2)\} = [-2, -1]^2$
$(1,0)$	$\text{conv}\{(0, 0), (0, 0), (0, 0), (0, 0)\} = \{(0, 0)\}$
$(0,1)$	$\text{conv}\{(0, 0), (0, 0), (0, 0), (0, 0)\} = \{(0, 0)\}$
$D \cap X$	$\text{conv}\{(0, 0), (0, 0)\} = \{(0, 0)\}$
$B \cap X$	$\text{conv}\{(2x^1-2, 2x^1-2), (2x^1-2, x^1-1)\} = \{2x^1-2\} \times [2x^1-2, x^1-1]$
$F \cap X$	$\text{conv}\{(2x^2-2, 2x^2-2), (x^2-1, 2x^2-2)\} = [2x^2-2, x^2-1] \times \{2x^2-2\}$
$A \cap B$	$\text{conv}\{(-1, x^2-1), (x^2-2, x^2-1)\} = [x^2-2, -1] \times \{x^2-1\}$
$B \cap C$	$\text{conv}\{(x^2, x^2), (0, x^2)\} = [x^2, 0] \times \{x^2\}$
$C \cap D$	$\text{conv}\{(x^2, x^2), (x^2, x^2)\} = \{(x^2, x^2)\}$
$D \cap E$	$\text{conv}\{(x^1, x^1), (x^1, x^1)\} = \{(x^1, x^1)\}$
$E \cap F$	$\text{conv}\{(x^1, 0), (x^1, x^1)\} = \{x^1\} \times [x^1, 0]$
$A \cap F$	$\text{conv}\{(x^1-1, -1), (x^1-1, x^1-2)\} = \{x^1-1\} \times [x^1-2, -1]$

Table 4.2.: Clarke's generalized Jacobian

Table 4.2 shows that in the considered region, which in particular contains X , $0 \in \partial\bar{V}_{\alpha\beta}^0(x)$ is equivalent to $x \in \{x \in X \mid x^1 + x^2 = 1\}$, which is exactly the set of generalized Nash equilibria.

Now consider the upper approximation for Clarke's generalized Jacobian from [52, Theorem 4.9]:

$$\partial(-\bar{V}_\gamma(x^*)) \subseteq \{s \in \mathbb{R}^n \mid (s, 0) \in -\nabla\Psi_\gamma(x^*, \bar{y}_\gamma(x^*)) + N_{\text{graph}(\Omega)}(x^*, \bar{y}_\gamma(x^*))\},$$

with the Clarke normal cone $N_{\text{graph}(\Omega)}(x^*, \bar{y}_\gamma(x^*))$ on the graph of Ω in $(x^*, \bar{y}_\gamma(x^*))$. We have

$$\text{graph}(\Omega) = \{(x, y) \mid y \in \Omega(x)\} = \{(x, y) \mid A(x^1, x^2, y^1, y^2)^T - b \leq 0\}$$

with

$$A := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \quad \text{and} \quad b := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Further if $I(x) := \{i \in \{1, \dots, 6\} \mid A_i(x^1, x^2, y^1, y^2)^T - b_i = 0\}$ denotes the set of all active inequalities and A_i the i -th row of A , Clarke's normal cone is

$$N_{\text{graph}(\Omega)}(x^*, \bar{y}_\gamma(x^*)) = \left\{ \sum_{i \in I(x^*)} \lambda_i A_i^T \mid \lambda_i \geq 0 \right\}.$$

For the subsequent calculations we need the gradient of Ψ_γ on X , that is

$$\nabla_{(x,y)} \Psi_\gamma(x, \bar{y}_\gamma(x)) = \begin{pmatrix} x^1 - 2 + \gamma(1 - x^1 - x^2) \\ x^2 - 2 + \gamma(1 - x^1 - x^2) \\ x^2 + 1 - \gamma(1 - x^1 - x^2) \\ x^1 + 1 - \gamma(1 - x^1 - x^2) \end{pmatrix}.$$

Since we only want to discuss $x \in X$, we consider the following cases:

- (a) For $x \in X$ with $x^1 \neq 0, x^2 \neq 0$ we have $I(x) = \{5, 6\}$, A_5 and A_6 are linearly independent, and therefore the multipliers λ_5, λ_6 are unique and we get

$$\begin{pmatrix} \bar{\lambda}_6 \\ \bar{\lambda}_5 \end{pmatrix} = \nabla_y \Psi_\gamma(x, \bar{y}_\gamma(x)) = \begin{pmatrix} x^2 + 1 - \gamma(1 - x^1 - x^2) \\ x^1 + 1 - \gamma(1 - x^1 - x^2) \end{pmatrix}.$$

Hence the upper estimate is single-valued (implying differentiability)

$$\partial(-\bar{V}_\gamma(x)) = \left\{ -\nabla_x \Psi_\gamma(x, \bar{y}_\gamma(x)) + \begin{pmatrix} \bar{\lambda}_5 \\ \bar{\lambda}_6 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 3 - 2\gamma(1 - x^1 - x^2) \\ 3 - 2\gamma(1 - x^1 - x^2) \end{pmatrix} \right\}.$$

- (b) For $x = (0, 0)$ we have $I(x) = \{1, 2, 5, 6\}$, λ_5, λ_6 are unique, $\lambda_1, \lambda_2 \geq 0$ are arbitrary and this implies $\partial(-\bar{V}_\gamma(x)) \subseteq \begin{pmatrix} 3 - 2\gamma \\ 3 - 2\gamma \end{pmatrix} + \begin{pmatrix} \mathbb{R}_- \\ \mathbb{R}_- \end{pmatrix}$.

- (c) For $x = (1, 0)$ we have $I(x) = \{2, 4, 5, 6\}$, $\lambda_6 = 1, \lambda_5 - \lambda_4 = 2$ and $\lambda_2 \geq 0$ is arbitrary. This implies

$$\partial(-\bar{V}_\gamma(x)) \subseteq \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} \lambda_5 \\ 1 - \lambda_2 \end{pmatrix} \mid \lambda_5 \geq 2, \lambda_2 \geq 0 \right\} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} \mathbb{R}_+ \\ \mathbb{R}_- \end{pmatrix}.$$

(d) Analogous for $x = (0, 1)$ we have $\partial(-\bar{V}_\gamma(x)) \subseteq \begin{pmatrix} 3 \\ 3 \end{pmatrix} + \begin{pmatrix} \mathbb{R}_- \\ \mathbb{R}_+ \end{pmatrix}$.

(e) For $x = (x_1, 0)$ with $0 < x_1 < 1$ we have $I(x) = \{2, 5, 6\}$, λ_5, λ_6 are unique, $\lambda_2 \geq 0$ is arbitrary and $\partial(-\bar{V}_\gamma(x)) \subseteq \begin{pmatrix} 3 - 2\gamma(1 - x^1) \\ 3 - 2\gamma(1 - x^1) \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbb{R}_- \end{pmatrix}$.

(f) Analogous for $x = (0, x_2)$ with $0 < x_2 < 1$: $\partial(-\bar{V}_\gamma(x)) \subseteq \begin{pmatrix} 3 - 2\gamma(1 - x^2) \\ 3 - 2\gamma(1 - x^2) \end{pmatrix} + \begin{pmatrix} \mathbb{R}_- \\ 0 \end{pmatrix}$.

Using $\partial\bar{V}_{\alpha\beta}^0(x) \subseteq \partial(-\bar{V}_\beta(x)) - \partial(-\bar{V}_\alpha(x))$ we get

$$\partial\bar{V}_{\alpha\beta}^0(x) \subseteq (\beta - \alpha) \left\{ \begin{array}{ll} (2(x^1 + x^2 - 1), 2(x^1 + x^2 - 1)), & x \in X, x^1 \neq 0, x^2 \neq 0, \\ (\mathbb{R}, \mathbb{R}), & x = (0, 0), \\ (\mathbb{R}, \mathbb{R}), & x = (1, 0), \\ (\mathbb{R}, \mathbb{R}), & x = (0, 1), \\ (2x^1 - 2, 2x^1 - 2) + (0, \mathbb{R}), & x \in X, x^2 = 0, x^1 \notin \{0, 1\}, \\ (2x^2 - 2, 2x^2 - 2) + (\mathbb{R}, 0), & x \in X, x^1 = 0, x^2 \notin \{0, 1\}. \end{array} \right.$$

The upper estimate for Clarke's generalized Jacobian of $\bar{V}_{\alpha\beta}^0$ contains 0 not only for the set of global minima $\{x \in X \mid x^1 + x^2 = 1\}$ but also for the point $(0, 0)$. This shows that even in simple cases like the one discussed above, the known upper estimates from literature are not good enough for an appropriate condition for stationary points to be global minima. \diamond

This example indicates that known upper estimates for Clarke's generalized Jacobian might be not good enough to get a condition for stationary points to be global minima. Thus it is a challenging problem to obtain such a result, since probably new estimates for the generalized Jacobian have to be developed.

5. Solving the KKT System of a GNEP

While for the previous considerations the Nikaido-Isoda function played a central role, we are now going a completely different way by using the KKT conditions of a GNEP. Let \bar{x} be a solution of the GNEP. Assuming any standard constraint qualification holds, the following KKT conditions will be satisfied for every player $\nu = 1, \dots, N$:

$$\begin{aligned} \nabla_{x^\nu} \theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) + \sum_{i=1}^{m_\nu} \lambda_i^\nu \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) &= 0, \\ \lambda_i^\nu \geq 0, \quad g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq 0, \quad \lambda_i^\nu g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) &= 0 \quad \forall i = 1, \dots, m_\nu, \end{aligned} \quad (5.1)$$

where $\lambda^\nu \in \mathbb{R}^{m_\nu}$ is the vector of Lagrange multipliers of player ν . Vice versa, recalling that the player's problems are convex by the Assumption 1.2, we have: if a point \bar{x} together with a suitable vector of multipliers $\lambda := (\lambda^1, \lambda^2, \dots, \lambda^N)$ satisfies the KKT conditions (5.1) for every $\nu = 1, \dots, N$, then \bar{x} is a solution of the GNEP. Thus one may try to solve the GNEP by solving the system obtained by concatenating the N systems (5.1). In order to use a more compact notation, we introduce some further definitions. We denote by

$$L^\nu(x, \lambda^\nu) := \theta_\nu(x^\nu, x^{-\nu}) + \sum_{i=1}^{m_\nu} \lambda_i^\nu g_i^\nu(x^\nu, x^{-\nu})$$

the Lagrangian of player ν . If we set

$$F(x, \lambda) := (\nabla_{x^\nu} L^\nu(x, \lambda^\nu))_{\nu=1}^N \quad \text{and} \quad g(x) := (g^\nu(x))_{\nu=1}^N,$$

the concatenated KKT system can be written as

$$F(x, \lambda) = 0, \quad \lambda \geq 0, \quad g(x) \leq 0, \quad \lambda^T g(x) = 0. \quad (5.2)$$

There exists a lot of literature on reformulating the KKT conditions of an optimization problem or of a variational inequality as a (constrained) system of equations or as a (constrained) optimization problem; and these reformulations are the basis for many efficient algorithms for the solution of these problems, see [31]. However up to now there are no meaningful results showing if and when these techniques will lead to useful results in the case of the KKT system of a GNEP. Therefore the main aim of this section is to derive theoretical results related to system (5.2) and to find some new solution methods. More specifically, we will analyse a merit function approach and an interior point method for the solution of the GNEP KKT system (5.2). These two approaches can be viewed as natural extensions of the corresponding methods for the solution of the KKT system of an optimization problem. We will explore the theoretical properties of the methods here. The presented results, except the finite termination property for the interior point algorithm, were published in [19].

5.1. Merit Function Approach

In order to solve the concatenated KKT system, an approach that has been very widely used in the optimization and VI communities and that has led to very useful developments, see [23, 31], is to reduce it to a system of equations through the use of a *complementarity function*. More specifically, let $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be any function such that $\phi(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$, and $ab = 0$. Then, it one can immediately see that the concatenated KKT system can be rewritten as

$$F(x, \lambda) = 0, \quad \Phi(x, \lambda) = 0,$$

where

$$\Phi(x, \lambda) := \begin{pmatrix} \phi(\lambda_1^1, -g_1^1(x)) \\ \vdots \\ \phi(\lambda_{m_1}^1, -g_{m_1}^1(x)) \\ \phi(\lambda_1^2, -g_1^2(x)) \\ \vdots \\ \phi(\lambda_{m_N}^N, -g_{m_N}^N(x)) \end{pmatrix} \in \mathbb{R}^m.$$

There exist many types of complementarity functions ϕ , but the two most prominent ones are the minimum-function $\phi(a, b) := \min\{a, b\}$ and the Fischer-Burmeister function

$$\phi(a, b) := \sqrt{a^2 + b^2} - (a + b).$$

The minimum-function is used in the developments of local Newton methods discussed in [26]. However, when it comes to the development of globally convergent algorithms, the Fischer-Burmeister function has the distinctive advantage of giving rise to continuously differentiable merit functions. There exist also variants of the latter complementarity function, like the penalized Fischer-Burmeister function, defined for an arbitrary parameter $\gamma \in (0, 1)$ by

$$\phi(a, b) := \gamma \left(\sqrt{a^2 + b^2} - (a + b) \right) - (1 - \gamma) \max\{0, a\} \max\{0, b\},$$

which was introduced in [11]. By [11, Proposition 1] the penalized Fischer-Burmeister function is a strongly semismooth complementarity function, which is continuously differentiable on $\mathbb{R}^2 \setminus \{(a, b) \mid a \geq 0, b \geq 0, ab = 0\}$. We will use both the standard and the penalized Fischer-Burmeister function. If a result holds for both functions we simply write ϕ . Otherwise we will explicitly mention which function we use.

Once the concatenated KKT system has been reformulated as a system of equations, we can solve the resulting system by finding a (global) minimum of the natural merit function

$$\Theta(x, \lambda) := \frac{1}{2} \left\| \begin{pmatrix} F(x, \lambda) \\ \Phi(x, \lambda) \end{pmatrix} \right\|^2.$$

Note that Φ (using the (penalized) Fischer-Burmeister function) is not differentiable in general, because the (penalized) Fischer-Burmeister complementarity function is nondifferentiable. However, it is very well known that Θ is once (though not twice) continuously differentiable for the Fischer-Burmeister function and also for the penalized one, c.f. [11]. Hence we can use standard optimization software to attempt to (globally) minimize Θ and find in this way a solution of the GNEP.

This is a well-established path and it is well understood that the two key issues that need to be addressed are

- conditions guaranteeing that unconstrained stationary points of Θ are global solutions, and
- conditions under which Θ can be shown to be coercive.

Once this has been done, one can safely attempt to solve the KKT system (5.2) by performing the unconstrained minimization of Θ . Unfortunately, while in the optimization and VI fields “reasonable” conditions guaranteeing the above mentioned results can be identified, see [31], the situation becomes much more involved in the case of system (5.2).

5.1.1. Stationarity Conditions

For the sake of notational simplicity, it is useful to introduce the matrix

$$E(x) := \begin{pmatrix} \nabla_{x^1} g^1(x) & & 0 \\ & \ddots & \\ 0 & & \nabla_{x^N} g^N(x) \end{pmatrix} \quad \text{with} \quad \nabla_{x^\nu} g^\nu(x) \in \mathbb{R}^{n_\nu \times m_\nu}. \quad (5.3)$$

Using the chain rule from [12] and some standard calculations, we obtain that the gradient of Θ is given by

$$\nabla \Theta(x, \lambda) = \begin{pmatrix} J_x F(x, \lambda) & E(x) \\ -D_g(x, \lambda) J_x g(x) & D_\lambda(x, \lambda) \end{pmatrix}^T \begin{pmatrix} F(x, \lambda) \\ \Phi(x, \lambda) \end{pmatrix},$$

where the matrices D_λ and D_g are $m \times m$ diagonal matrices

$$\begin{aligned} D_\lambda(x, \lambda) &:= \text{diag} (a^1(x, \lambda^1), \dots, a^N(x, \lambda^N)), \\ D_g(x, \lambda) &:= \text{diag} (b^1(x, \lambda^1), \dots, b^N(x, \lambda^N)), \end{aligned}$$

with vectors $a^\nu(x, \lambda^\nu), b^\nu(x, \lambda^\nu) \in \mathbb{R}^{m_\nu}$ whose entries are given by elements of the generalized Jacobian of the complementarity function, i.e., for all $i = 1, \dots, m_\nu$ and for all $\nu = 1, \dots, N$ we have

$$(a_i^\nu(x, \lambda_i^\nu), b_i^\nu(x, \lambda_i^\nu)) \begin{cases} = \frac{(\lambda_i^\nu, -g_i^\nu(x))}{\sqrt{(\lambda_i^\nu)^2 + g_i^\nu(x)^2}} - (1, 1), & \text{if } (\lambda_i^\nu, -g_i^\nu(x)) \neq (0, 0), \\ \in \text{cl}(\mathbb{B}_1(0, 0)) - (1, 1), & \text{if } (\lambda_i^\nu, -g_i^\nu(x)) = (0, 0) \end{cases}$$

for the Fischer-Burmeister function, and

$$(a_i^\nu(x, \lambda_i^\nu), b_i^\nu(x, \lambda_i^\nu)) \in \begin{cases} \left(\gamma \left(\frac{\lambda_i^\nu}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right) - (1 - \gamma) \max\{0, -g_i^\nu(x)\} \partial(\lambda_i^\nu)_+, \right. \\ \left. \gamma \left(\frac{-g_i^\nu(x)}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right) - (1 - \gamma) \max\{0, \lambda_i^\nu\} \partial(-g_i^\nu(x))_+ \right), & \text{if } (\lambda_i^\nu, -g_i^\nu(x)) \neq (0, 0), \\ \text{cl}(\mathbb{B}_\gamma(0, 0)) - (\gamma, \gamma), & \text{if } (\lambda_i^\nu, -g_i^\nu(x)) = (0, 0), \end{cases}$$

with

$$\partial(c)_+ := \begin{cases} 1, & \text{if } c > 0, \\ [0, 1], & \text{if } c = 0, \\ 0, & \text{if } c < 0, \end{cases}$$

for the penalized Fischer-Burmeister function with parameter $\gamma \in (0, 1)$, cf. [11]. For both complementarity functions $a_i^\nu(x, \lambda_i^\nu)$ and $b_i^\nu(x, \lambda_i^\nu)$ are nonpositive for all $i = 1, \dots, m_\nu$ and for all $\nu = 1, \dots, N$ and $a_i^\nu(x, \lambda_i^\nu) = 0$ or $b_i^\nu(x, \lambda_i^\nu) = 0$ implies $\phi(\lambda_i^\nu, -g_i^\nu(x)) = 0$. Therefore, in spite of the fact that the matrix appearing in the expression of $\nabla\Theta$ is not uniquely defined, the gradient of Θ itself is uniquely determined because the possibly multivalued elements of the generalized Jacobian are cancelled by corresponding zero entries in $\Phi(x, \lambda)$.

Based on this expression it is possible to establish a result, giving a sufficient condition for a stationary point of Θ to be a solution of the GNEP.

Theorem 5.1 *Let $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ be a stationary point of Θ , and suppose that $J_x F(\bar{x}, \bar{\lambda})$ is nonsingular and*

$$M(\bar{x}, \bar{\lambda}) := J_x g(\bar{x}) J_x F(\bar{x}, \bar{\lambda})^{-1} E(\bar{x})$$

is a P_0 -matrix. Then \bar{x} is a solution of the GNEP.

Proof. Since $(\bar{x}, \bar{\lambda})$ is a stationary point of Θ , it holds that

$$\nabla_x F(\bar{x}, \bar{\lambda}) F(\bar{x}, \bar{\lambda}) - \nabla_x g(\bar{x}) D_g(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}) = 0, \quad (5.4)$$

$$E(\bar{x})^T F(\bar{x}, \bar{\lambda}) + D_\lambda(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}) = 0. \quad (5.5)$$

By the nonsingularity of $\nabla_x F(\bar{x}, \bar{\lambda})$, we obtain from (5.4)

$$F(\bar{x}, \bar{\lambda}) = \nabla_x F(\bar{x}, \bar{\lambda})^{-1} \nabla_x g(\bar{x}) D_g(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}), \quad (5.6)$$

and substituting this into (5.5), we get

$$\begin{aligned} 0 &= E(\bar{x})^T \nabla_x F(\bar{x}, \bar{\lambda})^{-1} \nabla_x g(\bar{x}) D_g(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}) + D_\lambda(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}) \\ &= [M(\bar{x}, \bar{\lambda})^T D_g(\bar{x}, \bar{\lambda}) + D_\lambda(\bar{x}, \bar{\lambda})] \Phi(\bar{x}, \bar{\lambda}). \end{aligned} \quad (5.7)$$

Now recall that $a_i^\nu(\bar{x}, \bar{\lambda}_i^\nu), b_i^\nu(\bar{x}, \bar{\lambda}_i^\nu)$ are nonpositive with $(a_i^\nu(\bar{x}, \bar{\lambda}_i^\nu), b_i^\nu(\bar{x}, \bar{\lambda}_i^\nu)) \neq (0, 0)$ for all i, ν , and that $a_i^\nu(\bar{x}, \bar{\lambda}_i^\nu) = 0$ or $b_i^\nu(\bar{x}, \bar{\lambda}_i^\nu) = 0$ can happen only if $\phi(\bar{\lambda}_i^\nu, -g_i^\nu(\bar{x})) = 0$. Since in the previous equations both elements $a_i^\nu(\bar{x}, \bar{\lambda}_i^\nu)$ and $b_i^\nu(\bar{x}, \bar{\lambda}_i^\nu)$ are always post-multiplied by

$\phi(\bar{\lambda}_i^\nu, -g_i^\nu(\bar{x})) = 0$, we do not change these equations if we assume without loss of generality that both diagonal matrices $D_\lambda(\bar{x}, \bar{\lambda})$ and $D_g(\bar{x}, \bar{\lambda})$ are negative definite. Since $M(\bar{x}, \bar{\lambda})$ is assumed to be a P_0 -matrix, it follows that $(M(\bar{x}, \bar{\lambda})^T D_g(\bar{x}, \bar{\lambda}) + D_\lambda(\bar{x}, \bar{\lambda}))$ is nonsingular. Hence $\Phi(\bar{x}, \bar{\lambda}) = 0$ by (5.7), and this immediately implies $F(\bar{x}, \bar{\lambda}) = 0$ by (5.6). Therefore $(\bar{x}, \bar{\lambda})$ is a solution. \square

This result is particularly simple to verify when the constraints of the problem are all linear. In fact, in this case the matrix $M(x, \lambda)$ does not actually depend on the values of the multipliers. The situation becomes still simpler for games with quadratic objective functions and linear constraints, because in this case the matrix $M(x, \lambda)$ is actually independent of (x, λ) and the condition in the theorem reduces to the verification of the nonsingularity and P_0 property of two matrices.

Example 5.2 Consider a GNEP with three players $\nu = 1, 2, 3$, where player ν controls the single variable $x^\nu \in \mathbb{R}$, and the problem is given by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1 - 1)^2 - x^1 x^2 \quad \text{s.t. } x^1 + x^2 + x^3 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2 - 1)^2 + x^1 x^2 \quad \text{s.t. } x^1 + x^2 + x^3 \leq 1, \\ \text{Player 3: } & \min_{x^3} \frac{1}{2}(x^3 - 1)^2 \quad \text{s.t. } 0 \leq x^3 \leq x^1 + x^2. \end{aligned}$$

Then we have a nonsingular matrix

$$J_x F(x, \lambda) = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and we get

$$M(x, \lambda) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & -1 \\ -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & -1 & -1 & 1 \end{bmatrix}.$$

An elementary calculation shows that $\det(M(x, \lambda)_{\alpha\alpha}) \geq 0$ holds for all $\alpha \subseteq \{1, 2, 3, 4\}$, hence $M(x, \lambda)$ is a P_0 -matrix. Consequently, Theorem 5.1 can be applied and guarantees that every stationary point of Θ is a solution of the GNEP. \diamond

This example also indicates a limitation of Theorem 5.1, if the constraints are not linear. In this case, the nonsingularity of $J_x F(x, \lambda)$ and the P_0 property of $M(x, \lambda)$ must hold even for negative values of λ , and it is apparent that this won't be the case in general. In fact, $J_x F(x, \lambda)$ will contain block-diagonal terms of the type $\lambda_i^\nu \nabla_{x^\nu x^\nu}^2 g_i^\nu(x)$, which will be negative definite if λ_i^ν is negative, and can lead to a singular matrix $J_x F(x, \lambda)$ as in the following example.

Example 5.3 Consider a 2-player game where each player controls a single variable, given by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1)^2 + \frac{32}{5}x^1 & \text{s.t. } & \frac{1}{6}(x^1)^2 + x^2 - \frac{5}{2} \leq 0, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2)^2 + x^1x^2 - \frac{4}{5}x^2 & \text{s.t. } & x^2 \in \mathbb{R}. \end{aligned}$$

Then we have

$$J_x F(x, \lambda) = \begin{bmatrix} 1 + \frac{1}{3}\lambda & 0 \\ 1 & 1 \end{bmatrix},$$

which is nonsingular for all $\lambda \neq -3$. But if we consider the point $\bar{x} = (3, -3)$ together with $\bar{\lambda} = -3$, we obtain (using the Fischer-Burmeister function)

$$\begin{aligned} \nabla \Theta(\bar{x}, \bar{\lambda}) &= \begin{bmatrix} 1 + \frac{1}{3}\bar{\lambda} & 1 & -\frac{1}{3}\bar{x}^1 b(\bar{x}, \bar{\lambda}) \\ 0 & 1 & -b(\bar{x}, \bar{\lambda}) \\ \frac{1}{3}\bar{x}^1 & 0 & a(\bar{x}, \bar{\lambda}) \end{bmatrix} \begin{pmatrix} \bar{x}^1 + \frac{32}{5} + \frac{1}{3}\bar{x}^1 \bar{\lambda} \\ \bar{x}^2 + \bar{x}^1 - \frac{4}{5} \\ \phi(\bar{\lambda}, -\frac{1}{6}(\bar{x}^1)^2 - \bar{x}^2 + \frac{5}{2}) \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & \frac{1}{5} \\ 0 & 1 & \frac{1}{5} \\ 1 & 0 & -\frac{8}{5} \end{pmatrix} \begin{pmatrix} \frac{32}{5} \\ -\frac{4}{5} \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

Hence we have a stationary point that is certainly not a solution of the GNEP, since $\Theta(\bar{x}, \bar{\lambda}) = \frac{1}{2} \|(\frac{32}{5}, -\frac{4}{5}, 4)\|^2 \neq 0$. \diamond

This example might suggest that negativity of the multipliers is the reason for the failure of a stationary point being a solution. Therefore one could wish to solve the problem by considering a *constrained* minimization of Θ , that is by solving the problem

$$\min \Theta(x, \lambda) \quad \text{s.t.} \quad \lambda \geq 0. \quad (5.8)$$

This leads to successful results in the optimization/VI case, see [24, 31]. Unfortunately, the approach leads to problems in our game setting, as illustrated by the following example.

Example 5.4 Consider an apparently well-behaved game where each player controls a single variable, and the players' problems are given by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} x^1 & \text{s.t. } & (x^1)^2 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2)^2 & \text{s.t. } & x^2 \in \mathbb{R}. \end{aligned}$$

For player 2 the global minimum $x^2 = 0$ of the cost function is always feasible, and thus it is attained at the solution. For player 1 this implies that $x^1 = -1$ is the constrained optimum, hence the point $(-1, 0)$ is the only generalized Nash equilibrium. The corresponding Lagrange multiplier is $\lambda = \frac{1}{2}$. Since

$$\nabla \Theta(x, \lambda) = \begin{pmatrix} 2\lambda & 0 & -2x^1 b(x, \lambda) \\ 0 & 1 & -b(x, \lambda) \\ 2x^1 & 0 & a(x, \lambda) \end{pmatrix} \begin{pmatrix} 1 + 2x^1 \lambda \\ x^2 \\ \phi(\lambda, 1 - (x^1)^2 - x^2) \end{pmatrix},$$

the point $\bar{x} = (0, 0)$ together with $\bar{\lambda} = 0$ is an unconstrained stationary point of Θ , and also a stationary point of the constrained problem (5.8), but \bar{x} is not a solution. \diamond

In case the feasible sets of the players do not depend on the rivals' strategies, so that we have a standard Nash equilibrium problem (NEP), we can obtain a result that looks more familiar.

Theorem 5.5 *Consider a NEP. Let $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m$ be a stationary point of Θ , and suppose that $J_x F(\bar{x}, \bar{\lambda})$ is positive semidefinite and it holds that*

$$d^T J_x F(\bar{x}, \bar{\lambda}) d > 0 \quad \text{for all } d \in \{d \in \mathbb{R}^n \setminus \{0\} \mid E(\bar{x})^T d = 0\}.$$

Then \bar{x} is a solution of the NEP.

Proof. In a NEP we have $\nabla_x g(x) = E(x)$. Taking the two stationarity conditions (5.4) and (5.5), multiplying the first with $F(\bar{x}, \bar{\lambda})^T$ and substituting the second one in the resulting expression, we get

$$F(\bar{x}, \bar{\lambda})^T \nabla_x F(\bar{x}, \bar{\lambda}) F(\bar{x}, \bar{\lambda}) + \Phi(\bar{x}, \bar{\lambda})^T D_\lambda(\bar{x}, \bar{\lambda}) D_g(\bar{x}, \bar{\lambda}) \Phi(\bar{x}, \bar{\lambda}) = 0.$$

By the positive semidefiniteness of $J_x F(\bar{x}, \bar{\lambda})$ and since we may assume, without loss of generality, that both diagonal matrices $D_\lambda(\bar{x}, \bar{\lambda})$ and $D_g(\bar{x}, \bar{\lambda})$ have negative entries (cf. the proof of Theorem 5.1), we get $\Phi(\bar{x}, \bar{\lambda}) = 0$. Then equations (5.4) and (5.5), together with $d^T J_x F(\bar{x}, \bar{\lambda}) d > 0$ for all $d \in \{d \in \mathbb{R}^n \setminus \{0\} \mid E(\bar{x})^T d = 0\}$, imply $F(\bar{x}, \bar{\lambda}) = 0$, which completes the proof. \square

At first glance, the previous result looks very standard. We stress, however, that this is not so since the tangent cone in the assumptions of the theorem $\{d \mid E(\bar{x})^T d = 0\}$ is (in general) much smaller than the usual tangent cone. To this end, note that this tangent cone may be rewritten as

$$\mathcal{T}(x) = \{d = (d^1, \dots, d^N) \mid \nabla g_i^\nu(x^\nu)^T d^\nu = 0, \forall i = 1, \dots, m_\nu \forall \nu = 1, \dots, N\},$$

meaning that this set contains all vectors d whose block components d^ν are orthogonal to the gradients of *all* constraints $g_i^\nu(x^\nu) \leq 0$ and not just to the active ones. Hence the requirement in Theorem 5.5 is significantly weaker than the usual one.

5.1.2. Coercivity

The previous results provide conditions under which a stationary point of Θ is a solution of the underlying GNEP. Now, suppose we use a suitable descent method for the minimization of Θ . Any reasonable method has the property that each of its accumulation points is a stationary point of Θ and, therefore, a global minimum under the conditions given in our previous results. Hence, the main question remaining to be answered, at least from a

theoretical point of view, is under which assumptions a sequence $\{(x^k, \lambda^k)\}$, generated by a descent method, is guaranteed to be bounded, so that an accumulation point exists. A sufficient condition would be the boundedness of the level sets of Θ . Unfortunately, these level sets are typically unbounded, even under very restrictive assumptions. However, a closer look at the merit function Θ shows that this has mainly to do with the behavior of the sequence $\{\lambda^k\}$ which, in particular, might be unbounded.

From now on the analysis depends on the specific choice of the complementarity function. Let us first choose the Fischer-Burmeister function. Then it is possible to show that the sequence $\{x^k\}$ remains bounded under very reasonable assumptions. To this end, consider a GNEP that is defined via the optimization problems

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t.} \quad g^\nu(x^\nu, x^{-\nu}) \leq 0, \quad h^\nu(x^\nu) \leq 0, \quad \nu = 1, \dots, N,$$

with functions $h_j^\nu : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}$ and $g_i^\nu : \mathbb{R}^n \rightarrow \mathbb{R}$ for $j = 1, \dots, p_\nu, i = p_\nu + 1, \dots, m_\nu$, that are assumed to be convex in x^ν . Here we distinguish, for each player $\nu = 1, \dots, N$, between those constraints h^ν that depend on his own variables x^ν only, and those constraints g^ν that are allowed to depend on all variables. Consider the set

$$X_0 := \{x \in \mathbb{R}^n \mid h^\nu(x^\nu) \leq 0 \quad \forall \nu = 1, \dots, N\}.$$

This set is closed and convex since the constraints h^ν are convex by assumption. Assuming boundedness of the set X_0 , one can show boundedness of the x part of the iterates.

Proposition 5.6 *Suppose $h_j^\nu : \mathbb{R}^{n_\nu} \rightarrow \mathbb{R}$ is convex for all $\nu = 1, \dots, N, j = 1, \dots, p_\nu$ and the set X_0 is nonempty and bounded. Furthermore, let $\{(x^k, \lambda^k)\}$ be any sequence such that $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$ for all $k \in \mathbb{N}$. Then the sequence $\{x^k\}$ is bounded.*

Proof. Let us define

$$h_{\max}(x) := \max \{h_1^1(x^1), \dots, h_{p_1}^1(x^1), h_1^2(x^2), \dots, h_{p_N}^N(x^N)\}.$$

Being the maximum of convex functions, it follows that h_{\max} itself is also convex. Moreover

$$h_j^\nu(x^\nu) \leq \gamma \quad \forall j = 1, \dots, p_\nu \quad \forall \nu = 1, \dots, N \iff h_{\max}(x) \leq \gamma$$

for any given $\gamma \in \mathbb{R}$. In particular, we can rewrite the set X_0 as

$$X_0 = \{x \in \mathbb{R}^n \mid h_{\max}(x) \leq 0\}.$$

Since h_{\max} is a single convex function, it follows from our boundedness assumption on X_0 together with [66, Corollary 8.7.1] that the level sets

$$X_\gamma := \{x \in \mathbb{R}^n \mid h_{\max}(x) \leq \gamma\} = \{x \in \mathbb{R}^n \mid h_j^\nu(x^\nu) \leq \gamma \quad \forall j = 1, \dots, p_\nu \quad \forall \nu = 1, \dots, N\}$$

are also bounded for any $\gamma \in \mathbb{R}$.

Now, assume that the sequence $\{x^k\}$ is unbounded, say $\{\|x^k\|\} \rightarrow \infty$. Since X_γ is bounded for each $\gamma \in \mathbb{R}$, we can therefore find, for any given $\gamma = k, k \in \mathbb{N}$, an index $\ell(k) \in \mathbb{N}$ such that $x^{\ell(k)} \notin X_k$. This means that, for every $k \in \mathbb{N}$, there are indices $\nu(k) \in \{1, \dots, N\}$ and $j(k) \in \{1, \dots, p_{\nu(k)}\}$ such that $h_{j(k)}^{\nu(k)}(x^{\ell(k)}) > k$. Since there is only a finite number of players and constraints, there exist fixed indices $\nu \in \{1, \dots, N\}$ and $j \in \{1, \dots, p_\nu\}$, independent of $k \in \mathbb{N}$, such that $h_j^\nu(x^{\ell(k)}) > k$ on a suitable subsequence, say, for all $k \in K$. Exploiting this fact, it follows from the definition of the Fischer-Burmeister function that

$$\begin{aligned} \phi((\lambda^{\ell(k)})_j^\nu, -h_j^\nu(x^{\ell(k)})) &= \sqrt{(h_j^\nu(x^{\ell(k)}))^2 + ((\lambda^{\ell(k)})_j^\nu)^2} - (\lambda^{\ell(k)})_j^\nu + h_j^\nu(x^{\ell(k)}) \\ &\geq h_j^\nu(x^{\ell(k)}) > k, \end{aligned}$$

and thus we obtain

$$\Theta(x^{\ell(k)}, \lambda^{\ell(k)}) \geq \frac{1}{2} \phi^2((\lambda^{\ell(k)})_j^\nu, -h_j^\nu(x^{\ell(k)})) > \frac{1}{2} k^2.$$

Hence we have $\Theta(x^{\ell(k)}, \lambda^{\ell(k)}) \rightarrow_K \infty$ for $k \rightarrow_K \infty$, contradicting the assumption that $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$ for all $k \in \mathbb{N}$. \square

Using the penalized Fischer-Burmeister function with parameter $\gamma \in (0, 1)$ (and different assumptions) it is possible to show boundedness not only for the sequence $\{x^k\}$ but also for the sequence $\{\lambda^k\}$. Before stating this result, let us first recall some properties of the penalized Fischer-Burmeister function. By [11, Proposition 1] the penalized Fischer-Burmeister function is a strongly semismooth complementarity function, which is continuously differentiable on $\mathbb{R}^2 \setminus \{(a, b) \mid a \geq 0, b \geq 0, ab = 0\}$. Further we have $|\phi(a^k, b^k)| \rightarrow \infty$ for any sequences $\{a^k\}, \{b^k\} \subset \mathbb{R}$ with $(\max\{0, a^k\} \max\{0, b^k\}) \rightarrow +\infty$ or $a^k \rightarrow -\infty$ or $b^k \rightarrow -\infty$.

Theorem 5.7 *Assume that*

- (a) $\lim_{\|x\| \rightarrow \infty} \|g_+(x)\| = +\infty$, where $g_+(x) := \max\{0, g(x)\}$;
- (b) the *Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ)* holds for each player, i.e., for all $\nu = 1, \dots, N$ and for all $x \in \mathbb{R}^n$ there exists a $d^\nu \in \mathbb{R}^{n_\nu}$ such that

$$\nabla_{x^\nu} g_i^\nu(x)^T d^\nu < 0 \quad \text{for all } i \in I_{\geq}^\nu(x), \quad (5.9)$$

where $I_{\geq}^\nu(x) := \{i \in \{1, \dots, m_\nu\} \mid g_i^\nu(x) \geq 0\}$ denotes the set of active or violated constraints for player ν .

Furthermore, let $\{(x^k, \lambda^k)\}$ be any sequence such that $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$ for all $k \in \mathbb{N}$. Then the sequence $\{(x^k, \lambda^k)\}$ is bounded.

Proof. First assume that the sequence $\{x^k\}$ is unbounded, i.e., for some $K \subseteq \mathbb{N}$ we have $\|x^k\| \rightarrow_K \infty$. Then by (a) there exist a $\nu \in \{1, \dots, N\}$ and an $i \in \{1, \dots, m_\nu\}$ such that

$\{g_i^\nu(x^k)\} \rightarrow_K +\infty$. The properties of the penalized Fischer-Burmeister function imply $|\phi((\lambda^k)_i^\nu, -g_i^\nu(x^k))| \rightarrow_K +\infty$, which, in turn, implies $\Theta(x^k, \lambda^k) \rightarrow_K +\infty$. This contradicts $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$ and thus shows that $\{x^k\}$ is bounded. Therefore, without loss of generality, the sequence $\{x^k\}$ converges to some \bar{x} .

Now assume that $\{\lambda^k\}$ is unbounded. Then there exist indices $\nu \in \{1, \dots, N\}$ and $i \in \{1, \dots, m_\nu\}$ with $(\lambda^k)_i^\nu \rightarrow_K +\infty$ or $(\lambda^k)_i^\nu \rightarrow_K -\infty$ on a subset $K \subseteq \mathbb{N}$. The latter one again implies $|\phi((\lambda^k)_i^\nu, -g_i^\nu(x^k))| \rightarrow_K +\infty$ and hence contradicts $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$. Thus we have $(\lambda^k)_i^\nu \rightarrow_K +\infty$. For $g_i^\nu(\bar{x}) < 0$, we obtain $(\max\{0, (\lambda^k)_i^\nu\} \max\{0, -g_i^\nu(x^k)\}) \rightarrow_K +\infty$ and the properties of ϕ imply $|\phi((\lambda^k)_i^\nu, -g_i^\nu(x^k))| \rightarrow_K +\infty$, which contradicts our assumptions. Hence we have $g_i^\nu(\bar{x}) \geq 0$.

Let J be the set of all $i \in \{1, \dots, m_\nu\}$ with $(\lambda^k)_i^\nu \rightarrow_K +\infty$ on an appropriate subsequence K . By the above we know $g_i^\nu(\bar{x}) \geq 0$ for all $i \in J$. Using EMFCQ from (5.9), there exists a vector d^ν such that $\nabla_{x^\nu} g_i^\nu(\bar{x})^T d^\nu < 0$ for all $i \in J$. This yields

$$\begin{aligned} \lim_{k \in K} \nabla_{x^\nu} L^\nu(x^k, (\lambda^k)^\nu)^T d^\nu &= \lim_{k \in K} \left(\nabla_{x^\nu} \theta_\nu(x^k) + \sum_{i \notin J} (\lambda^k)_i^\nu \nabla_{x^\nu} g_i^\nu(x^k) \right)^T d^\nu \\ &\quad + \lim_{k \in K} \left(\sum_{i \in J} (\lambda^k)_i^\nu \nabla_{x^\nu} g_i^\nu(x^k) \right)^T d^\nu = -\infty, \end{aligned}$$

since the first term is bounded (because of $\{x^k\} \rightarrow_K \bar{x}$, the continuity of the functions $\nabla_{x^\nu} \theta_\nu$ and $\nabla_{x^\nu} g^\nu$ and because all sequences $(\lambda^k)_i^\nu$ for $i \notin J$ are bounded by the definition of the index set J), whereas the second term is unbounded since $(\lambda^k)_i^\nu \rightarrow_K +\infty$ and $\nabla_{x^\nu} g_i^\nu(\bar{x})^T d^\nu < 0$ for all $i \in J$. Using the Cauchy-Schwarz inequality, we therefore obtain

$$\|\nabla_{x^\nu} L^\nu(x^k, (\lambda^k)^\nu)\| \|d^\nu\| \geq |\nabla_{x^\nu} L^\nu(x^k, (\lambda^k)^\nu)^T d^\nu| \rightarrow_K +\infty.$$

Since d^ν is a fixed vector, this implies $\|\nabla_{x^\nu} L^\nu(x^k, (\lambda^k)^\nu)\| \rightarrow_K +\infty$ which, in turn, contradicts $\Theta(x^k, \lambda^k) \leq \Theta(x^0, \lambda^0)$. Thus also the sequence $\{\lambda^k\}$ is bounded. \square

Note that condition (a) in the theorem above is a mild boundedness assumption on the feasible sets of the players. In particular, (a) holds in the setting of Proposition 5.6. Also condition (b) is rather mild and common in an optimization context.

Inspecting the proof of Theorem 5.7, one can see that condition (a) is also sufficient to show boundedness of the sequence $\{x^k\}$ for the Fischer-Burmeister function. However, the boundedness of the λ -sequence can only be obtained by the penalized Fischer-Burmeister function, because the penalty term is needed to ensure the crucial part $g_i^\nu(\bar{x}) \geq 0$ in the proof.

The given results show that the penalized Fischer-Burmeister function has better theoretical properties than the usual one.

5.2. Interior Point Method

As we have seen in Section 5.1.1, there are some problems with the “ λ -part” of the variables. We already discussed, see (5.8), that a straightforward treatment of the sign constraints for the multipliers is not likely to be helpful in the merit function approach. Another suitable alternative seems to be an interior point approach to the solution of the GNEP KKT system (5.2), since these methods are well known to be efficient methods for solving KKT systems arising from optimization or VI problems. We will develop an inexact potential reduction algorithm out of an exact version and analyse its properties. To this end, we formulate the GNEP KKT system as a constrained nonlinear system of equations (CE or, more precisely, $\text{CE}(H, Z)$) of the form

$$H(z) = 0, \quad z \in Z \quad (5.10)$$

for a given function $H : \mathbb{R}^l \rightarrow \mathbb{R}^l$ and a given set $Z \subseteq \mathbb{R}^l$ that we define below.

We introduce slack variables $w := (w^\nu)_{\nu=1}^N$, where $w^\nu \in \mathbb{R}^{m_\nu}$, and set

$$\lambda \circ w := (\lambda_1^1 w_1^1, \dots, \lambda_{m_N}^N w_{m_N}^N)^T.$$

Then we define

$$H(z) := H(x, \lambda, w) := \begin{pmatrix} F(x, \lambda) \\ g(x) + w \\ \lambda \circ w \end{pmatrix} \quad (5.11)$$

and

$$Z := \{z = (x, \lambda, w) \mid x \in \mathbb{R}^n, \lambda \in \mathbb{R}_+^m, w \in \mathbb{R}_+^m\}. \quad (5.12)$$

It is immediate to verify that a point (x, λ) solves the KKT system (5.2) if and only if this point, together with a suitable w , solves the constrained equation defined by (5.11) and (5.12). In order to solve this constrained equation problem, we use an interior point approach that generates points in the interior of Z . In other words, our method will generate a sequence (x^k, λ^k, w^k) with $\lambda^k > 0$ and $w^k > 0$ for every k . The particular method that we base our analysis on is the potential reduction method from [54], also discussed in detail in [31]. We generalize this potential reduction method by allowing inexact solutions of the subproblems and study in detail its implication in the case of our specific system (5.11) and (5.12). Note that we want to keep the notation from [31]. Thus we define the set

$$S := \mathbb{R}^n \times \mathbb{R}_+^{2m},$$

which coincides with the set Z in our case, as well as a potential function on $\text{int}(S)$

$$p(u, v) := \zeta \log(\|u\|^2 + \|v\|^2) - \sum_{i=1}^{2m} \log(v_i), \quad (u, v) \in \mathbb{R}^n \times \mathbb{R}_{++}^{2m}, \quad \zeta > m.$$

The properties of this function are well known from the literature on interior point methods. Basically, the function p is defined in the interior of S and penalizes points that are near the boundary of S , but are far from the origin. Some further properties we will use are stated in the following lemma.

Lemma 5.8 Define $a := (0_n^T, 1_{2m}^T)^T$. Then it holds that

$$\nabla p(u, v)^T \begin{pmatrix} u \\ v \end{pmatrix} = 2(\zeta - m) > 0, \quad (5.13)$$

$$\begin{pmatrix} a^T \begin{pmatrix} u \\ v \end{pmatrix} \end{pmatrix} (a^T \nabla p(u, v)) \leq \|a\|^2 \nabla p(u, v)^T \begin{pmatrix} u \\ v \end{pmatrix} \quad (5.14)$$

for all $(u, v) \in \mathbb{R}^n \times \mathbb{R}_{++}^{2m}$.

Proof. Equation (5.13) follows with $v^{-1} := (v_1^{-1}, \dots, v_{2m}^{-1})$ from the definition of p by

$$\begin{aligned} \nabla p(u, v)^T \begin{pmatrix} u \\ v \end{pmatrix} &= \begin{pmatrix} \zeta \frac{2u}{\|u\|^2 + \|v\|^2} \\ \zeta \frac{2v}{\|u\|^2 + \|v\|^2} - v^{-1} \end{pmatrix}^T \begin{pmatrix} u \\ v \end{pmatrix} \\ &= \zeta \frac{2\|u\|^2}{\|u\|^2 + \|v\|^2} + \zeta \frac{2\|v\|^2}{\|u\|^2 + \|v\|^2} - \underbrace{v^T v^{-1}}_{=2m} \\ &= 2\zeta - 2m. \end{aligned}$$

Further, taking into account the definition of a and $\|v\|_1 = \sum_{i=1}^{2m} v_i$, it follows that

$$\begin{aligned} \begin{pmatrix} a^T \begin{pmatrix} u \\ v \end{pmatrix} \end{pmatrix} (a^T \nabla p(u, v)) &= (1_{2m}^T v) (1_{2m}^T \nabla_v p(u, v)) \\ &= \left(\sum_{i=1}^{2m} v_i \right) \left(\sum_{i=1}^{2m} \left[\zeta \frac{2v_i}{\|u\|^2 + \|v\|^2} - \frac{1}{v_i} \right] \right) \\ &= \frac{2\zeta \|v\|_1^2}{\|u\|^2 + \|v\|^2} - \left(\sum_{i=1}^{2m} v_i \right) \left(\sum_{i=1}^{2m} \frac{1}{v_i} \right). \end{aligned}$$

It is known that $\|v\|_1 \leq \sqrt{2m} \|v\|$ and, by the arithmetic-geometric mean inequality,

$$\begin{aligned} \frac{1}{2m} \sum_{i=1}^{2m} v_i &\geq \sqrt[2m]{v_1 \cdots v_{2m}}, \\ \frac{1}{2m} \sum_{i=1}^{2m} \frac{1}{v_i} &\geq \sqrt[2m]{\frac{1}{v_1} \cdots \frac{1}{v_{2m}}}. \end{aligned}$$

This can be used in order to estimate the previous expression by

$$\begin{aligned} \begin{pmatrix} a^T \begin{pmatrix} u \\ v \end{pmatrix} \end{pmatrix} (a^T \nabla p(u, v)) &\leq \frac{2\zeta 2m \|v\|^2}{\|u\|^2 + \|v\|^2} - 4m^2 \underbrace{\sqrt[2m]{v_1 \cdots v_{2m}} \sqrt[2m]{\frac{1}{v_1} \cdots \frac{1}{v_{2m}}}}_{=1} \\ &\leq 2\zeta 2m - 4m^2 \end{aligned}$$

$$\begin{aligned}
 &= 2m2(\zeta - m) \\
 &= \|a\|^2 \nabla p(u, v)^T \begin{pmatrix} u \\ v \end{pmatrix},
 \end{aligned}$$

where the last equality follows from (5.13) and $\|a\|^2 = \|1_{2m}\|^2 = 2m$. \square

Based on p , we obtain a potential function for the CE(H, Z) which is defined on the nonempty set

$$Z_I := H^{-1}(\text{int}(S)) \cap \text{int}(Z)$$

by setting

$$\psi(z) := p(H(z)) \quad \text{for } z \in Z_I.$$

Throughout this section, p and ψ always denote these two potential functions.

We are now in the position to formulate our interior point method. The core of this approach is the calculation of a Newton-type direction for the system $H(z) = 0$. According to standard procedures in interior point methods, the Newton direction is “bent” in order to follow the central path. Operatively this means that the search direction used in this method is the solution of the system

$$H(z^k) + JH(z^k)d^k = \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \quad (5.15)$$

with a defined by $a^T := (0_n^T, 1_{2m}^T)$. Once this direction has been calculated, a line-search is performed by using the potential function ψ . The version we describe and analyse below is a variant where we allow the possibility of an inaccurate solution of system (5.15).

Algorithm 5.9 (*Inexact Potential Reduction Method for GNEPs*)

(S.0) Choose $z^0 \in Z_I, \beta, \gamma \in (0, 1)$, and set $k := 0, \bar{\sigma} = 1, a^T = (0_n^T, 1_{2m}^T)$.

(S.1) If $H(z^k) = 0$: STOP.

(S.2) Choose $\sigma_k \in [0, \bar{\sigma}), \eta_k \geq 0$, and compute a vector $d^k \in \mathbb{R}^l$ such that

$$\left\| H(z^k) + JH(z^k)d^k - \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \right\| \leq \eta_k \|H(z^k)\| \quad \text{and} \quad (5.16)$$

$$\nabla \psi(z^k)^T d^k < 0. \quad (5.17)$$

(S.3) Compute a stepsize $t_k := \max \{ \beta^\ell \mid \ell = 0, 1, 2, \dots \}$ such that

$$z^k + t_k d^k \in Z_I \quad \text{and} \quad (5.18)$$

$$\psi(z^k + t_k d^k) \leq \psi(z^k) + \gamma t_k \nabla \psi(z^k)^T d^k. \quad (5.19)$$

(S.4) Set $z^{k+1} := z^k + t_k d^k, k \leftarrow k + 1$, and go to (S.1).

Remark 5.10 (a) By construction, all iterates z^k generated by Algorithm 5.9 belong to the set Z_I , hence we have $z^k \in \text{int}(Z)$ and $H(z^k) \in \text{int}(S)$ for all $k \in \mathbb{N}$.

(b) If $JH(z^k) \in \mathbb{R}^{l \times l}$ is a nonsingular matrix for all k , it follows that the linear system of equations (5.15) always has an exact solution \hat{d}^k . In particular, this exact solution satisfies the inexactness requirement from (5.16) for an arbitrary number $\eta_k \geq 0$. Furthermore, this exact solution also satisfies the descent property $\nabla\psi(z^k)^T \hat{d}^k < 0$, as shown below in Lemma 5.11, cf. [31]. It therefore follows that one can always find a vector d^k satisfying the two requirements (5.16) and (5.17), i.e., (S.2) is well-defined. Numerically this can be done by applying an iterative linear solver, like GMRES or one of its variants, to the exact equation system (5.15). Thus a suitable vector satisfying the inexactness requirement (5.16) and the descent property (5.17) can be found after a finite number of iterations.

(c) Since, by construction, we have $z^k \in Z_I$ for an arbitrary fixed iteration $k \in \mathbb{N}$ and since Z_I is an open set, we see that the test (5.18) holds for all sufficiently small stepsizes t_k . Furthermore, the Armijo line search from (5.19) is eventually satisfied since d^k is a descent direction of the potential function ψ in view of the construction in (S.2), cf. (5.17). In particular, this means that (S.3) is also well-defined.

Now we want to show that the exact solution of (5.16), that is the solution of (5.15), is under some assumptions on the sequence $\{\sigma_k\}$ a descent direction of the potential function ψ .

Lemma 5.11 *If $z^k \in \text{int}(Z)$, d^k satisfies the exact equation (5.15), and the sequence $\{\sigma_k\}$ satisfies $\sigma_k \in [0, 1)$ for all $k \in \mathbb{N}$, it holds that*

$$\nabla\psi(z^k)^T d^k \leq 2(\zeta - m)(\sigma_k - 1) < 0,$$

in particular, d^k is a direction of descent of ψ at z^k .

Proof. Using the definition of ψ and the Newton-type equation (5.15), we have

$$\begin{aligned} \nabla\psi(z^k)^T d^k &= \nabla p(H(z^k))^T JH(z^k) d^k \\ &= \nabla p(H(z^k))^T \left[-H(z^k) + \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \right] \\ &= -\nabla p(H(z^k))^T H(z^k) + \sigma_k \frac{(a^T H(z^k)) (\nabla p(H(z^k))^T a)}{\|a\|^2}. \end{aligned}$$

Writing $(u, v) := H(z^k) \in \mathbb{R}^{n+2m}$, applying (5.13) to the first term as well as (5.14) and thereafter (5.13) to the second term, we obtain

$$\nabla\psi(z^k)^T d^k \leq -2(\zeta - m) + \sigma_k 2(\zeta - m) = 2(\zeta - m)(\sigma_k - 1) < 0,$$

which is the desired result. □

The following is the main convergence result for Algorithm 5.9, where, implicitly, we assume that Algorithm 5.9 does not terminate within a finite number of iterations with a solution of the constrained nonlinear system $\text{CE}(H, Z)$.

Theorem 5.12 *Assume that $JH(z)$ is nonsingular for all $z \in Z_I$, and that the two sequences $\{\sigma_k\}$ and $\{\eta_k\}$ from (S.2) of Algorithm 5.9 satisfy the conditions*

$$\limsup_{k \rightarrow \infty} \sigma_k < \bar{\sigma} \quad \text{and} \quad \lim_{k \rightarrow \infty} \eta_k = 0. \quad (5.20)$$

Let $\{z^k\}$ be any sequence generated by Algorithm 5.9. Then:

- (a) The sequence $\{H(z^k)\}$ is bounded.
- (b) Any accumulation point of $\{z^k\}$ is a solution of (5.10).

Proof. We first note that our assumptions together with Remark 5.10 (b), (c) guarantee that Algorithm 5.9 is at least well-defined. Throughout this proof, we use the abbreviation $u^k := H(z^k)$ for all $k \in \mathbb{N}$.

(a) Suppose that $\{u^k\}$ is unbounded. Subsequencing if necessary, we may assume without loss of generality that $\lim_{k \rightarrow \infty} \|u^k\| = \infty$. Since $\{u^k\} \subseteq \text{int}(S)$ in view of Remark 5.10 (a), an elementary calculation then shows that $\lim_{k \rightarrow \infty} p(u^k) = \infty$. However, since d^k is a descent step for ψ , it follows from the definition of the potential function ψ together with the line search rule from (5.19) that

$$p(u^k) = p(H(z^k)) = \psi(z^k) < \psi(z^{k-1}) < \dots < \psi(z^0),$$

and this contradiction completes the proof of part (a).

(b) Let z^∞ be an accumulation point of the sequence $\{z^k\}$, and let $\{z^k\}_K$ be a corresponding subsequence converging to z^∞ . Since $z^k \in \text{int}(Z)$ for all $k \in \mathbb{N}$, cf. Remark 5.10 (a), it follows that $z^\infty \in Z$ since Z is a closed set. Define $u^\infty := H(z^\infty)$ and assume, by contradiction, that $u^\infty \neq 0$. In view of part (a) and assumption (5.20), we may assume without loss of generality that

$$\lim_{k \in K} \sigma_k = \sigma_\infty \text{ for some } \sigma_\infty \in [0, \bar{\sigma}) \quad \text{and} \quad \lim_{k \in K} u^k = u^\infty \neq 0.$$

Hence there exists an $\varepsilon > 0$ such that $\|u^k\| \geq \varepsilon$ holds for all $k \in K$. Furthermore, the proof of part (a) also shows that $p(u^k) \leq \delta$ for all $k \in K$ with $\delta := \psi(z^0)$. This means that the sequence $\{u^k\}$ belongs to the set

$$\Lambda(\varepsilon, \delta) := \{u \in \text{int}(S) \mid p(u) \leq \delta, \|u\| \geq \varepsilon\}$$

which is a compact set. Hence we have $u^\infty = H(z^\infty) \in \Lambda(\varepsilon, \delta) \subseteq \text{int}(S)$. Consequently, we have $z^\infty \in H^{-1}(\text{int}(S)) \cap Z$. However, since $H^{-1}(\text{int}(S)) \cap \text{bd}(Z) \subseteq \text{int}(Z) \cap \text{bd}(Z) = \emptyset$, it therefore follows that z^∞ belongs to the set $H^{-1}(\text{int}(S)) \cap \text{int}(Z) = Z_I$.

We now claim that the subsequence $\{d^k\}_{k \in K}$ is also bounded. To this end, let us define the residuals

$$r^k := H(z^k) + JH(z^k)d^k - \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \quad \text{for all } k \in \mathbb{N}. \quad (5.21)$$

Then the inexactness requirement (5.16) can be written as

$$\|r^k\| \leq \eta_k \|H(z^k)\| \quad \text{for all } k \in \mathbb{N}. \quad (5.22)$$

Since the Jacobian $JH(z^k)$ is nonsingular at $z^k \in Z_I$, we obtain from (5.21) that

$$d^k = JH(z^k)^{-1} \left[r^k - H(z^k) + \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \right] \quad \text{for all } k \in \mathbb{N}. \quad (5.23)$$

Since $\{z^k\}_{k \in K} \rightarrow z^\infty$, the continuity of the Jacobian implies that $\{JH(z^k)\}_{k \in K} \rightarrow JH(z^\infty)$. However, since we already know that z^∞ belongs to the set Z_I , it follows that $JH(z^\infty)$ is nonsingular. This implies that there exists a constant $\omega > 0$ such that $\|JH(z^k)^{-1}\| \leq \omega$ for all $k \in K$ sufficiently large. We then obtain from (5.23) and the Cauchy-Schwarz inequality that

$$\|d^k\| \leq \omega(\eta_k + 1 + \sigma_k) \|H(z^k)\|$$

for all $k \in K$ sufficiently large. Since $\{\|H(z^k)\|\}$ is bounded by part (a), we immediately get from (5.20) that the sequence $\{d^k\}_{k \in K}$ is also bounded. Without loss of generality, we may therefore assume that $\lim_{k \in K} d^k = d^\infty$ for some vector d^∞ . Using statement (a) once again together with $\eta_k \rightarrow 0$, it follows from (5.22) that $r^k \rightarrow 0$. On the other hand, using the definition of the residuum r^k and taking the limit $k \rightarrow \infty$ on the subset $K \subseteq \mathbb{N}$, it follows that

$$0 = H(z^\infty) + JH(z^\infty)d^\infty - \sigma_\infty \frac{a^T H(z^\infty)}{\|a\|^2} a.$$

Recalling that $z^\infty \in Z_I$ by assumption, we obtain $\nabla\psi(z^\infty)^T d^\infty < 0$ by Lemma 5.11. The convergence of $\{z^k\}_{k \in K}$ to z^∞ together with the continuity of ψ on the set Z_I implies that the subsequence $\{\psi(z^k)\}_{k \in K}$ also converges. On the other hand, the Armijo rule (5.19) implies that the entire sequence $\{\psi(z^k)\}_{k \in \mathbb{N}}$ is monotonically decreasing. This shows that the whole sequence $\{\psi(z^k)\}_{k \in \mathbb{N}}$ converges. Using the Armijo line search rule (5.19) once more, we have

$$\psi(z^{k+1}) - \psi(z^k) \leq \gamma t_k \nabla\psi(z^k)^T d^k < 0$$

for all $k \in \mathbb{N}$. Since the left-hand side converges to zero, $\lim_{k \rightarrow \infty} t_k \nabla\psi(z^k)^T d^k = 0$ must hold. This, in turn, implies $\lim_{k \in K} t_k = 0$ since

$$\lim_{k \in K} \nabla\psi(z^k)^T d^k = \nabla\psi(z^\infty)^T d^\infty < 0.$$

Let $\ell_k \in \mathbb{N}_0$ be the unique index such that $t_k = \beta^{\ell_k}$ holds in (S.3) for all $k \in \mathbb{N}$. With $\lim_{k \in K} t_k = 0$, we also have $\lim_{k \in K} \frac{t_k}{\beta} = 0$. Since the limit point z^∞ belongs to the open

set Z_I , it therefore follows that the sequence $\{z^k + \frac{t_k}{\beta} d^k\}_{k \in K}$ also belongs to this set, at least for all sufficiently large $k \in K$. Consequently, for these $k \in K$, the line search test in (5.19) fails for the stepsize $\frac{t_k}{\beta} = \beta^{\ell_k - 1}$. We therefore have

$$\frac{\psi(z^k + \beta^{\ell_k - 1} d^k) - \psi(z^k)}{\beta^{\ell_k - 1}} > \gamma \nabla \psi(z^k)^T d^k$$

for all $k \in K$ sufficiently large. Taking the limit $k \rightarrow \infty$ on the subset K , the continuous differentiability of the potential function ψ on the set Z_I then gives

$$\nabla \psi(z^\infty)^T d^\infty \geq \gamma \nabla \psi(z^\infty)^T d^\infty.$$

Since $\nabla \psi(z^\infty)^T d^\infty < 0$, this is only possible if $\gamma \geq 1$, a contradiction to the choice of $\gamma \in (0, 1)$. Consequently, we have $0 = u^\infty = H(z^\infty)$, and z^∞ is a solution of the constrained system of nonlinear equations (5.10). \square

Note that the previous convergence result requires the Jacobian matrices $JH(z)$ to be nonsingular for all $z \in Z_I$ (an assumption that will be discussed in the next section), however, it does not state any assumptions for the limit points that might not belong to Z_I . In fact, the above convergence result also holds when the Jacobian is singular at a limit point. This singularity of the Jacobian, however, also indicates that in general we cannot expect local fast convergence of our interior point method since Newton-type methods for nonlinear systems typically require a nonsingular Jacobian at the solution in order to achieve a local superlinear/quadratic rate of convergence. This sounds like a disadvantage compared to some other Newton-type methods, however, we recall that these Newton-type methods also have severe troubles in basically all interesting situations where at least one joint constraint is active at the solution since then singularity problems arise, cf. [26]. Hence, also these Newton methods are not quadratically convergent, and the rate of convergence may actually slow down dramatically.

5.2.1. Nonsingularity Conditions

The critical issue in applying Theorem 5.12 is establishing the nonsingularity of JH . This section is devoted to this issue. We will see that while the conditions we will use in order to establish the nonsingularity of JH are similar to those obtained in the equation reformulation approach in Section 5.1.1 they only need to be valid for positive values of λ .

The structure of $JH(z)$ is the following

$$JH(z) := \begin{pmatrix} J_x F(x, \lambda) & E(x) & 0 \\ J_x g(x) & 0 & I_m \\ 0 & W & \Lambda \end{pmatrix}, \quad (5.24)$$

with

$$W := \text{diag}(w), \quad \Lambda := \text{diag}(\lambda),$$

and E defined in (5.3). In order to analyse the nonsingularity of this matrix, we first introduce the following terminology, cf. [31].

Definition 5.13 *A matrix $Q = (M_1 M_2 M_3)$ is said to have the mixed P_0 -property if M_3 has full column rank and*

$$\left. \begin{array}{l} M_1 u + M_2 v + M_3 s = 0 \\ (u, v) \neq 0 \end{array} \right\} \implies u_i v_i \geq 0 \text{ for some } i \text{ such that } |u_i| + |v_i| > 0.$$

Note that the matrix M_3 in the previous definition might vanish. Then it is easy to see that a square matrix M is a P_0 -matrix if and only if the pair $[M \quad -I]$ (with a vacuous M_3 -part) has the mixed P_0 -property. Therefore Definition 5.13 generalizes the standard notion of a P_0 -matrix. A useful characterization of the mixed P_0 -property is given in [31, Lemma 11.4.3] and restated in the following result.

Lemma 5.14 *Let M_1 and M_2 be matrices of order $(n + m) \times m$ and M_3 be a matrix of order $(n + m) \times n$. The matrix $Q = (M_1 M_2 M_3)$ has the mixed P_0 -property if and only if for every pair of $m \times m$ diagonal matrices D_1 and D_2 both having positive diagonal entries, the $(2m + n) \times (2m + n)$ square matrix*

$$M := \begin{pmatrix} D_1 & D_2 & 0 \\ M_1 & M_2 & M_3 \end{pmatrix}$$

is nonsingular.

Note that this Lemma is immediately applicable to (5.24) and gives a necessary and sufficient condition for the nonsingularity of JH when $\lambda > 0$ and $w > 0$. However, the mixed P_0 -property is difficult to interpret and to verify. Therefore we now give some sufficient conditions which are derived taking into account the GNEP structure and which lead more easily to verification and comparison with previous results. The proofs of these results may be carried out by referring to Lemma 5.14, however, we prefer to give direct proofs to be independent of that result, and because the direct proofs are not really longer than those based on Lemma 5.14.

The following theorem gives a first nonsingularity result.

Theorem 5.15 *Let $z = (x, \lambda, w) \in \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ be given such that $J_x F(x, \lambda)$ is nonsingular and*

$$M(x, \lambda) := J_x g(x) J_x F(x, \lambda)^{-1} E(x) \tag{5.25}$$

is a P_0 -matrix. Then the Jacobian $JH(z)$ is nonsingular.

Proof. Using the structure of $JH(z)$ the homogeneous linear system $JH(z)q = 0$, with $q = (q^{(1)}, q^{(2)}, q^{(3)})$ being partitioned in a suitable way, can be rewritten in the following way:

$$J_x F(x, \lambda) q^{(1)} + E(x) q^{(2)} = 0, \tag{5.26}$$

$$J_x g(x)q^{(1)} + q^{(3)} = 0, \quad (5.27)$$

$$Wq^{(2)} + \Lambda q^{(3)} = 0. \quad (5.28)$$

Since $J_x F(x, \lambda)$ is nonsingular by assumption, (5.26) yields

$$q^{(1)} = -J_x F(x, \lambda)^{-1} E(x)q^{(2)}.$$

Hence we obtain

$$q^{(3)} = -J_x g(x)q^{(1)} = J_x g(x)J_x F(x, \lambda)^{-1} E(x)q^{(2)} = M(x, \lambda)q^{(2)}$$

from (5.27) and the definition of $M(x, \lambda)$. Substituting this expression into (5.28) gives

$$[W + \Lambda M(x, \lambda)]q^{(2)} = 0.$$

Since $M(x, \lambda)$ is a P_0 -matrix by assumption and $w, \lambda > 0$, it follows that $[W + \Lambda M(x, \lambda)]$ is nonsingular and hence $q^{(2)} = 0$. This, in turn, implies $q^{(1)} = 0$ and $q^{(3)} = 0$. Consequently, $JH(z)$ is nonsingular. \square

Note that the conditions of Theorem 5.15 are identical to the assumptions for the stationarity condition in Theorem 5.1. The difference is that the multipliers are now guaranteed to be positive in the interior point approach, whereas this condition was crucial in the equation/merit function approach, cf. the corresponding discussion in Section 5.1.1.

To illustrate this point, let us consider once again Example 5.4: It is now easy to see that this example satisfies the conditions of Theorem 5.15:

$$J_x F(x, \lambda) = \begin{pmatrix} 2\lambda & 0 \\ 0 & 1 \end{pmatrix}$$

is nonsingular for all $\lambda > 0$ and $M(x, \lambda) = \frac{2(x^1)^2}{\lambda} \geq 0$ for all $\lambda > 0$. Hence this example is no longer a counterexample for our interior point approach.

The following theorem gives another sufficient condition for the nonsingularity of JH . This condition is stronger than that in Theorem 5.15, nevertheless it is interesting because it gives a quantitative insight into what is necessary to guarantee the nonsingularity of JH .

Theorem 5.16 *Let $z = (x, \lambda, w) \in \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ be given such that $J_x F(x, \lambda)$ is nonsingular and*

$$\begin{aligned} \text{eig}_{\min} \left(\frac{1}{2} E(x)^T (J_x F(x, \lambda)^{-1} + J_x F(x, \lambda)^{-T}) E(x) \right) \geq \\ \|J_x g(x) - E(x)^T\| \|J_x F(x, \lambda)^{-1}\| \|E(x)\|. \end{aligned}$$

Then the Jacobian $JH(z)$ is nonsingular.

Proof. For all $u \in \mathbb{R}^m$ we have

$$\begin{aligned}
 u^T E(x)^T J_x F(x, \lambda)^{-1} E(x) u &= \frac{1}{2} u^T \left(E(x)^T (J_x F(x, \lambda)^{-1} + J_x F(x, \lambda)^{-T}) E(x) \right) u \\
 &\geq \operatorname{eig}_{\min} \left(\frac{1}{2} E(x)^T (J_x F(x, \lambda)^{-1} + J_x F(x, \lambda)^{-T}) E(x) \right) \|u\|^2 \\
 &\geq \|J_x g(x) - E(x)^T\| \|J_x F(x, \lambda)^{-1}\| \|E(x)\| \|u\|^2 \\
 &\geq |u^T (J_x g(x) - E(x)^T) J_x F(x, \lambda)^{-1} E(x) u| \\
 &\geq -u^T (J_x g(x) - E(x)^T) J_x F(x, \lambda)^{-1} E(x) u.
 \end{aligned}$$

Using the matrix $M(x, \lambda)$ from (5.25), this implies that

$$u^T M(x, \lambda) u = u^T J_x g(x) J_x F(x, \lambda)^{-1} E(x) u \geq 0$$

for all $u \in \mathbb{R}^m$. Therefore $M(x, \lambda)$ is positive semidefinite, hence a P_0 -matrix, and Theorem 5.15 guarantees nonsingularity of $JH(z)$. \square

In the case of a NEP, if $J_x F(x, \lambda)$ is positive definite, the matrix $M(x, \lambda)$ from (5.25) is automatically P_0 (actually, positive semidefinite) since $J_x g(x) = E(x)^T$ in this case. Then it may be interesting to see that in the case of a NEP we can relax a bit the nonsingularity assumption on $J_x F(x, \lambda)$ and still get nonsingularity of $JH(z)$. In fact, we have the following counterpart of the stationary point condition from Theorem 5.5.

Theorem 5.17 *Consider a NEP, and let $z = (x, \lambda, w) \in \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ be given such that $J_x F(x, \lambda)$ is positive semidefinite and it holds that*

$$d^T J_x F(x, \lambda) d > 0, \quad \text{for all } d \in \{d \in \mathbb{R}^n \setminus \{0\} \mid E(\bar{x})^T d = 0\}.$$

Then the Jacobian $JH(z)$ is nonsingular.

Proof. Consider once again the homogeneous linear system $JH(z)q = 0$, so that (5.26)–(5.28) hold with $J_x g(x) = E(x)^T$, since we are in the NEP case. Since $\lambda \in \mathbb{R}_{++}^m$, (5.28) can be solved for $q^{(3)}$ and we obtain

$$\begin{aligned}
 0 &\stackrel{(5.26)}{=} (q^{(1)})^T J_x F(x, \lambda) q^{(1)} + (q^{(1)})^T E(x) q^{(2)} \\
 &\stackrel{(5.27)}{=} (q^{(1)})^T J_x F(x, \lambda) q^{(1)} - (q^{(3)})^T q^{(2)} \\
 &\stackrel{(5.28)}{=} (q^{(1)})^T J_x F(x, \lambda) q^{(1)} + (q^{(2)})^T (W \Lambda^{-1}) q^{(2)}.
 \end{aligned}$$

Positive semidefiniteness of $J_x F(x, \lambda)$, together with $w > 0, \lambda > 0$, implies $q^{(2)} = 0$ and thus also $q^{(3)} = 0$ by (5.28). Then we have from (5.26) and (5.27)

$$(q^{(1)})^T J_x F(x, \lambda) q^{(1)} = 0 \quad \text{and} \quad E(x)^T q^{(1)} = 0,$$

and the assumptions show $q^{(1)} = 0$, hence nonsingularity of $JH(z)$. \square

In spite of the result above, it should be pointed out that in general, in Theorem 5.15, we do not need the matrix $J_x F(x, \lambda)$ to be positive (semi-) definite. This is illustrated by the following example.

Example 5.18 Consider a GNEP with two players, each controlling a single variable, that is given by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1)^2 - 2x^1 & \text{s.t. } & (x^1)^2 + x^2 \leq 0, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2)^2 + (2 - (x^1)^2)x^2 & \text{s.t. } & x^2 \in \mathbb{R}. \end{aligned}$$

It is easy to see that

$$J_x F(x, \lambda) = \begin{pmatrix} 1 + 2\lambda & 0 \\ -2x^1 & 1 \end{pmatrix}$$

is nonsingular for all $x \in \mathbb{R}^2$ and all $\lambda > 0$ but it is not positive semidefinite everywhere. However, since

$$M(x, \lambda) = J_x g(x) J_x F(x, \lambda)^{-1} E(x) = (2x^1, 1) \begin{pmatrix} \frac{1}{1+2\lambda} & 0 \\ \frac{2x^1}{1+2\lambda} & 1 \end{pmatrix} \begin{pmatrix} 2x^1 \\ 0 \end{pmatrix} = \frac{8(x^1)^2}{(1+2\lambda)} \geq 0,$$

it follows that the conditions from Theorem 5.15 are satisfied. \diamond

As we have seen in the previous sections, nonsingularity of $J_x F(x, \lambda)$ and the P_0 -condition on the matrix $M(x, \lambda)$ guarantee both that stationary points of the merit function are solutions of the GNEP and that the matrix $JH(z)$ is nonsingular. In the case of NEPs we obtain these properties by some semi-definiteness assumptions on $J_x F(x, \lambda)$. Let us recall that in the context of the interior point approach, all conditions only have to hold for positive λ and, therefore, are less restrictive than in the merit function context.

To conclude this subsection, we remark that the matrix $JH(z)$ is not positive semidefinite.

Remark 5.19 A necessary condition for positive semidefiniteness of

$$JH(z) := \begin{pmatrix} J_x F(x, \lambda) & E(x) & 0 \\ J_x g(x) & 0 & I_m \\ 0 & W & \Lambda \end{pmatrix}$$

is positive semidefiniteness of the lower right block or equivalently of its symmetric part

$$\frac{1}{2} \left(\begin{pmatrix} 0 & I_m \\ W & \Lambda \end{pmatrix} + \begin{pmatrix} 0 & I_m \\ W & \Lambda \end{pmatrix}^T \right) = \begin{pmatrix} 0 & \frac{1}{2}(I_m + W) \\ \frac{1}{2}(I_m + W) & \Lambda \end{pmatrix}.$$

But elementary calculations show that the eigenvalues of this matrix are given by

$$\left\{ \frac{\lambda_i \pm \sqrt{(\lambda_i)^2 + (1 + w_i)^2}}{2} \mid i = 1, \dots, m \right\}.$$

Therefore, if $w_i = -1$ and $\lambda_i \geq 0$ for all $i = 1, \dots, m$ does not hold, there are negative eigenvalues and the matrix $JH(z)$ is not positive semidefinite. In particular for all $z^k \in \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ the matrix $JH(z^k)$ is never positive semidefinite.

Since the matrix $JH(z)$ is not positive semidefinite, the mapping H is not monotone. Therefore a VI approach for the solution of the constrained equation $\text{CE}(H, Z)$ is difficult, since it often requires some monotonicity assumption.

5.2.2. Boundedness

Note that Theorem 5.12 does not guarantee the existence of an accumulation point of the sequence generated by Algorithm 5.9. The following result therefore considers precisely this question and provides conditions under which the entire sequence remains bounded.

Theorem 5.20 *Assume that*

- (a) $JH(z)$ is nonsingular for all $z \in Z_I$;
- (b) $\lim_{\|x\| \rightarrow \infty} \|g_+(x)\| = +\infty$;
- (c) the Extended Mangasarian-Fromovitz Constraint Qualification (EMFCQ) holds for each player, i.e., for all $\nu = 1, \dots, N$ and for all $x \in \mathbb{R}^n$ there exists a $d^\nu \in \mathbb{R}^{n_\nu}$ such that

$$\nabla_{x^\nu} g_i^\nu(x)^T d^\nu < 0 \quad \text{for all } i \in I_{\geq}^\nu(x) := \{i \in \{1, \dots, m_\nu\} \mid g_i^\nu(x) \geq 0\}.$$

Then any sequence $\{(x^k, \lambda^k, w^k)\}$ generated by Algorithm 5.9 remains bounded.

Proof. Assume that there exists a sequence $\{(x^k, \lambda^k, w^k)\} \subseteq Z_I$ that is unbounded, i.e. $\lim_{k \rightarrow \infty} \|(x^k, \lambda^k, w^k)\| = \infty$. We will show that this implies $\|H(x^k, \lambda^k, w^k)\| \rightarrow \infty$ for $k \rightarrow \infty$, contradicting part (a) of Theorem 5.12. We consider two cases.

Case 1: $\|(x^k, w^k)\| \rightarrow \infty$. Then either $\{x^k\}$ is bounded, or $\|x^k\| \rightarrow \infty$, without loss of generality on the entire sequence. If $\{x^k\}$ is bounded, then, subsequencing if necessary, we have $\|w^k\| \rightarrow \infty$. By the continuity of g the sequence $\{g(x^k)\}$ is bounded. We therefore obtain $\|g(x^k) + w^k\| \rightarrow \infty$. This, in turn, implies $\|H(x^k, \lambda^k, w^k)\| \rightarrow \infty$.

On the other hand, if $\|x^k\| \rightarrow \infty$, it follows from assumption (b) that $\|g_+(x^k)\| \rightarrow \infty$. Moreover, since all components of the vector w^k are positive, this also implies $\|g(x^k) + w^k\| \rightarrow \infty$, and it follows once again that $\|H(x^k, \lambda^k, w^k)\| \rightarrow \infty$ also in this (sub-) case.

Case 2: $\|(x^k, w^k)\|$ is bounded. Then we have $\|\lambda^k\| \rightarrow \infty$. Let ν be a player such that $\|(\lambda^k)^\nu\| \rightarrow \infty$, and let J be the set of indices such that $(\lambda^k)_j^\nu \rightarrow \infty$, whereas, subsequencing if necessary, we can assume that the remaining components stay bounded. Without loss of generality, we may also assume that $x^k \rightarrow \bar{x}$ and $w^k \rightarrow \bar{w}$. If, for some $j \in J$, we have $\bar{w}_j^\nu > 0$, it follows that $(\lambda^k)_j^\nu (w^k)_j^\nu \rightarrow +\infty$, and therefore $\|H(x^k, \lambda^k, w^k)\| \rightarrow \infty$. Hence it remains to consider the case where $\bar{w}_j^\nu = 0$ for all $j \in J$. Since (x^k, λ^k, w^k) belongs to Z_I , we have $g_j^\nu(x^k) + (w^k)_j^\nu > 0$ and, therefore, $g_j^\nu(\bar{x}) \geq 0$ for all $j \in J$. Now we can use the EMFCQ condition and repeat the corresponding part of the proof of Theorem 5.7 to get $\|\nabla_{x^\nu} L^\nu(x^k, (\lambda^k)^\nu)\| \rightarrow +\infty$ which, in turn, implies $\|H(x^k, \lambda^k, w^k)\| \rightarrow \infty$ for $k \rightarrow \infty$, also in this case. \square

5.2.3. A Finite Termination Result

In this section we want to find a relation between the number of iterations until the termination criterion $\|H(z^k)\| \leq \varepsilon$ is satisfied and the accuracy $\varepsilon > 0$. For this purpose we assume throughout this section that the exact version of Algorithm 5.9 is implemented, that is we assume $\eta_k = 0$ for all $k \in \mathbb{N}$. Let us begin with a relation between the two quantities $\|H(z)\|$ and $\Psi(z)$.

Lemma 5.21 *It holds that*

$$\|H(z)\|^2 \leq \exp\left(\frac{\psi(z)}{\zeta - m}\right).$$

Proof. By definition and $\ln(x^2) = 2 \ln(x)$, we have

$$\begin{aligned} \psi(z) &= \zeta \ln(\|H(z)\|^2) - \sum_{i=1}^{2m} \ln(H_{n+i}(z)) \\ &= (\zeta - m) \ln(\|H(z)\|^2) + 2m \ln(\|H(z)\|) - \sum_{i=1}^{2m} \ln(H_{n+i}(z)) \\ &\geq (\zeta - m) \ln(\|H(z)\|^2), \end{aligned}$$

which immediately implies the statement. \square

Using this Lemma we can prove the following relation between $\|H(z^k)\|$ and the stepsizes.

Lemma 5.22 *If $\eta_k = 0$ for all $k \in \mathbb{N}$ and $\tilde{\sigma} := \sup_{k \in \mathbb{N}} \sigma_k < 1$ we have*

$$\|H(z^k)\|^2 \leq \exp\left(\frac{\psi(z^0)}{\zeta - m} - 2\gamma(1 - \tilde{\sigma}) \sum_{\ell=0}^{k-1} t_\ell\right)$$

for all $k \in \mathbb{N}$.

Proof. By the Armijo line-search rule, Lemma 5.11, and $\tilde{\sigma} := \sup_{k \in \mathbb{N}} \sigma_k < 1$ we have

$$\begin{aligned} \psi(z^{k+1}) - \psi(z^k) &\leq \gamma t_k \nabla \psi(z^k)^T d^k \\ &\leq 2\gamma(\zeta - m)t_k(\sigma_k - 1) \\ &\leq -2\gamma(\zeta - m)(1 - \tilde{\sigma})t_k. \end{aligned}$$

Summing up this inequality for $k = 0, 1, 2, \dots$ yields

$$\psi(z^k) - \psi(z^0) \leq -2\gamma(\zeta - m)(1 - \tilde{\sigma}) \sum_{\ell=0}^{k-1} t_\ell.$$

Lemma 5.21 therefore gives

$$\begin{aligned} \|H(z^k)\|^2 &\leq \exp\left(\frac{\psi(z^k)}{\zeta - m}\right) \\ &\leq \exp\left(\frac{\psi(z^0)}{\zeta - m} - 2\gamma(1 - \tilde{\sigma}) \sum_{\ell=0}^{k-1} t_\ell\right). \end{aligned}$$

□

In view of Lemma 5.22 we see that for $\varepsilon > 0$ the termination criterion $\|H(z^k)\| \leq \varepsilon$ is certainly satisfied for some $k \in \mathbb{N}$ if

$$\exp\left(\frac{\psi(z^0)}{\zeta - m} - 2\gamma(1 - \tilde{\sigma}) \sum_{\ell=0}^{k-1} t_\ell\right) \leq \sqrt{\varepsilon}$$

holds. This is equivalent to

$$\sum_{\ell=0}^{k-1} t_\ell \geq \frac{\psi(z^0)}{2\gamma(\zeta - m)(1 - \tilde{\sigma})} - \frac{\ln \varepsilon}{4\gamma(1 - \tilde{\sigma})},$$

which will be satisfied for some $k \in \mathbb{N}$ if $\sum_{\ell \in \mathbb{N}} t_\ell = \infty$. In particular, it will be satisfied if there is a lower bound $\bar{t} > 0$ on the sequence of stepsizes, such that $t_k \geq \bar{t}$ for all $k \in \mathbb{N}$. From this stronger assumption we similarly obtain that

$$k\bar{t} \geq \frac{\psi(z^0)}{2\gamma(\zeta - m)(1 - \tilde{\sigma})} - \frac{\ln \varepsilon}{4\gamma(1 - \tilde{\sigma})}$$

implies $\|H(z^k)\| \leq \varepsilon$. This proves the following result, where the notation $\lceil a \rceil$ is used for the smallest integer that is larger than $a \in \mathbb{R}$.

Theorem 5.23 *Let $\{z^k\}$ be a sequence generated by Algorithm 5.9 with $\eta_k = 0$ for all $k \in \mathbb{N}$, and suppose that $\tilde{\sigma} := \sup_{k \in \mathbb{N}} \sigma_k < 1$ and $t_k \geq \bar{t}$ holds for all $k \in \mathbb{N}$ for some lower bound $\bar{t} > 0$. Then the termination criterion $\|H(z^k)\| \leq \varepsilon$ is satisfied after at most*

$$k = \left\lceil \frac{\psi(z^0)}{2\gamma(\zeta - m)(1 - \tilde{\sigma})\bar{t}} - \frac{\ln \varepsilon}{4\gamma(1 - \tilde{\sigma})\bar{t}} \right\rceil$$

iterations.

For the application of this theorem we have to prove the existence of a lower bound $\bar{t} > 0$ for the sequence of the stepsizes. This is the content of the remaining part of this section. Therefore let us first introduce some notation. Since we consider the exact version of Algorithm 5.9, there is a linear system of equations to be solved at iteration k . Partitioning the vector $d^k := ((d_x^k)^T, (d_\lambda^k)^T, (d_w^k)^T)^T \in \mathbb{R}^{n+m+m}$, this system becomes

$$J_x F(x^k, \lambda^k) d_x^k + E(x^k) d_\lambda^k = -F(x^k, \lambda^k), \quad (5.29)$$

$$J_x g(x^k) d_x^k + d_w^k = -g(x^k) - w^k + \sigma_k \mu_k \mathbf{1}_m, \quad (5.30)$$

$$W^k d_\lambda^k + \Lambda^k d_w^k = -\Lambda^k W^k \mathbf{1}_m + \sigma_k \mu_k \mathbf{1}_m, \quad (5.31)$$

where we used $W^k := \text{diag}(w^k) \in \mathbb{R}_{++}^{m \times m}$, $\Lambda^k := \text{diag}(\lambda^k) \in \mathbb{R}_{++}^{m \times m}$, and

$$\mu_k := \frac{a^T H(z^k)}{\|a\|^2}.$$

Further define

$$\frac{d_\lambda^k}{\lambda^k} := (\Lambda^k)^{-1} d_\lambda^k = \left(\frac{d_{\lambda,1}^k}{\lambda_1^k}, \dots, \frac{d_{\lambda,m}^k}{\lambda_m^k} \right)^T \in \mathbb{R}^m,$$

$$\frac{d_w^k}{w^k} := (W^k)^{-1} d_w^k = \left(\frac{d_{w,1}^k}{w_1^k}, \dots, \frac{d_{w,m}^k}{w_m^k} \right)^T \in \mathbb{R}^m.$$

By (S.3) of Algorithm 5.9 the stepsize t_k must satisfy two conditions. The first one is $z^k + t_k d^k \in Z_I$, which requires

$$\begin{aligned} \lambda_i^k + t_k d_{\lambda,i}^k &> 0, \\ w_i^k + t_k d_{w,i}^k &> 0, \\ (\lambda_i^k + t_k d_{\lambda,i}^k) (w_i^k + t_k d_{w,i}^k) &> 0, \\ g_i(x^k + t_k d_x^k) + (w_i^k + t_k d_{w,i}^k) &> 0 \end{aligned}$$

for all $i = 1, \dots, m$. Obviously if two of the first three inequalities hold also the third one holds. Assuming convexity of the constraint functions and boundedness of some sequences we can show the following theorem.

Theorem 5.24 *Let the constraint functions $g_i, i = 1, \dots, m$ be convex and let $\{z^k\}$ be a sequence generated by Algorithm 5.9 with $\eta_k = 0$ for all $k \in \mathbb{N}$. Assume there are constants $c_\lambda > 0$ and $c_w > 0$ such that*

$$\frac{d_{\lambda,i}^k}{\lambda_i^k} \leq c_\lambda \quad \text{and} \quad \frac{d_{w,i}^k}{w_i^k} \leq c_w \quad \text{for all } k \in \mathbb{N}, i = 1, \dots, m. \quad (5.32)$$

Then we have $z^k + t_k d^k \in Z_I$ for all $t_k \in [0, \tilde{t}]$ with $\tilde{t} := \min \left\{ \frac{1}{1+c_\lambda}, \frac{1}{1+c_w} \right\}$.

Proof. Writing equation (5.31) for each component and solving it by $d_{w,i}^k$ we obtain for all $i \in \{1, \dots, m\}$

$$\begin{aligned} w_i^k + t_k d_{w,i}^k &= w_i^k + t_k \left[-w_i^k \left(1 + \frac{d_{\lambda,i}^k}{\lambda_i^k} \right) + \frac{\sigma_k \mu_k}{\lambda_i^k} \right] \\ &= w_i^k \left[1 - t_k \left(1 + \frac{d_{\lambda,i}^k}{\lambda_i^k} \right) \right] + t_k \frac{\sigma_k \mu_k}{\lambda_i^k}. \end{aligned}$$

If $\frac{d_{\lambda,i}^k}{\lambda_i^k} \leq 0$ we certainly have $1 - t_k \left(1 + \frac{d_{\lambda,i}^k}{\lambda_i^k} \right) \geq 0$ for all $t_k \in [0, 1]$ and hence $w_i^k + t_k d_{w,i}^k > 0$. Otherwise we have, by the definition of \tilde{t} and $\frac{d_{\lambda,i}^k}{\lambda_i^k} \leq c_\lambda$,

$$t_k \left(1 + \frac{d_{\lambda,i}^k}{\lambda_i^k} \right) \leq \frac{1}{1 + c_\lambda} (1 + c_\lambda) = 1$$

and hence $w_i^k + t_k d_{w,i}^k > 0$ for all $t_k \in [0, \tilde{t}]$. Similarly we obtain for all $i \in \{1, \dots, m\}$

$$\lambda_i^k + t_k d_{\lambda,i}^k = \lambda_i^k \left[1 - t_k \left(1 + \frac{d_{w,i}^k}{w_i^k} \right) \right] + t_k \frac{\sigma_k \mu_k}{w_i^k},$$

and hence using $\frac{d_{w,i}^k}{w_i^k} \leq c_w$ we can show $\lambda_i^k + t_k d_{\lambda,i}^k > 0$ for all $t_k \in [0, \tilde{t}]$.

Furthermore we get

$$(w_i^k + t_k d_{w,i}^k) (\lambda_i^k + t_k d_{\lambda,i}^k) > 0 \quad \text{for all } t_k \in [0, \tilde{t}].$$

Using that all $g_i, i = 1, \dots, m$ are continuously differentiable and convex together with equation (5.30) we obtain

$$\begin{aligned} g_i(x^k + t_k d_x^k) + (w_i^k + t_k d_{w,i}^k) &\geq g_i(x^k) + t_k \nabla g_i(x^k)^T d_x^k + (w_i^k + t_k d_{w,i}^k) \\ &\stackrel{(5.30)}{=} (1 - t_k) (g_i(x^k) + w_i^k) + t_k \sigma_k \mu_k \end{aligned}$$

which is positive for all $t_k \in [0, 1]$ and all $i = 1, \dots, m$, since $z^k \in Z_I$ implies $g_i(x^k) + w_i^k > 0$. Altogether this proves $z^k + t_k d^k \in Z_I$ for all $t_k \in [0, \tilde{t}]$. \square

In Theorem 5.24 we assume boundedness (from above) for each component of $\left\{ \frac{d_\lambda^k}{\lambda^k} \right\}$. This assumption holds under suitable conditions, which we will derive next.

If $J_x F(x^k, \lambda^k)$ is nonsingular for all $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}_{++}^m$ one can solve (5.29) by d_x^k , (5.31) by d_w^k and insert this in (5.30), to obtain the equation

$$(M^k \Lambda^k + W^k) \frac{d_\lambda^k}{\lambda^k} = r^k \tag{5.33}$$

with

$$\begin{aligned} M^k &:= M(x^k, \lambda^k) = J_x g(x^k) J_x F(x^k, \lambda^k)^{-1} E(x^k), \\ r^k &:= -J_x g(x^k) J_x F(x^k, \lambda^k)^{-1} F(x^k, \lambda^k) + g(x^k) + \sigma_k \mu_k ((\Lambda^k)^{-1} - I_m) \mathbf{1}_m. \end{aligned} \quad (5.34)$$

As we have seen in Theorem 5.15 the nonsingularity of $J_x F(x^k, \lambda^k)$ and the P_0 property of M^k on $\mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ imply nonsingularity of $JH(z^k)$ and also of $(M^k \Lambda^k + W^k)$ for all $k \in \mathbb{N}$. Therefore we can use Cramer's Rule to get

$$\frac{d_{\lambda_i}^k}{\lambda_i^k} = \frac{\det \left(\tilde{M}_i^k \tilde{\Lambda}_i^k + \tilde{W}_i^k \right)}{\det (M^k \Lambda^k + W^k)}, \quad (5.35)$$

with

$$\begin{aligned} \tilde{\Lambda}_i^k &:= \text{diag} (\lambda_1^k, \dots, \lambda_{i-1}^k, 1, \lambda_{i+1}^k, \dots, \lambda_m^k), \\ \tilde{W}_i^k &:= \text{diag} (w_1^k, \dots, w_{i-1}^k, 0, w_{i+1}^k, \dots, w_m^k), \\ \tilde{M}_i^k &:= (M_1^k \dots M_{i-1}^k r^k M_{i+1}^k \dots M_m^k), \end{aligned}$$

where M_j^k denotes the j -th column of the matrix M^k .

If we assume that $J_x F(x, \lambda)$ is nonsingular for all (x, λ) in $\mathbb{R}^n \times \mathbb{R}_+^m$ (not only $\mathbb{R}^n \times \mathbb{R}_{++}^m$), and $\{z^k\}$ and $\sigma_k \mu_k (\Lambda^k)^{-1} \mathbf{1}_m$ are bounded, we get boundedness of r^k . Moreover, assuming strict complementarity at all solutions of the GNEP and M^k to be a P -matrix, implies nonsingularity of any limiting matrix of $\{M^k \Lambda^k + W^k\}$ and hence boundedness of $\left\{ \frac{d_{\lambda_i}^k}{\lambda_i^k} \right\}$. Unfortunately if there are repeated constraints for two or more players, which is in particular the case for jointly convex GNEPs, the matrix M^k has identical rows, and hence can not be nonsingular or even a P -matrix. Nevertheless, as we will see next, it is possible that the sequence $\left\{ \frac{d_{\lambda_i}^k}{\lambda_i^k} \right\}$ stays bounded.

Let us first introduce some notation. For a set $\alpha \subseteq \{1, \dots, m\}$, define

$$\bar{\alpha} := \{1, \dots, m\} \setminus \alpha.$$

In the subsequent analysis we consider a solution $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{w})$ of the constrained equation $H(z) = 0, z \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$ and we define the index sets

$$I_{<}(\bar{x}) := \{j \mid g_j(\bar{x}) < 0\} \quad \text{and} \quad I_{=}(\bar{x}) := \{j \mid g_j(\bar{x}) = 0\}.$$

All constraints with indices belonging to $I_{=}(\bar{x})$ are called active constraints, the ones with indices belonging to $I_{<}(\bar{x})$ are inactive constraints at \bar{x} . It is possible that we have a joint constraint for different players, meaning $g_i \equiv g_j$ for $i \neq j$. We are in particular interested in index sets, where all these repeated active constraints are dropped except one. Thus let us define the maximum number $q(\bar{x}) \leq m$ of different active constraints that we have, by

$$q(\bar{x}) := \max \{|\alpha| \mid \alpha \subseteq I_{=}(\bar{x}), \quad g_i \not\equiv g_j \quad \forall i, j \in \alpha\}.$$

Depending on $q(\bar{x})$ and $|I_=(\bar{x})|$ define the following sets of index sets:

$$\begin{aligned}
 I_{=}^0(\bar{x}) &:= \{\alpha \subseteq I_=(\bar{x}) \mid |\alpha| = q(\bar{x}), \quad g_i \not\equiv g_j \quad \forall i, j \in \alpha\}; \\
 I_{=}^{-k}(\bar{x}) &:= \{\alpha \subseteq I_=(\bar{x}) \mid \exists \beta_\alpha \in I_{=}^0(\bar{x}) \exists i_1, \dots, i_k \in \beta_\alpha : \alpha = \beta_\alpha \setminus \{i_1, \dots, i_k\}\}, \\
 &\quad \text{for } 1 \leq k \leq q(\bar{x}); \\
 I_{=}^1(\bar{x}) &:= \{\alpha \subseteq I_=(\bar{x}) \mid \exists \beta_\alpha \in I_{=}^0(\bar{x}) \exists i, j \in \alpha, i \neq j : g_i \equiv g_j, \alpha \setminus \{i\} \subseteq \beta_\alpha\}, \\
 &\quad \text{if } |I_=(\bar{x})| > q(\bar{x}); \\
 I_{=}^k(\bar{x}) &:= \{\alpha \subseteq I_=(\bar{x}) \mid \alpha \not\subseteq I_{=}^{k-1}(\bar{x}), \exists \beta_\alpha \in I_{=}^{k-1}(\bar{x}) \exists i \in \alpha : \alpha \setminus \{i\} = \beta_\alpha\}, \\
 &\quad \text{for } 2 \leq k \leq |I_=(\bar{x})| - q(\bar{x}).
 \end{aligned}$$

$I_{=}^0(\bar{x})$ contains all those index sets, where every active constraint is exactly once included, and there are no repeated constraints. This set is assumed to be nonempty since otherwise there are no active constraints at all and hence (almost) nothing to prove. Further, if we take an index set of $I_{=}^0(\bar{x})$ and drop exactly k elements, with $1 \leq k \leq q(\bar{x})$, we obtain an index set in $I_{=}^{-k}(\bar{x})$. On the other hand we get for each element $\alpha \in I_{=}^{-k}(\bar{x})$ a related element $\beta_\alpha \in I_{=}^0(\bar{x})$ with $\alpha \subseteq \beta_\alpha$. Therefore we have no repeated active constraints in any index set contained in $I_{=}^{-k}(\bar{x})$ for $1 \leq k \leq q(\bar{x})$ and the superscript indicates that k different constraints are missing to get an element of $I_{=}^0(\bar{x})$. In particular, the index sets for different nonpositive superscripts are disjoint and $I_{=}^{-q(\bar{x})}(\bar{x}) = \{\emptyset\}$.

The index sets with positive superscript are defined recursively. $I_{=}^1(\bar{x})$ contains all those index sets where exactly one active repeated constraint is included twice and all the others at most once. Therefore, by dropping one of the two repeated active constraints, we get an index set that does no longer contain repeated active constraints. Hence it is a subset of some index set $\beta_\alpha \in I_{=}^0(\bar{x})$, but it might be a strict subset, since it must not contain all different active constraints.

Any set $\alpha \in I_{=}^2(\bar{x})$ is not contained in $I_{=}^1(\bar{x})$ and by dropping one element of α we must obtain a set in $I_{=}^1(\bar{x})$. This means that $I_{=}^2(\bar{x})$ contains all those index sets, where exactly two constraints have to be dropped in order to get only different active constraints. Hence we either have three different indices $i_1, i_2, i_3 \in \alpha$ with $g_{i_1} \equiv g_{i_2} \equiv g_{i_3}$ or two pairs of different indices $i_1, i_2 \in \alpha$ and $j_1, j_2 \in \alpha$ with $g_{i_1} \equiv g_{i_2} \not\equiv g_{j_1} \equiv g_{j_2}$.

Further, the recursive definition implies that having any index set in $I_{=}^k(\bar{x})$, we must drop exactly k indices to obtain an index set from $I_{=}^0(\bar{x})$. The entire set $I_=(\bar{x})$ is contained in $I_{=}^{|I_=(\bar{x})|-q(\bar{x})}(\bar{x})$. Note that if we consider GNEPs where all constraints of all players are different, we have $|I_=(\bar{x})| = q(\bar{x})$ and we do not have the index sets with positive superscript.

Altogether we have a disjoint partition of the power set $\mathcal{P}(I_=(\bar{x}))$ of $I_=(\bar{x})$,

$$\mathcal{P}(I_=(\bar{x})) = \bigcup_{-q(\bar{x}) \leq k \leq m - q(\bar{x})} I_{=}^k(\bar{x}). \tag{5.36}$$

Let us illustrate the defined index sets in an example.

Example 5.25 Consider the 2-player game

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1 - 1)^2 \quad \text{s.t.} \quad x^1 \geq 0, x^1 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2 - 1)^2 \quad \text{s.t.} \quad x^2 \geq 0, x^1 + x^2 \leq 1. \end{aligned}$$

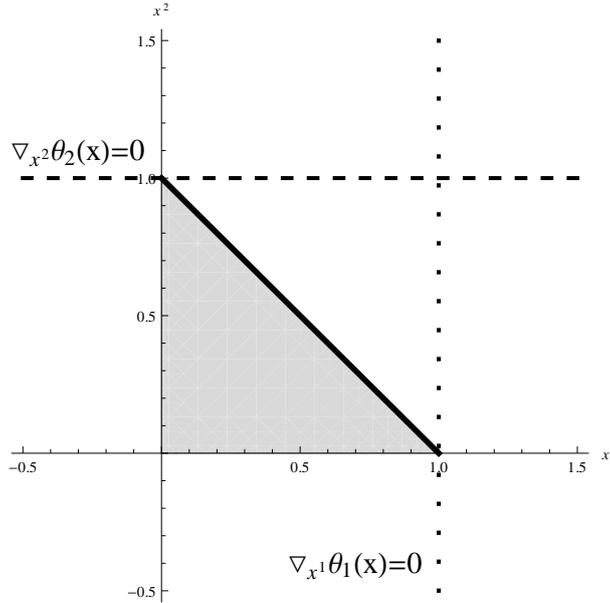


Figure 5.1.: Graphical solution of Example 5.25

Using Theorem 2.1 and Figure 5.1 we get the solution set

$$\{(a, 1 - a) \mid a \in [0, 1]\}.$$

Define the constraint function

$$g(x) := (-x^1, x^1 + x^2 - 1, -x^2, x^1 + x^2 - 1)^T.$$

Then we have $g_2 \equiv g_4$ and for the sets of active constraints at the solutions $I_=(\bar{x})$ and the index sets $I_=(\bar{x})^k$ we get:

$(0, 1)$	$(1, 0)$	$(a, 1 - a), a \in (0, 1)$
$I_=(0, 1) = \{1, 2, 4\}$	$I_=(1, 0) = \{2, 3, 4\}$	$I_=(a, 1 - a) = \{2, 4\}$
$I_=(0, 1)^0 = \{\{1, 2\}, \{1, 4\}\}$	$I_=(1, 0)^0 = \{\{2, 3\}, \{3, 4\}\}$	$I_=(a, 1 - a)^0 = \{\{2\}, \{4\}\}$
$I_=(0, 1)^1 = \{\{2, 4\}, \{1, 2, 4\}\}$	$I_=(1, 0)^1 = \{\{2, 4\}, \{2, 3, 4\}\}$	$I_=(a, 1 - a)^1 = \{\{2, 4\}\}$
$I_=(0, 1)^{-1} = \{\{1\}, \{2\}, \{4\}\}$	$I_=(1, 0)^{-1} = \{\{2\}, \{3\}, \{4\}\}$	$I_=(a, 1 - a)^{-1} = \{\emptyset\}$
$I_=(0, 1)^{-2} = \{\emptyset\}$	$I_=(1, 0)^{-2} = \{\emptyset\}$	

◇

5. Solving the KKT System of a GNEP

The above index sets will play a crucial role in the proofs later on. For the set of assumptions that are used to show boundedness of $\left\{\frac{d^k}{\lambda^k}\right\}$ we only need the set $I_{\leq}^0(\bar{x})$.

Assumption 5.26 (A1) *The sequence $\{z^k\}$ is bounded.*

(A2) *If there are identical constraints $g_i \equiv g_j$ with $i, j \in \{1, \dots, m\}, i \neq j$, then we have $w_i^0 = w_j^0$ for the starting vector w^0 .*

(A3) *At every iteration k , the parameter σ_k is sufficiently small to guarantee boundedness of $\frac{\sigma_k \mu_k}{w_j^k \lambda_j^k}$ for all $j = 1, \dots, m$.*

(A4) *The matrix $J_x F(x^k, \lambda^k)$ is nonsingular for all $(x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}_+^m$.*

(A5) *At every solution \bar{z} of $H(z) = 0, z \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$ there is a $\beta \in I_{\leq}^0(\bar{x})$, such that $M_{\beta\beta}(\bar{x}, \bar{\lambda})$ is nonsingular and $\bar{\lambda}_j > 0$ for all $j \in \beta$. Further $\det(M_{\alpha\alpha}(\bar{x}, \bar{\lambda}))$ is zero or has the same sign for all $\alpha \in I_{\leq}^0(\bar{x})$.*

Sufficient conditions for Assumption (A1), the boundedness of the sequence $\{z^k\}$, are given in Theorem 5.20. Assumption (A2), the choice of equal starting parameters for equal constraints $g_i \equiv g_j, i \neq j$ is very natural. A simple consequence of (A2) is stated in the next Lemma.

Lemma 5.27 *Let (A2) be satisfied. If there are indices $i, j \in \{1, \dots, m\}$ with $g_i \equiv g_j$, we have $w_i^k = w_j^k$ for all $k \in \mathbb{N}$. Furthermore*

$$\left(\prod_{i \in \bar{\alpha}_1 \setminus I_{<}(\bar{x})} w_i^k \right) = \left(\prod_{i \in \bar{\alpha}_2 \setminus I_{<}(\bar{x})} w_i^k \right)$$

holds for all $\alpha_1, \alpha_2 \in I_{\leq}^0(\bar{x})$ and all $k \in \mathbb{N}$, that is the product is independent of the choice of $\alpha \in I_{\leq}^0(\bar{x})$.

Proof. Let $\alpha_1, \alpha_2 \in I_{\leq}^0(\bar{x})$. By definition $\bar{\alpha}_1 \setminus I_{<}(\bar{x})$ and $\bar{\alpha}_2 \setminus I_{<}(\bar{x})$ contain only active constraints that are joined by several players, and they contain all but one of each joined active constraint. Thus (A2) implies

$$\left(\prod_{i \in \bar{\alpha}_1 \setminus I_{<}(\bar{x})} w_i^0 \right) = \left(\prod_{i \in \bar{\alpha}_2 \setminus I_{<}(\bar{x})} w_i^0 \right).$$

If we have indices $i, j \in \{1, \dots, m\}$ with $g_i \equiv g_j$ and $w_i^k = w_j^k$ for some $k \in \mathbb{N}_0$, (5.30) yields

$$\begin{aligned} d_{w,i}^k &= -w_i^k - g_i(x^k) - \nabla g_i(x^k)^T d_x^k + \sigma_k \mu_k \\ &= -w_j^k - g_j(x^k) - \nabla g_j(x^k)^T d_x^k + \sigma_k \mu_k = d_{w,j}^k, \end{aligned}$$

and hence

$$w_i^{k+1} = w_i^k + t_k d_{w,i}^k = w_j^k + t_k d_{w,j}^k = w_j^{k+1},$$

which implies the first assertion. Furthermore, by induction, we obtain

$$\left(\prod_{i \in \bar{\alpha}_1 \setminus I_{<}(\bar{x})} w_i^{k+1} \right) = \left(\prod_{i \in \bar{\alpha}_2 \setminus I_{<}(\bar{x})} w_i^{k+1} \right),$$

and this completes the proof. \square

Assumption (A3) is not restrictive but rather gives an idea, how to choose the parameter σ_k at every iteration. Note, that if $\min_j \{w_j^k \lambda_j^k\}$, the smallest of the last m components of $H(z^k)$, vanishes faster than the average $\mu^k = \frac{a^T H(z^k)}{\|a\|^2} = \frac{1}{2m} \sum_{i=1}^{2m} H_{n+i}(z^k)$ of the last $2m$ components of $H(z^k)$, (A3) implies $\sigma_k \rightarrow 0$.

The assumptions (A4) and (A5) restrict the number of GNEPs under consideration. Assumption (A4) also plays an important role for the nonsingularity of $JH(z^k)$. But note that nonsingularity is required for $\lambda \geq 0$ and not only for $\lambda > 0$. It will be needed to guarantee the existence of the limiting matrices $M(\bar{x}, \bar{\lambda})$. Further we can use (A4) together with (A1), (A3), and the continuity of g , $J_x g$, and F to ensure that r^k , defined in (5.34), is bounded. For the nonsingularity of $M_{\beta\beta}(\bar{x}, \bar{\lambda})$ in (A5) it is clear that at a solution \bar{x} all different active constraints have to be linearly independent, to get full row rank of $J_x g_\beta(\bar{x})$. Further the columns of $E(\bar{x})$ with indices in β must be linearly independent. If we do not have any joint constraints, this requires that LICQ holds for each player, i.e., $\{\nabla_{x^\nu} g_i^\nu(\bar{x}) \mid g_i^\nu(\bar{x}) = 0\}$ has to be linearly independent for each player $\nu = 1, \dots, N$. For joint active constraints the condition is more relaxed, since we then have $\beta \subset I_=(\bar{x})$, and thus for some players less gradients to be linearly independent. The second part of (A5), $\bar{\lambda}_j > 0$ for all $j \in \beta$, is strict complementarity, if there are no repeated active constraints, since then $\beta = I_=(\bar{x})$. Having repeated constraints this is a weaker assumption, since the Lagrange multiplier has to be positive only for one player and not for all players sharing a common constraint.

If for a given sequence (x^k, λ^k) that converges to a solution $(\bar{x}, \bar{\lambda})$, the matrices $M(x^k, \lambda^k) = J_x g(x^k) J_x F(x^k, \lambda^k)^{-1} E(x^k)$ are P_0 -matrices, which is the sufficient condition for nonsingularity of $JH(z^k)$ from Theorem 5.15, we have $\det(M_{\alpha\alpha}(x^k, \lambda^k)) \geq 0$ for all $\alpha \subseteq \{1, \dots, m\}$. Using continuity of $J_x g$ and E , and (A4), which guarantees continuity of $J_x F(x^k, \lambda^k)^{-1}$ for $(x^k, \lambda^k) \rightarrow (\bar{x}, \bar{\lambda})$, we get $\det(M_{\alpha\alpha}(\bar{x}, \bar{\lambda})) \geq 0$ for all $\alpha \subseteq \{1, \dots, m\}$. Therefore also the matrix $M(\bar{x}, \bar{\lambda})$ has the P_0 property. In this case the third part of (A5), that the determinants $\det(M_{\alpha\alpha}(\bar{x}, \bar{\lambda}))$ are zero or have the same sign for all $\alpha \in I_=(\bar{x})$, is satisfied.

Now we want to give an example showing that the assumptions (A1) to (A5) can be satisfied without $JH(\bar{z})$ being nonsingular at a solution \bar{z} , or M being a P -matrix.

Example 5.28 Consider the 2-player game, defined via

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1 - 1)^2 \quad \text{s.t.} \quad -x^1 \leq 1, x^1 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2 - 1)^2 \quad \text{s.t.} \quad -x^2 \leq 1, x^1 + x^2 \leq 1. \end{aligned}$$

Using Theorem 2.1 we can determine the solutions \bar{x} of the game, see Figure 5.2. By setting

$$g(x) := (-x^1 - 1, x^1 + x^2 - 1, -x^2 - 1, x^1 + x^2 - 1)^T,$$

we can compute the corresponding Lagrange multipliers $\bar{\lambda}$ and the slack variables \bar{w} . The solution set is

$$\{(\bar{x}, \bar{\lambda}, \bar{w}) \mid \bar{x} = (a, 1 - a), \bar{\lambda} = (0, 1 - a, 0, a), \bar{w} = (1 + a, 0, 2 - a, 0), a \in [0, 1]\}.$$

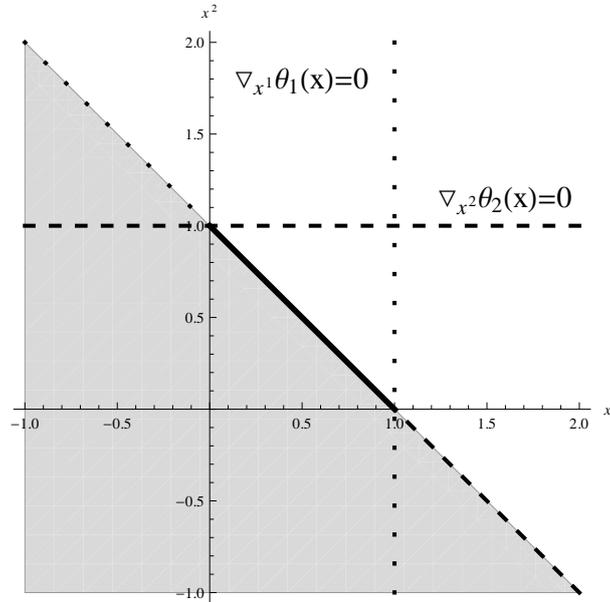


Figure 5.2.: Graphical solution of Example 5.28

We have $J_x F(x, \lambda) = I_2$, which is nonsingular for all $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m$. Hence (A4) is satisfied and further

$$M = M(x, \lambda) = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ 0 & -1 \\ 1 & 1 \end{pmatrix} I_2 \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}.$$

Now we can see that M is a P_0 -matrix independent of $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+^m$, but not a P -matrix. We have $I_{\bar{x}}^0 = \{\{2\}, \{4\}\}$ and $\det(M_{\{2\}, \{2\}}) = 1 = \det(M_{\{4\}, \{4\}})$. Further one of the

multipliers $\bar{\lambda}_2 = 1 - a$ or $\bar{\lambda}_4 = a$ is positive for all $a \in [0, 1]$ and therefore (A5) is satisfied. With M being a P_0 -matrix we get nonsingularity of $JH(z^k)$ for all $z^k \in \mathbb{R}^n \times \mathbb{R}_{++}^m \times \mathbb{R}_{++}^m$ by Theorem 5.15. Since all constraints are linear and the feasible set is bounded, we obtain boundedness of the sequence $\{z^k\}$ from Theorem 5.20. Thus also (A1) is satisfied. If we choose equal starting vectors $w_2^0 = w_4^0$ and σ_k sufficiently small at each iteration, also (A2) and (A3) and hence (A1) to (A5) is satisfied in this example. But, if we consider the solutions

$$\begin{aligned} \bar{x} &= (0, 1), & \bar{\lambda} &= (0, 1, 0, 0), & \bar{w} &= (1, 0, 2, 0), & \text{or} \\ \bar{x} &= (1, 0), & \bar{\lambda} &= (0, 0, 0, 1), & \bar{w} &= (2, 0, 1, 0), \end{aligned}$$

where strict complementarity does not hold, we get a singular matrix

$$JH(\bar{x}, \bar{\lambda}, \bar{w}) = \begin{pmatrix} 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \bar{w}_1 & 0 & 0 & 0 & \bar{\lambda}_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{w}_2 & 0 & 0 & 0 & \bar{\lambda}_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{w}_3 & 0 & 0 & 0 & \bar{\lambda}_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & \bar{w}_4 & 0 & 0 & 0 & \bar{\lambda}_4 \end{pmatrix},$$

since the 10th or 8th row contains only zeros. \diamond

Using the Cramer Rule, see (5.35), we will show that the sequence $\left\{\frac{d_\lambda^k}{\lambda^k}\right\}$ is bounded under the Assumptions (A1) to (A5). This will be done in three Lemmata. First we prove a lower bound for the denominator of the Cramer Rule formula, then we will show that any subsequence of $\left\{\frac{d_\lambda^k}{\lambda^k}\right\}$ converging to a solution stays bounded, and finally we prove the statement. For the denominator we get the following bound.

Lemma 5.29 *Suppose Assumptions (A2), (A4) and (A5) hold. Consider a subsequence $\{z^k\}_{k \in K}$ converging to a solution $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{w})$ of the constrained equation $H(z) = 0, z \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$. Let $\beta \in I_-^0(\bar{x})$ be defined by (A5). Then there is a constant $C > 0$, such that*

$$|\det(M^k \Lambda^k + W^k)| \geq C \left(\prod_{i \in \beta \setminus I_<(\bar{x})} w_i^k \right)$$

for all $z^k, k \in K$, sufficiently close to \bar{z} .

Proof. Let \bar{z} be a solution of the constrained equation and let $\beta \in I_-^0(\bar{x})$ satisfy Assumption (A5). Exploiting the diagonal structure of W^k we can use a known determinant

formula, see [13, p. 60], to get

$$\begin{aligned}
 \det(M^k \Lambda^k + W^k) &= \sum_{\alpha \subseteq \{1, \dots, m\}} \det(W_{\bar{\alpha}\bar{\alpha}}^k) \det((M^k \Lambda^k)_{\alpha\alpha}) \\
 &= \sum_{\alpha \subseteq \{1, \dots, m\}} \det(W_{\bar{\alpha}\bar{\alpha}}^k) \det(M_{\alpha\alpha}^k) \det(\Lambda_{\alpha\alpha}^k) \\
 &= \sum_{\alpha \subseteq \{1, \dots, m\}} \left(\prod_{i \in \bar{\alpha}} w_i^k \right) \left(\prod_{i \in \alpha} \lambda_i^k \right) \det(M_{\alpha\alpha}^k).
 \end{aligned}$$

By (5.36), for each $\alpha \subseteq \{1, \dots, m\}$ there is a $\ell \in \mathbb{Z}$ with $-q(\bar{x}) \leq \ell \leq |I_{=}(\bar{x})| - q(\bar{x})$ such that $\alpha \cap I_{=}(\bar{x}) \in I_{=}^{\ell}(\bar{x})$. If $\ell > 0$ we have two indices $i, j \in \alpha$ with $i \neq j$ and $g_i \equiv g_j$, resulting in two identical rows in $M_{\alpha\alpha}^k = J_x g_{\alpha}(x^k) J_x F(x^k, \lambda^k)^{-1} (E(x^k))_{\{1, \dots, m\}, \alpha}$. This implies $\det(M_{\alpha\alpha}^k) = 0$ and hence these summands vanish. If $\ell \leq 0$ the definition of $I_{=}^{\ell}(\bar{x})$ implies the existence of an index set $\beta_{\alpha} \in I_{=}^0(\bar{x})$ such that $\alpha \setminus I_{<}(\bar{x}) \subseteq \beta_{\alpha}$, and hence $\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x}) \subseteq \bar{\alpha} \setminus I_{<}(\bar{x})$ must hold. Then by splitting $\bar{\alpha}$ via

$$\bar{\alpha} = (\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})) \cup ((\bar{\alpha} \setminus I_{<}(\bar{x})) \setminus (\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x}))) \cup (\bar{\alpha} \cap I_{<}(\bar{x})),$$

and using that $\left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right) = \left(\prod_{i \in \bar{\beta} \setminus I_{<}(\bar{x})} w_i^k \right)$ by Lemma 5.27 and (A2) we get

$$\begin{aligned}
 &\sum_{\alpha \subseteq \{1, \dots, m\}} \left(\prod_{i \in \bar{\alpha}} w_i^k \right) \left(\prod_{i \in \alpha} \lambda_i^k \right) \det(M_{\alpha\alpha}^k) = \\
 &\left(\prod_{i \in \bar{\beta} \setminus I_{<}(\bar{x})} w_i^k \right) \sum_{\substack{\alpha \subseteq \{1, \dots, m\}, \\ \det(M_{\alpha\alpha}^k) \neq 0}} \left(\prod_{i \in (\bar{\alpha} \setminus I_{<}(\bar{x})) \setminus (\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x}))} w_i^k \right) \left(\prod_{i \in \bar{\alpha} \cap I_{<}(\bar{x})} w_i^k \right) \left(\prod_{i \in \alpha} \lambda_i^k \right) \det(M_{\alpha\alpha}^k).
 \end{aligned}$$

It is enough to show that the absolute value of the sum converges to some positive constant. Since $w_i^k \rightarrow_K 0$ for all $i \in (\bar{\alpha} \setminus I_{<}(\bar{x})) \setminus (\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x}))$, all summands where this set is nonempty converge to zero. Further $\lambda_i^k \rightarrow_K 0$ for all $i \in \alpha \cap I_{<}(\bar{x})$, which also implies a vanishing summand. If both cases do not occur, we must have $\bar{\alpha} \setminus I_{<}(\bar{x}) = \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})$ and $\alpha \cap I_{<}(\bar{x}) = \emptyset$, which implies $\alpha = \beta_{\alpha} \in I_{=}^0(\bar{x})$. Thus we only have to consider the sum

$$\sum_{\alpha \in I_{=}^0(\bar{x})} \left(\prod_{i \in I_{<}(\bar{x})} w_i^k \right) \left(\prod_{i \in \alpha} \lambda_i^k \right) \det(M_{\alpha\alpha}(x^k, \lambda^k)).$$

By Assumption (A5), $\det(M_{\alpha\alpha}(\bar{x}, \bar{\lambda}))$ is zero or has the same sign for all $\alpha \in I_{=}^0(\bar{x})$. This together with $(\lambda^k, w^k) \in \mathbb{R}_{++}^{2m}$ shows that all terms in the last sum are either nonnegative or nonpositive in the limit, so they have the same sign or converge to zero, whereas at least

one term ($\alpha = \beta$) is nonzero since by (A5) $\det(M_{\beta\beta}(\bar{x}, \bar{\lambda})) \neq 0$ and $\bar{\lambda}_j > 0$ for all $j \in \beta$. Thus we can find a constant $C > 0$ such that

$$\left| \sum_{\alpha \in I_{\leq}^0(\bar{x})} \left(\prod_{i \in I_{<}(\bar{x})} w_i^k \right) \left(\prod_{i \in \alpha} \lambda_i^k \right) \det(M_{\alpha\alpha}^k) \right| \rightarrow_K 2C.$$

This yields the statement,

$$|\det(M^k \Lambda^k + W^k)| \geq C \left(\prod_{i \in \beta \setminus I_{<}(\bar{x})} w_i^k \right)$$

for all $z^k, k \in K$, sufficiently close to \bar{z} . \square

Looking at the bound derived in this lemma, we see that it contains $\left(\prod_{i \in \beta \setminus I_{<}(\bar{x})} w_i^k \right)$, which converges to zero, if there are repeated active constraints. Therefore it is not enough to show that the numerator in the Cramer Rule is bounded by some constant, which can be obtained from the boundedness of z^k and the continuity of all involved functions, but that it stays bounded if divided by $\left(\prod_{i \in \beta \setminus I_{<}(\bar{x})} w_i^k \right)$. The next Lemma shows this and therefore boundedness of $\left\{ \frac{d_{\lambda}^k}{\lambda^k} \right\}$ on a subsequence converging to a solution.

Lemma 5.30 *Let Assumption 5.26 hold. Consider a subsequence $\{z^k\}_{k \in K}$ converging to a solution $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{w})$ of the constrained equation $H(z) = 0, z \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$. Then $\left\{ \frac{d_{\lambda}^k}{\lambda^k} \right\}_{k \in K}$ is bounded.*

Proof. Let $k \in K$. For an arbitrary $j \in \{1, \dots, m\}$ we will show boundedness of $\left\{ \frac{d_{\lambda,j}^k}{\lambda_j^k} \right\}$ using the representation

$$\frac{d_{\lambda,j}^k}{\lambda_j^k} = \frac{\det(\tilde{M}_j^k \tilde{\Lambda}_j^k + \tilde{W}_j^k)}{\det(M^k \Lambda^k + W^k)} = \sum_{\alpha \subseteq \{1, \dots, m\}} \frac{\det((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}}) \det((\tilde{M}_j^k)_{\alpha\alpha}) \det((\tilde{\Lambda}_j^k)_{\alpha\alpha})}{\det(M^k \Lambda^k + W^k)}, \quad (5.37)$$

obtained by Cramer's Rule, see (5.35), and the determinantal formula from [13]. From Lemma 5.29 we have a $\beta \in I_{\leq}^0(\bar{x})$ satisfying Assumption (A5), and a constant $C > 0$, such that the denominator satisfies

$$|\det(M^k \Lambda^k + W^k)| \geq C \left(\prod_{i \in \beta \setminus I_{<}(\bar{x})} w_i^k \right) \quad (5.38)$$

for all z^k sufficiently close to \bar{z} . All summands $\alpha \subseteq \{1, \dots, m\}$ in (5.37) with $j \in \bar{\alpha}$ are zero by construction of \tilde{W}_j^k . Thus we only consider α with $j \in \alpha$. By (5.36) there is a $\ell \in \mathbb{Z}$ with $-q(\bar{x}) \leq \ell \leq |I_{-}(\bar{x})| - q(\bar{x})$ such that $\alpha \cap I_{-}(\bar{x}) \in I_{\leq}^{\ell}(\bar{x})$. Now one of the following three cases must occur for each α :

- $j \in \alpha$ and $\ell \leq 0$: Then the definition of $I_{\leq}^{\ell}(\bar{x})$ implies the existence of an index set $\beta_{\alpha} \in I_{\leq}^0(\bar{x})$ such that $\alpha \setminus I_{<}(\bar{x}) \subseteq \beta_{\alpha}$, and hence $\bar{\beta}_{\alpha} \setminus I_{<}(\bar{x}) \subseteq \bar{\alpha} \setminus I_{<}(\bar{x})$. Therefore we get

$$\begin{aligned}
 & \left| \det \left((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}} \right) \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \det \left((\tilde{\Lambda}_j^k)_{\alpha\alpha} \right) \right| \\
 &= \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right) \left(\prod_{i \in \bar{\alpha} \setminus (\beta_{\alpha} \setminus I_{<}(\bar{x}))} w_i^k \right) \left(\prod_{\substack{i \in \alpha \\ i \neq j}} \lambda_i^k \right) \left| \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \right| \\
 &\leq \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right) B_{\alpha},
 \end{aligned}$$

where the existence of the constant $B_{\alpha} > 0$ follows from (A1), which guarantees boundedness of all iterates, together with (A3) and (A4) which imply boundedness of r^k and thus also of $\det \left((\tilde{M}_j^k)_{\alpha\alpha} \right)$. This together with (5.38) and Lemma 5.27 shows

$$\begin{aligned}
 \frac{\left| \det \left((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}} \right) \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \det \left((\tilde{\Lambda}_j^k)_{\alpha\alpha} \right) \right|}{\left| \det (M^k \Lambda^k + W^k) \right|} &\leq \frac{B_{\alpha} \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right)}{C \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right)} \\
 &= \frac{B_{\alpha}}{C},
 \end{aligned}$$

which means boundedness of these summands.

- $j \in \alpha$ and $\ell = 1$: Per definition we have an index set $\beta_{\alpha} \in I_{\leq}^0(\bar{x})$ and two indices $l_1, l_2 \in \alpha \setminus I_{<}(\bar{x})$ such that $l_1 \neq l_2$, $g_{l_1} \equiv g_{l_2}$, and $\alpha \setminus I_{<}(\bar{x}) \subseteq \beta_{\alpha} \cup \{l_1\}$, which implies $\bar{\beta}_{\alpha} \setminus (I_{<}(\bar{x}) \cup \{l_1\}) \subseteq \bar{\alpha} \setminus I_{<}(\bar{x})$. We can change the rows of $(\tilde{M}_j^k)_{\alpha\alpha}$ such that the row corresponding to l_1 is directly under the one corresponding to l_2 , and this does not change the absolute value $\left| \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \right|$. Then developing the determinant by the j -th column, and using that the l_1 -th and l_2 -th row of M^k are identical, we get

$$\begin{aligned}
 \left| \det (\tilde{M}_j^k)_{\alpha\alpha} \right| &= \left| \det (M_1^k \dots M_{j-1}^k r^k M_{j+1}^k \dots M_m^k)_{\alpha\alpha} \right| \\
 &= \left| r_{l_1}^k \det (M_{\alpha \setminus \{l_1, \alpha \setminus \{j\}}^k} - r_{l_2}^k \det (M_{\alpha \setminus \{l_2, \alpha \setminus \{j\}}^k) \right| \\
 &= \left| r_{l_1}^k - r_{l_2}^k \right| \left| \det (M_{\alpha \setminus \{l_1, \alpha \setminus \{j\}}^k) \right|.
 \end{aligned}$$

(A1) and (A4) imply that $\left| \det \left(M_{\alpha \setminus \{l_1, \alpha \setminus \{j\}}^k \right) \right|$ is bounded. Thus using once again

(A1) and $\bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\}) \subseteq \bar{\alpha} \setminus I_<(\bar{x})$, there exists a constant $B_\alpha > 0$ such that

$$\begin{aligned} & |\det \left((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}} \right) \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \det \left((\tilde{\Lambda}_j^k)_{\alpha\alpha} \right)| \\ &= \left(\prod_{i \in \bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\})} w_i^k \right) \left(\prod_{i \in \bar{\alpha} \setminus (\bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\}))} w_i^k \right) \left(\prod_{\substack{i \in \alpha \\ i \neq j}} \lambda_i^k \right) |\det \left((\tilde{M}_j^k)_{\alpha\alpha} \right)| \\ &\leq \left(\prod_{i \in \bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\})} w_i^k \right) |r_{l_1}^k - r_{l_2}^k| B_\alpha. \end{aligned}$$

By Lemma 5.27 we get

$$\left(\prod_{i \in \bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\})} w_i^k \right) = \left(\prod_{i \in \bar{\beta} \setminus (I_<(\bar{x}) \cup \{l_1\})} w_i^k \right),$$

since $\beta_\alpha, \beta \in I_\equiv(\bar{x})$. This together with (5.38) implies

$$\begin{aligned} \frac{|\det \left((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}} \right) \det \left((\tilde{M}_j^k)_{\alpha\alpha} \right) \det \left((\tilde{\Lambda}_j^k)_{\alpha\alpha} \right)|}{|\det(M^k \Lambda^k + W^k)|} &\leq \frac{|r_{l_1}^k - r_{l_2}^k| \left(\prod_{i \in \bar{\beta}_\alpha \setminus (I_<(\bar{x}) \cup \{l_1\})} w_i^k \right) B_\alpha}{C \left(\prod_{i \in \bar{\beta} \setminus I_<(\bar{x})} w_i^k \right)} \\ &= \frac{|r_{l_1}^k - r_{l_2}^k| B_\alpha}{w_{l_1}^k C}. \end{aligned}$$

Since $g_{l_1} \equiv g_{l_2}$ and, again by Lemma 5.27, $w_{l_1}^k = w_{l_2}^k$ holds, we have

$$\frac{|r_{l_1}^k - r_{l_2}^k|}{w_{l_1}^k} = \frac{\sigma_k \mu_k \left| \frac{1}{\lambda_{l_1}^k} - \frac{1}{\lambda_{l_2}^k} \right|}{w_{l_1}^k} \leq \frac{\sigma_k \mu_k}{\min \{w_{l_1}^k \lambda_{l_1}^k, w_{l_2}^k \lambda_{l_2}^k\}},$$

which is bounded by Assumption (A3). Therefore the summands of (5.37) are bounded in this case.

- $j \in \alpha$ and $\ell \geq 2$: Then there are two possibilities. Either we have three different indices $l_1, l_2, l_3 \in \alpha \setminus I_<(\bar{x})$ with $g_{l_1} \equiv g_{l_2} \equiv g_{l_3}$. Then the corresponding rows of $(\tilde{M}_j^k)_{\alpha\alpha}$ can only differ in one column (the one corresponding to j). Hence the rows are linearly dependent and $\det(\tilde{M}_j^k)_{\alpha\alpha} = 0$ which implies that the summand in (5.37) vanishes. Or there are four indices $j_1, j_2, l_1, l_2 \in \alpha \setminus I_<(\bar{x})$ with $j_1 \neq j_2, l_1 \neq l_2$ and $g_{j_1} \equiv g_{j_2} \neq g_{l_1} \equiv g_{l_2}$. Then the corresponding four rows of $(\tilde{M}_j^k)_{\alpha\alpha}$ are linearly dependent and again the summand vanishes in (5.37).

5. Solving the KKT System of a GNEP

Altogether we have shown that all summands in (5.37) are bounded, hence $\left\{ \frac{d_{\lambda_j}^k}{\lambda_j^k} \right\}_{k \in K}$ is bounded for all $j \in \{1, \dots, m\}$. \square

Now we are able to show the desired boundedness result not only on a subsequence but for the entire sequence.

Lemma 5.31 *Let the assumptions of Theorem 5.12 and Assumption 5.26 hold. Then the entire sequence $\left\{ \frac{d_{\lambda}^k}{\lambda^k} \right\}_{k \in \mathbb{N}}$ is bounded.*

Proof. Assume $\left\{ \frac{d_{\lambda}^k}{\lambda^k} \right\}_{k \in \mathbb{N}}$ is unbounded. Then there exists a $K \subseteq \mathbb{N}$, such that the subsequence $\left\{ \left\| \frac{d_{\lambda}^k}{\lambda^k} \right\| \right\}_{k \in K}$ is monotonically increasing and going to infinity. By (A1), the sequence $\{z^k\}_{k \in K}$ is bounded and hence has an accumulation point \bar{z} , which is by Theorem 5.12 a solution of $H(z) = 0, z \in \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$. Therefore we can find a subsequence $K_1 \subseteq K$ with $z^k \rightarrow_{K_1} \bar{z}$ and, using the monotonicity, $\left\| \frac{d_{\lambda}^k}{\lambda^k} \right\| \rightarrow_{K_1} \infty$. This contradicts Lemma 5.30 and shows that the entire sequence $\left\{ \frac{d_{\lambda}^k}{\lambda^k} \right\}_{k \in \mathbb{N}}$ remains bounded. \square

By this Lemma 5.31 we have boundedness of $\frac{d_{\lambda_i}^k}{\lambda_i^k}$ not only from above. Thus we get the existence of a constant $c_{\lambda} > 0$ such that $\left| \frac{d_{\lambda_i}^k}{\lambda_i^k} \right| \leq c_{\lambda}$ for all $i = 1, \dots, m$. Now using equation (5.31) we have

$$\frac{d_{w,i}^k}{w_i^k} = -1 - \frac{d_{\lambda,i}^k}{\lambda_i^k} + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k}. \quad (5.39)$$

By the boundedness of $\frac{\sigma_k \mu_k}{\lambda_i^k w_i^k}$ from (A3), we further get a constant $c_w > 0$ such that $\frac{d_{w,i}^k}{w_i^k} < c_w$ for all $i = 1, \dots, m$. Therefore by Theorem 5.24 and Lemma 5.31 we have shown the following result.

Corollary 5.32 *Let the assumptions of Theorem 5.12 and Assumptions 5.26 hold, and assume that all constraint functions $g_i, i = 1, \dots, m$ are convex. Then there exists a $\tilde{t} > 0$ such that $z^k + t_k d^k \in Z_I$ for all $t_k \in [0, \tilde{t}]$.*

The previous theorem is only the first part for a lower bound result for the stepsize t_k . The remaining part is to show that the Armijo linesearch rule

$$\Psi(z^k + t_k d^k) \leq \Psi(z^k) + \gamma t_k \nabla \Psi(z^k)^T d^k$$

is satisfied for $t_k \leq \bar{t}$ with a $\bar{t} \leq \tilde{t}$. Again, boundedness of some sequences, in particular of $\left\{ \frac{d_{\lambda}^k}{\|H(z^k)\|} \right\}$ is used to prove this. Therefore consider once again equation (5.33). Multiplying

this equation by $\frac{1}{\|H(z^k)\|}\Lambda^k$ and using the diagonal structure of the matrices W^k and Λ^k we obtain

$$(\Lambda^k M^k + W^k) \frac{d_\lambda^k}{\|H(z^k)\|} = \frac{\Lambda^k r^k}{\|H(z^k)\|}. \quad (5.40)$$

Similarly to the previous part we want to use the Cramer Rule to get boundedness of $\left\{ \frac{d_\lambda^k}{\|H(z^k)\|} \right\}$, that is we consider

$$\frac{d_{\lambda,i}^k}{\|H(z^k)\|} = \frac{\det(\Lambda^k \hat{M}_i^k + \tilde{W}_i^k)}{\det(\Lambda^k M^k + W^k)}, \quad (5.41)$$

with

$$\hat{M}_i^k := \left(M_1^k \dots M_{i-1}^k \frac{r^k}{\|H(z^k)\|} M_{i+1}^k \dots M_m^k \right).$$

Note that we can use the full matrix Λ^k here also in the numerator and the matrix \tilde{W}_i^k is the same as before in (5.35).

For our boundedness result we will need a stronger assumption on the parameter σ_k .

Assumption 5.33 (A3') *At every iteration k , the parameter σ_k is sufficiently small to guarantee boundedness of $\frac{\sigma_k \mu_k}{\|H(z^k)\| H_{n+j}(z^k)}$ for all $j = 1, \dots, 2m$.*

In contrast to (A3) the new assumption includes all elements $n+1$ to $n+2m$ from H and not only the last m elements. It is easy to see that (A3') implies (A3), but in fact it is much stronger, since we have the additional term $\|H(z^k)\|$ in the denominator, which is aimed to become zero. For any sequence $\{z^k\}$ with $\|H(z^k)\| \rightarrow 0$, (A3') implies

$$\frac{\sigma_k \mu_k}{\|H(z^k)\|} \rightarrow 0 \quad \text{and} \quad \frac{\sigma_k \mu_k}{H_{n+j}(z^k)} \rightarrow 0 \quad \text{for all } j = 1, \dots, 2m. \quad (5.42)$$

In particular we now have $\sigma_k \rightarrow 0$. Using this stronger assumption (A3') instead of (A3) we can obtain the following.

Lemma 5.34 *Let Assumption 5.26 with (A3') instead of (A3) hold. Then the sequences $\left\{ \frac{d_\lambda^k}{\|H(z^k)\|} \right\}$ and $\left\{ \frac{d_x^k}{\|H(z^k)\|} \right\}$ are bounded.*

Proof. This proof is very similar to the ones of Lemma 5.30 and 5.31. Let $k \in K$. For an arbitrary $j \in \{1, \dots, m\}$ we will show boundedness of $\left\{ \frac{d_{\lambda,j}^k}{\|H(z^k)\|} \right\}$ using the representation

$$\frac{d_{\lambda,j}^k}{\|H(z^k)\|} = \frac{\det(\Lambda^k \hat{M}_j^k + \tilde{W}_j^k)}{\det(\Lambda^k M^k + W^k)} = \sum_{\alpha \subseteq \{1, \dots, m\}} \frac{\det((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}}) \det(\Lambda_{\alpha\alpha}^k) \det((\hat{M}_j^k)_{\alpha\alpha})}{\det(\Lambda^k M^k + W^k)}, \quad (5.43)$$

obtained by Cramer's Rule, see (5.41). Since the determinants of the matrices $(\Lambda^k M^k + W^k)$ and $(M^k \Lambda^k + W^k)$ coincide, because

$$\det(M^k \Lambda^k + W^k) = \sum_{\alpha \subseteq \{1, \dots, m\}} \det(W_{\bar{\alpha}\bar{\alpha}}^k) \det(M_{\alpha\alpha}^k) \det(\Lambda_{\alpha\alpha}^k) = \det(\Lambda^k M^k + W^k),$$

Lemma 5.29 yields the existence of a $\beta \in I_{\bar{=}}^0(\bar{x})$ and a constant $C > 0$, such that

$$|\det(\Lambda^k M^k + W^k)| \geq C \left(\prod_{i \in \bar{\beta} \setminus I_{<}(\bar{x})} w_i^k \right)$$

for all z^k sufficiently close to a solution $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{w})$ of the constrained equation $\text{CE}(H, Z)$. All summands in (5.43) with $j \in \bar{\alpha}$ are zero by construction of \bar{W}_j^k . Thus consider only summands with $j \in \alpha$. By (5.36), for each $\alpha \subseteq \{1, \dots, m\}$ there is a $\ell \in \mathbb{Z}$ with $-q(\bar{x}) \leq \ell \leq |I_{=}(\bar{x})| - q(\bar{x})$ such that $\alpha \cap I_{=}(\bar{x}) = I_{=}^{\ell}(\bar{x})$. Now one of the following three cases must occur for each α :

- $j \in \alpha$ and $\ell \leq 0$: In order to get boundedness of

$$\left| \det\left((\hat{M}_j^k)_{\alpha\alpha}\right) \det\left((\Lambda^k)_{\alpha\alpha}\right) \right| = \left| \det\left((\Lambda^k \hat{M}_j^k)_{\alpha\alpha}\right) \right|,$$

we can use (A1), (A4) and further we only have to assure boundedness of

$$\begin{aligned} \frac{\lambda_j^k r_j^k}{\|H(z^k)\|} &= -\lambda_j^k \nabla g_j(x^k)^T J_x F(x^k, \lambda^k)^{-1} \frac{F(x^k, \lambda^k)}{\|H(z^k)\|} \\ &\quad + \frac{\lambda_j^k (g_j(x^k) + w_j^k) - \lambda_j^k w_j^k}{\|H(z^k)\|} + \frac{\sigma_k \mu_k (1 - \lambda_j^k)}{\|H(z^k)\|}. \end{aligned}$$

But this is bounded by (A1), (A3') and (A4). Now we get analogous to Lemma 5.30 an index set $\beta_{\alpha} \in I_{\bar{=}}^0(\bar{x})$ and a constant $B_{\alpha} > 0$ such that

$$\begin{aligned} \frac{\left| \det\left((\bar{W}_j^k)_{\bar{\alpha}\bar{\alpha}}\right) \det\left((\hat{M}_j^k)_{\alpha\alpha}\right) \det\left(\Lambda_{\alpha\alpha}^k\right) \right|}{\left| \det(\Lambda^k M^k + W^k) \right|} &\leq \frac{B_{\alpha} \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right)}{C \left(\prod_{i \in \bar{\beta}_{\alpha} \setminus I_{<}(\bar{x})} w_i^k \right)} \\ &= \frac{B_{\alpha}}{C}, \end{aligned}$$

which means boundedness of these summands.

- $j \in \alpha$ and $\ell = 1$: We can follow the corresponding part of the proof of Lemma 5.30. We have two indices $l_1, l_2 \in \alpha \setminus I_{<}(\bar{x})$ such that $l_1 \neq l_2$, $g_{l_1} \equiv g_{l_2}$ and we get

$$\left| \det(\hat{M}_j^k)_{\alpha\alpha} \right| = \frac{|r_{l_1}^k - r_{l_2}^k|}{\|H(z^k)\|} \left| \det(M_{\alpha \setminus \{l_1\}, \alpha \setminus \{j\}}^k) \right|.$$

Further there is some bound $B_\alpha > 0$ such that

$$\frac{|\det((\tilde{W}_j^k)_{\bar{\alpha}\bar{\alpha}}) \det((\hat{M}_j^k)_{\alpha\alpha}) \det(\Lambda^k)_{\alpha\alpha}|}{|\det(\Lambda^k M^k + W^k)|} \leq \frac{|r_{l_1}^k - r_{l_2}^k| B_\alpha}{\|H(z^k)\| w_{l_1}^k C}.$$

Finally we use

$$\frac{|r_{l_1}^k - r_{l_2}^k|}{\|H(z^k)\| w_{l_1}^k} = \frac{\sigma_k \mu_k \left| \frac{1}{\lambda_{l_1}^k} - \frac{1}{\lambda_{l_2}^k} \right|}{\|H(z^k)\| w_{l_1}^k} \leq \frac{\sigma_k \mu_k}{\|H(z^k)\| \min\{w_{l_1}^k \lambda_{l_1}^k, w_{l_2}^k \lambda_{l_2}^k\}},$$

which is bounded by Assumption (A3'), and obtain boundedness of the summands of (5.43) in this case.

- $j \in \alpha$ and $\ell \geq 2$: In this case we get in complete analogy to the proof of Lemma 5.30 that all summands vanish, since $\det(\hat{M}_j^k)_{\alpha\alpha} = 0$.

Altogether we have shown that all summands in (5.43) are bounded. Hence $\left\{ \frac{d_{\lambda,j}^k}{\|H(z^k)\|} \right\}$ is bounded for all $j \in \{1, \dots, m\}$ first on a subsequence and then using the same arguments as in the proof of Lemma 5.31 on the entire sequence.

Finally using (5.29) and the nonsingularity assumption from (A4) we obtain

$$\frac{d_x^k}{\|H(z^k)\|} = -J_x F(x^k, \lambda^k)^{-1} \left(E(x^k) \frac{d_\lambda^k}{\|H(z^k)\|} + \frac{F(x^k, \lambda^k)}{\|H(z^k)\|} \right),$$

which implies boundedness of the sequence $\left\{ \frac{d_x^k}{\|H(z^k)\|} \right\}$. \square

Before exploiting the boundedness properties from the previous Lemma for our main result, we want to prove a technical lemma giving a relation between $H(z^k + t_k d^k)$ and $H(z^k)$.

Lemma 5.35 *Suppose Assumption 5.26 with (A3') instead of (A3) holds. Let $\{z^k\}_{k \in K} = \{(x^k, \lambda^k, w^k)\}_{k \in K} \subseteq \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}_+^m$ be a sequence converging to a solution and $t_k \in [0, 1]$.*

Define $q(z^k, t_k) := \begin{pmatrix} q_1(z^k, t_k) \\ q_2(z^k, t_k) \\ q_3(z^k, t_k) \end{pmatrix} \in \mathbb{R}^{n+m+m}$ by

$$\begin{aligned} q_1(z^k, t_k) &:= \left(\int_0^1 (J_x F(x^k + st_k d_x^k, \lambda^k + st_k d_\lambda^k) - J_x F(x^k, \lambda^k)) ds \right) d_x^k \\ &\quad + \left(\int_0^1 (E(x^k + st_k d_x^k) - E(x^k)) ds \right) d_\lambda^k, \\ q_2(z^k, t_k) &:= \left(\int_0^1 (J_x g(x^k + st_k d_x^k) - J_x g(x^k)) ds \right) d_x^k + \sigma_k \mu_k 1_m, \\ q_3(z^k, t_k) &:= \sigma_k \mu_k 1_m + t_k (d_w^k \circ d_\lambda^k). \end{aligned}$$

Then we have

$$H(z^k + t_k d^k) = (1 - t_k)H(z^k) + t_k q(z^k, t_k),$$

and further if $t_k \rightarrow_K 0$ we get

$$\lim_{k \in K} \frac{\|q(z^k, t_k)\|}{\|H(z^k)\|} = 0.$$

Proof. Using the mean value theorem we obtain for the first n -dimensional part of H

$$\begin{aligned} F(x^k + t_k d_x^k, \lambda^k + t_k d_\lambda^k) &= F(x^k, \lambda^k) + t_k \left(J_x F(x^k, \lambda^k) d_x^k + E(x^k) d_\lambda^k \right) \\ &\quad + t_k \left(\int_0^1 \left(J_x F(x^k + s t_k d_x^k, \lambda^k + s t_k d_\lambda^k) - J_x F(x^k, \lambda^k) \right) ds \right) d_x^k \\ &\quad + t_k \left(\int_0^1 \left(E(x^k + s t_k d_x^k) - E(x^k) \right) ds \right) d_\lambda^k \\ &\stackrel{(5.29)}{=} (1 - t_k) F(x^k, \lambda^k) \\ &\quad + t_k \left(\int_0^1 \left(J_x F(x^k + s t_k d_x^k, \lambda^k + s t_k d_\lambda^k) - J_x F(x^k, \lambda^k) \right) ds \right) d_x^k \\ &\quad + t_k \left(\int_0^1 \left(E(x^k + s t_k d_x^k) - E(x^k) \right) ds \right) d_\lambda^k \\ &= (1 - t_k) F(x^k, \lambda^k) + t_k q_1(z^k, t_k), \end{aligned}$$

for the second m -dimensional part

$$\begin{aligned} g(x^k + t_k d_x^k) + w^k + t_k d_w^k &= g(x^k) + w^k + t_k \left(J_x g(x^k) d_x^k + d_w^k \right) \\ &\quad + t_k \left(\int_0^1 \left(J_x g(x^k + s t_k d_x^k) - J_x g(x^k) \right) ds \right) d_x^k \\ &\stackrel{(5.30)}{=} (1 - t_k) (g(x^k) + w^k) + t_k \sigma_k \mu_k 1_m \\ &\quad + t_k \left(\int_0^1 \left(J_x g(x^k + s t_k d_x^k) - J_x g(x^k) \right) ds \right) d_x^k \\ &= (1 - t_k) (g(x^k) + w^k) + t_k q_2(z^k, t_k), \end{aligned}$$

and for the third m -dimensional part

$$\begin{aligned} (\lambda^k + t_k d_\lambda^k) \circ (w^k + t_k d_w^k) &= \lambda^k \circ w^k + t_k (\lambda^k \circ d_w^k + w^k \circ d_\lambda^k) + t_k^2 (d_w^k \circ d_\lambda^k) \\ &\stackrel{(5.31)}{=} (1 - t_k) \lambda^k \circ w^k + t_k \sigma_k \mu_k 1_m + t_k^2 (d_w^k \circ d_\lambda^k) \\ &= (1 - t_k) \lambda^k \circ w^k + t_k q_3(z^k, t_k). \end{aligned}$$

This proves

$$H(z^k + t_k d^k) = (1 - t_k)H(z^k) + t_k q(z^k, t_k).$$

Now assume $t_k \rightarrow_K 0$. Since z^k converges to a solution, $\frac{\sigma_k \mu_k}{\|H(z^k)\|} \rightarrow_K 0$ is a consequence of (A3'), see (5.42). Further from Lemma 5.31 and equation (5.39) we obtain that $\frac{|d_{\lambda,i}^k d_{w,i}^k|}{\|H(z^k)\|} \leq \frac{|d_{\lambda,i}^k d_{w,i}^k|}{\lambda_i^k w_i^k}$ is bounded, and thus $t_k \rightarrow_K 0$ yields $t_k \frac{|d_{\lambda,i}^k d_{w,i}^k|}{\|H(z^k)\|} \rightarrow_K 0$. From Lemma 5.34 we know that the sequences $\left\{ \frac{d_\lambda^k}{\|H(z^k)\|} \right\}$ and $\left\{ \frac{d_x^k}{\|H(z^k)\|} \right\}$ are bounded. Therefore $\lim_{k \in K} \frac{\|q(z^k, t_k)\|}{\|H(z^k)\|} = 0$ follows if we can show that all the integrals in the definition of $q(z^k, t_k)$ converge to zero. With $\left\{ \frac{d_\lambda^k}{\|H(z^k)\|} \right\}$ and $\left\{ \frac{d_x^k}{\|H(z^k)\|} \right\}$ also d_λ^k and d_x^k are bounded. Thus using the boundedness assumption (A1) we can find a compact set C such that $(x^k + st_k d_x^k, \lambda^k + st_k d_\lambda^k) \in C$ for all $s \in [0, 1]$ and all $k \in K$. Since $J_x F$ is continuous on \mathbb{R}^n , it is uniformly continuous on C . Hence $t_k \rightarrow 0$ implies

$$\begin{aligned} \lim_{k \in K} \left\| \int_0^1 (J_x F(x^k + st_k d_x^k, \lambda^k + st_k d_\lambda^k) - J_x F(x^k, \lambda^k)) ds \right\| \\ \leq \lim_{k \in K} \max_{s \in [0,1]} \|J_x F(x^k + st_k d_x^k, \lambda^k + st_k d_\lambda^k) - J_x F(x^k, \lambda^k)\| = 0. \end{aligned}$$

Using the continuity of E and $J_x g$ the same argument shows that all the integrals in the definition of $q(z^k, t_k)$ converge to zero, which completes the proof. \square

Finally we are in the position to prove the boundedness of the stepsize from below.

Theorem 5.36 *Let the assumptions of Theorem 5.12 and Assumptions 5.26 with (A3') instead of (A3) hold. Further assume that all constraint functions $g_i, i = 1, \dots, m$ are convex. Then the sequence of the stepsizes $\{t_k\}$ generated by Algorithm 5.9 with $\eta_k = 0$ for all $k \in \mathbb{N}$ is bounded away from zero, that is, there exists a $\bar{t} > 0$ such that $t_k \geq \bar{t}$ for all $k \in \mathbb{N}$.*

Proof. By Corollary 5.32, there is a $\tilde{t} \in (0, 1]$ such that $z^k + t_k d^k \in Z_I$ for all $t_k \in [0, \tilde{t}]$. Since d^k is a direction of descent of Ψ by Lemma 5.11, we can always find a $t_k = \beta^{l_k} = \max\{\beta^l \mid \beta^l \leq \tilde{t}, l = 0, 1, 2, \dots\}$ such that the Armijo linesearch rule

$$\Psi(z^k + t_k d^k) \leq \Psi(z^k) + \gamma t_k \nabla \Psi(z^k)^T d^k$$

is satisfied. We have to show that this sequence $\{t_k\}$ is bounded away from zero, because this immediately implies that the stepsize computed by Algorithm 5.9 with $\eta_k = 0$ for all $k \in \mathbb{N}$ is bounded away from zero. Assume by contradiction that there is a subsequence $K \subseteq \mathbb{N}$ with $t_k = \beta^{l_k} \rightarrow_K 0$ and then also $\beta^{l_k-1} \rightarrow_K 0$. Subsequencing, if necessary, we may assume that $\{z^k\}$ converges to a solution, i.e., $H(z^k) \rightarrow_K 0$. For all $k \in K$ sufficiently large we obtain $\beta^{l_k-1} \leq \tilde{t}$, hence $z^k + \beta^{l_k-1} d^k \in Z_I$, and from the linesearch rule

$$\frac{\Psi(z^k + \beta^{l_k-1} d^k) - \Psi(z^k)}{\beta^{l_k-1}} > \gamma \nabla \Psi(z^k)^T d^k.$$

On the one hand, using the definitions and Lemma 5.8 we have

$$\begin{aligned}
 \lim_{k \in K} \nabla \Psi(z^k)^T d^k &= \lim_{k \in K} \nabla p(H(z^k))^T JH(z^k) d^k \\
 &= \lim_{k \in K} \left[-\nabla p(H(z^k))^T H(z^k) + \sigma_k \frac{(a^T H(z^k)) (\nabla p(H(z^k))^T a)}{\|a\|^2} \right] \\
 &= \lim_{k \in K} [-2(\zeta - m) + \sigma_k \mu_k \nabla p(H(z^k))^T a] \\
 &= \lim_{k \in K} \left[-2(\zeta - m) + 2\zeta \sum_{i=1}^{2m} \frac{\sigma_k \mu_k H_{n+i}(z^k)}{\|H(z^k)\|^2} - \sum_{i=1}^{2m} \frac{\sigma_k \mu_k}{H_{n+i}(z^k)} \right] \\
 &= -2(\zeta - m),
 \end{aligned}$$

where for the last equality we used (5.42) which is a consequence of (A3').

On the other hand we consider

$$\begin{aligned}
 \liminf_{k \in K} \frac{\Psi(z^k + \beta^{l_k-1} d^k) - \Psi(z^k)}{\beta^{l_k-1}} &= \liminf_{k \in K} \frac{1}{\beta^{l_k-1}} \left[\zeta \ln(\|H(z^k + \beta^{l_k-1} d^k)\|^2) - \zeta \ln(\|H(z^k)\|^2) \right. \\
 &\quad - \sum_{i=1}^m (\ln((\lambda_i^k + \beta^{l_k-1} d_{\lambda,i}^k)(w_i^k + \beta^{l_k-1} d_{w,i}^k)) - \ln(\lambda_i^k w_i^k)) \\
 &\quad \left. - \sum_{i=1}^m (\ln(g_i(x^k + \beta^{l_k-1} d_x^k) + w_i^k + \beta^{l_k-1} d_{w,i}^k) - \ln(g_i(x^k) + w_i^k)) \right] \\
 &= \liminf_{k \in K} \frac{1}{\beta^{l_k-1}} \left[2\zeta \ln \frac{\|H(z^k + \beta^{l_k-1} d^k)\|}{\|H(z^k)\|} \right. \\
 &\quad - \sum_{i=1}^m \ln \frac{(\lambda_i^k + \beta^{l_k-1} d_{\lambda,i}^k)(w_i^k + \beta^{l_k-1} d_{w,i}^k)}{\lambda_i^k w_i^k} \\
 &\quad \left. - \sum_{i=1}^m \ln \frac{g_i(x^k + \beta^{l_k-1} d_x^k) + w_i^k + \beta^{l_k-1} d_{w,i}^k}{g_i(x^k) + w_i^k} \right].
 \end{aligned}$$

Using Lemma 5.35, the monotonicity of \ln , and $\ln(1 + y) \leq y$ for all $y > -1$, we get for the first term

$$\begin{aligned}
 2\zeta \ln \frac{\|H(z^k + \beta^{l_k-1} d^k)\|}{\|H(z^k)\|} &= 2\zeta \ln \frac{\|(1 - \beta^{l_k-1})H(z^k) + \beta^{l_k-1}q(z^k, \beta^{l_k-1})\|}{\|H(z^k)\|} \\
 &\leq 2\zeta \ln \left(1 - \beta^{l_k-1} + \beta^{l_k-1} \frac{\|q(z^k, \beta^{l_k-1})\|}{\|H(z^k)\|} \right) \\
 &\leq 2\zeta \beta^{l_k-1} \left(-1 + \frac{\|q(z^k, \beta^{l_k-1})\|}{\|H(z^k)\|} \right).
 \end{aligned}$$

For the other terms the following approximation, that is valid for all $t \in [0, 1]$ and $y \in \mathbb{R}$ with $t|y| < 1$, can be used. From the Taylor series of $\ln(1 + ty)$ and the properties of a geometric series one gets

$$-\ln(1 + ty) = -\sum_{j=1}^{\infty} (-1)^{j+1} \frac{(ty)^j}{j} \leq \sum_{j=1}^{\infty} t^j |y|^j = t|y| \sum_{j=0}^{\infty} t^j |y|^j = \frac{t|y|}{1 - t|y|}. \quad (5.44)$$

By (5.42), $\frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} \rightarrow_K 0$, and, as we have seen in the proof of Lemma 5.35, $\beta^{l_k-1} \frac{d_{\lambda,i}^k d_{w,i}^k}{\lambda_i^k w_i^k} \rightarrow_K 0$, since $\beta^{l_k-1} \rightarrow_K 0$. Hence we obtain for the second term for all $i = 1, \dots, m$

$$\begin{aligned} -\ln \frac{(\lambda_i^k + \beta^{l_k-1} d_{\lambda,i}^k)(w_i^k + \beta^{l_k-1} d_{w,i}^k)}{\lambda_i^k w_i^k} &\stackrel{(5.31)}{=} -\ln \frac{(1 - \beta^{l_k-1})\lambda_i^k w_i^k + \beta^{l_k-1} \sigma_k \mu_k + (\beta^{l_k-1})^2 d_{\lambda,i}^k d_{w,i}^k}{\lambda_i^k w_i^k} \\ &= -\ln \left(1 + \beta^{l_k-1} \left(-1 + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} + \beta^{l_k-1} \frac{d_{\lambda,i}^k d_{w,i}^k}{\lambda_i^k w_i^k} \right) \right) \\ &\stackrel{(5.44)}{\leq} \frac{\beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} + \beta^{l_k-1} \frac{d_{\lambda,i}^k d_{w,i}^k}{\lambda_i^k w_i^k} \right|}{1 - \beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} + \beta^{l_k-1} \frac{d_{\lambda,i}^k d_{w,i}^k}{\lambda_i^k w_i^k} \right|}. \end{aligned}$$

For the third term we can use convexity of g_i , i.e.,

$$g_i(x^k + \beta^{l_k-1} d_x^k) \geq g_i(x^k) + \beta^{l_k-1} \nabla g_i(x^k)^T d_x^k,$$

monotonicity of \ln , and $\frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \rightarrow_K 0$ by (5.42) to obtain

$$\begin{aligned} -\ln \frac{g_i(x^k + \beta^{l_k-1} d_x^k) + w_i^k + \beta^{l_k-1} d_{w,i}^k}{g_i(x^k) + w_i^k} &\leq -\ln \frac{g_i(x^k) + w_i^k + \beta^{l_k-1} \left(\nabla g_i(x^k)^T d_x^k + d_{w,i}^k \right)}{g_i(x^k) + w_i^k} \\ &\stackrel{(5.30)}{=} -\ln \left(1 + \beta^{l_k-1} \left(-1 + \frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \right) \right) \\ &\stackrel{(5.44)}{\leq} \frac{\beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \right|}{1 - \beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \right|} \end{aligned}$$

for all $i = 1, \dots, m$. Altogether yields

$$\begin{aligned} \liminf_{k \in K} \frac{\Psi(z^k + \beta^{l_k-1} d^k) - \Psi(z^k)}{\beta^{l_k-1}} &= \liminf_{k \in K} \frac{1}{\beta^{l_k-1}} \left[2\zeta \ln \frac{\|H(z^k + \beta^{l_k-1} d_k)\|}{\|H(z^k)\|} \right. \\ &\quad - \sum_{i=1}^m \ln \frac{(\lambda_i^k + \beta^{l_k-1} d_{\lambda,i}^k)(w_i^k + \beta^{l_k-1} d_{w,i}^k)}{\lambda_i^k w_i^k} \\ &\quad \left. - \sum_{i=1}^m \ln \frac{g_i(x^k + \beta^{l_k-1} d_x^k) + w_i^k + \beta^{l_k-1} d_{w,i}^k}{g_i(x^k) + w_i^k} \right] \end{aligned}$$

$$\begin{aligned}
 &\leq \liminf_{k \in K} \left[2\zeta \left(-1 + \frac{\|q(z^k, \beta^{l_k-1})\|}{\|H(z^k)\|} \right) \right. \\
 &\quad + \sum_{i=1}^m \frac{\left| -1 + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} + \beta^{l_k-1} \frac{d_{\lambda_i, i}^k d_{w, i}^k}{\lambda_i^k w_i^k} \right|}{1 - \beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} + \beta^{l_k-1} \frac{d_{\lambda_i, i}^k d_{w, i}^k}{\lambda_i^k w_i^k} \right|} \\
 &\quad \left. + \sum_{i=1}^m \frac{\left| -1 + \frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \right|}{1 - \beta^{l_k-1} \left| -1 + \frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \right|} \right] \\
 &= -2(\zeta - m),
 \end{aligned}$$

where the last equality follows since $\beta^{l_k-1} \rightarrow_K 0$, $\frac{\|q(z^k, \beta^{l_k-1})\|}{\|H(z^k)\|} \rightarrow_K 0$ by Lemma 5.35, $\frac{\sigma_k \mu_k}{\lambda_i^k w_i^k} \rightarrow_K 0$, $\beta^{l_k-1} \frac{d_{\lambda_i, i}^k d_{w, i}^k}{\lambda_i^k w_i^k} \rightarrow_K 0$, and $\frac{\sigma_k \mu_k}{g_i(x^k) + w_i^k} \rightarrow_K 0$. Thus we finally get

$$-2(\zeta - m) \geq \liminf_{k \in K} \frac{\Psi(z^k + \beta^{l_k-1} d^k) - \Psi(z^k)}{\beta^{l_k-1}} \geq \gamma \lim_{k \in K} \nabla \psi(z^k)^T d^k = -2\gamma(\zeta - m),$$

which contradicts $\gamma < 1$. Therefore the assumption $t_k \rightarrow_K 0$ was wrong, and we have shown that there exists a $\bar{t} > 0$ such that $t_k \geq \bar{t}$ for all $k \in \mathbb{N}$. Thus the sequence $\{t_k\}$ is bounded away from zero. \square

With $\sigma_k \rightarrow 0$ by (A3') we can define $\sigma_k < \frac{1}{2}$ for all $k \in \mathbb{N}$. Then, having a lower bound for the stepsizes, the assumptions of Theorem 5.23 are satisfied and we can guarantee $\|H(z^k)\| \leq \varepsilon$ after at most

$$k = \left\lceil \frac{\psi(z^0)}{\gamma(\zeta - m)\bar{t}} - \frac{\ln \varepsilon}{2\gamma\bar{t}} \right\rceil$$

iterations. This is the desired relation between the number of iterations and the accuracy. The formula gives an idea of the rate of convergence of our algorithm. However, we only proved the existence of the lower bound $\bar{t} > 0$ for the stepsizes and there is no estimate for its size yet, which results in a limited practical applicability of the formula.

6. Numerical Methods

The theoretical results in this thesis suggest algorithms or foundations for them which have to be tested in practice. In the following sections the implementational details for the specific algorithms are described and a comparison of the algorithms is given. The implementation of all algorithms was done in MATLAB[®] (version 7.12.0.635 (R2011a)–64bit from March 18, 2011).

The considered methods can be divided into three groups. One group is the globalized Newton method which is restricted to jointly convex GNEPs and can only find normalized solutions. It is a globalization of the local method from [47] and therefore it is compared to the local variant with different parameters. A second group is the gradient sampling algorithm for solving the unconstrained optimization reformulation using $\bar{V}_{\alpha\beta}^c$. In this case all GNEPs satisfying Assumption 4.4 can be solved, in particular all jointly convex ones. This method is much more expensive at each iteration than the methods of the third group, but due to the sampling strategy the computed solutions spread over the solution set. Finally, all the remaining algorithms form a third group, since they all solve the concatenated KKT system of a player convex GNEP. The main computational effort for all these methods is the solution of a linear equation system at each iteration. Thus the effort of the algorithms in this group is comparable and a performance profile of the iteration numbers, as described below, draws a fair picture of the performance of these methods. Note that also numerical tests for the constrained problem (5.8)

$$\min \Theta(x, \lambda) \quad \text{s.t.} \quad \lambda \geq 0$$

were performed, but the results were not competitive to the other methods. Some more details are given at the end of the Sections 6.3 and 6.6.

As introduced in [17] performance profiles are defined in the following way. Consider a set A of n_a algorithms, a set P of n_p problems and a performance measure $m_{p,a}$ (e.g., number of iterations, function evaluations). The performance on problem $p \in P$ by algorithm $a \in A$ is compared with the best performance by any algorithm on this problem using the *performance ratio*

$$r_{p,a} = \frac{m_{p,a}}{\min\{m_{p,\bar{a}} \mid \bar{a} \in A\}}.$$

Then, one obtains an overall assessment of the performance of the algorithm by defining the value

$$\rho_a(\tau) = \frac{|\{p \in P \mid r_{p,a} \leq \tau\}|}{n_p},$$

which represents the probability for algorithm $a \in A$ that the performance ratio $r_{p,a}$ is within a factor $\tau \in \mathbb{R}$ of the best possible ratio. The function ρ_a represents the distribution

function for the performance ratio. Thus $\rho_a(1)$ gives the fraction of problems for which the algorithm a was the most effective, $\rho_a(2)$ gives the fraction of problems for which the algorithm a is within a factor of 2 of the best algorithm, and so on.

The test problems under consideration are the commonly used ones in GNEP literature. Most of them are taken from the numerical test library in [28]. Additionally we used further problems, namely Harker's problem (Harker) described in [42], an electricity market problem (Heu) from [43], two small problems (NTF1, NTF2) from [56], a transportation problem from [35] in different dimensions (Tr1, Tr2, Tr3), a Spam-filtering problem (Spam) which is a multi-player version of the 2-player game described in [7], and a model for a lobbying process (Lob), see [71]. A detailed description of all the problems can be found in the references. The Table in Appendix A.1 gives an overview about the problem dimensions and the structure of the cost and constraint functions. Note, however, that some of the nonlinear cost functions are not defined everywhere, which can cause trouble, if the algorithms compute iterates there. All the test problems were used for the third group of algorithms, since all methods in this group can deal with player convex GNEPs in its most general form. For the globalized Newton method and its local counterpart only the jointly convex GNEPs can be used. Since the function evaluation of $\bar{V}_{\alpha\beta}^c$ needs a lot of time for complicated problems the gradient sampling algorithm is only tested on the lower dimensional problems. In order to show the distribution of the solutions computed by this method there are four 2-player games, where each player controls one variable, with 100 randomly chosen starting vectors considered. These are the following four examples, the two jointly convex GNEPs NTF1 and NTF2 and two further academic not jointly convex problems, where the solution set can be found graphically using Theorem 2.1.

Example 6.1 This is the NTF1 example from [56], defined by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} (x^1)^2 - x^1 x^2 - x^1 & \text{s.t. } & x^1 \geq 0, x^1 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} (x^2)^2 - \frac{1}{2} x^1 x^2 - 2x^2 & \text{s.t. } & x^2 \geq 0, x^1 + x^2 \leq 1. \end{aligned}$$

The solution set is given by

$$\left\{ (\lambda, 1 - \lambda) \mid \lambda \in \left[0, \frac{2}{3} \right] \right\},$$

see Figure 6.1, and the starting vectors are chosen randomly in $[0, 1]^2$. \diamond

Example 6.2 This is the NTF2 problem, a modification of the previous Example 6.1, that is also from [56].

$$\begin{aligned} \text{Player 1: } & \min_{x^1} (x^1)^2 - x^1 x^2 - x^1 & \text{s.t. } & x^1 \geq 0, (x^1)^2 + (x^2)^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} (x^2)^2 - \frac{1}{2} x^1 x^2 - 2x^2 & \text{s.t. } & x^2 \geq 0, (x^1)^2 + (x^2)^2 \leq 1. \end{aligned}$$

By Figure 6.1 the solution set is

$$\left\{ (\lambda, \sqrt{1 - \lambda^2}) \mid \lambda \in \left[0, \frac{4}{5} \right] \right\},$$

and starting vectors are randomly taken from $[0, 1] \times [0, 2]$. \diamond

Example 6.3 This 2-player game is defined via:

$$\begin{aligned} \text{Player 1: } & \min_{x^1} (x^1 - 2)^2 \quad \text{s.t.} \quad x^1 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} (x^2 - 2)^2 \quad \text{s.t.} \quad x^1 + x^2 \leq 1, x^2 - x^1 \leq 0. \end{aligned}$$

It can be shown, see Figure 6.1, that the solution set is given by

$$\left\{ (\lambda, 1 - \lambda) \mid \lambda \in \left[\frac{1}{2}, 2 \right] \right\}.$$

The starting points are chosen randomly in $[-10, 10]^2$. \diamond

Example 6.4 This problem is given by

$$\begin{aligned} \text{Player 1: } & \min_{x^1} \frac{1}{2}(x^1)^2 + 3x^1x^2 \quad \text{s.t.} \quad x^1 - 3x^2 \leq 2, -3x^1 + x^2 \leq 2, x^1 + x^2 \leq 1, \\ \text{Player 2: } & \min_{x^2} \frac{1}{2}(x^2)^2 + 3x^1x^2 \quad \text{s.t.} \quad x^1 - 3x^2 \leq 2, -3x^1 + x^2 \leq 2. \end{aligned}$$

The corresponding solution set is by Figure 6.1 the set

$$\{(0, 0), (-1, -1)\} \cup \left\{ (\lambda, -\frac{2}{3} + \frac{1}{3}\lambda) \mid \lambda \in \left[\frac{1}{5}, 1 \right] \right\} \cup \left\{ (\lambda, 2 + 3\lambda) \mid \lambda \in \left[-\frac{3}{5}, -\frac{1}{3} \right] \right\}.$$

The starting vectors are taken randomly from $[-2, 2]^2$. \diamond

6.1. Globalized Newton Method via $V_{\alpha\beta}$

In the implementation of Algorithm 3.7 the subproblem

$$\max_y \Psi_\gamma(x, y) \quad \text{s.t.} \quad g(y) \leq 0$$

is solved by the SNOPT solver from the TOMLAB[®] package (version 7.7.0–64bit) at every iteration. The algorithm stops if $\|F_\beta(x^k)\| < \varepsilon$, or if the maximum number of iterations $kmax$ is reached. The algorithm parameters are $\varepsilon = 10^{-6}$, $kmax = 100$, $s = 2.1$, $\rho = 10^{-8}$, $\tau = 0.5$, $\sigma = 10^{-2}$, and for the function $V_{\alpha\beta}$, $\alpha = 10^{-2}$ and $\beta = 1$ are used.

The Table in Appendix A.2 reports the details of the test runs for the globalized Newton method and the original local variant for two different parameters $\gamma \in \{\alpha, \beta\}$. The local method has the same stopping criteria as the globalized one. Additionally the local algorithm stops if the Newton equation (3.11) cannot be solved sufficiently accurate in the sense that $\|H_k d^k + F_\gamma(x^k)\| > 10^{-2}$ for the computed solution d^k . The results show that the local method is very sensitive to the choice of the parameter $\gamma \in \{\alpha, \beta\}$.

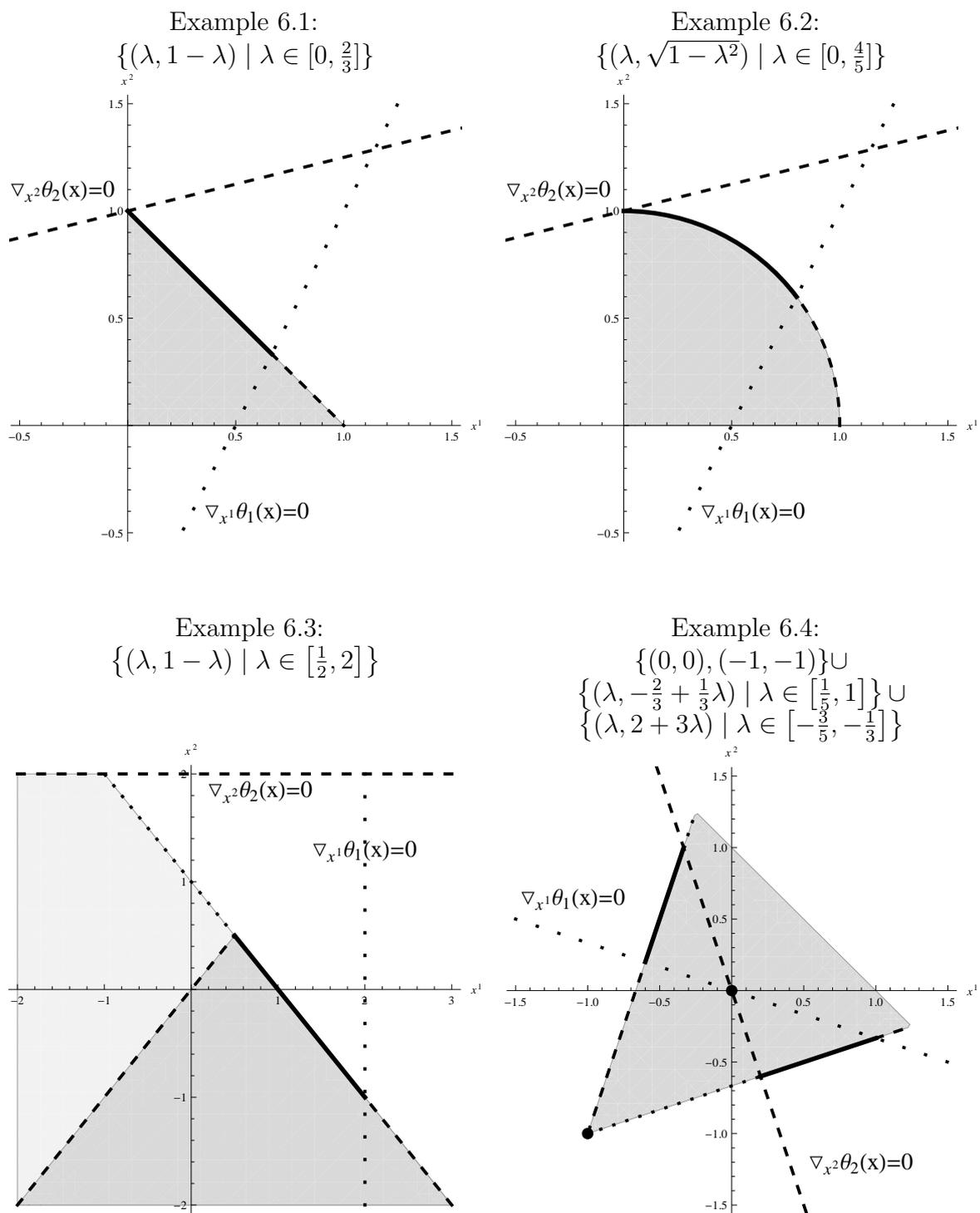


Figure 6.1.: Graphically solutions of the 2-dimensional GNEPs

For the Examples A11–A17, Harker, NTF1, NTF2 and Spam there is not much difference between the local and the globalized Newton methods. Note, however, that for Examples A16a, A16c and A16d, there are failures for the starting point $x^0 = 1000$ for the local method with the smaller parameter $\alpha = 10^{-2}$. Moreover, Example A18 suffers from the singularity of matrices H_k , and hence the globalized method does not use the Newton direction. Therefore the globalized method is just a gradient method, and so it is slowly convergent. The local methods have the same singularity problem in A18, but nevertheless try to compute the Newton direction. Although this fails in most cases, the method is successful in one case (for the smaller parameter α and starting point $x^0 = 1$), since the Newton equation is solved with sufficient accuracy without H_k being nonsingular! Example Heu shows the expected behavior. The globalized method first takes some gradient steps until it is close to a solution, and then switches to Newton steps for fast local convergence. The local method (with the small parameter α) is successful and much faster than the global one for the two starting points closer to the solution, but can not find a solution within the maximum number of iterations for the more far away one. For the larger parameter β , however, the local method can not solve the Newton equation with sufficiently high accuracy for any of the starting points.

The results show that the globalized Newton method is, as expected, more reliable than the local one, because it has the additional option to switch to the anti-gradient step. In fact, the globalized Newton method is able to solve all test problems, whereas the local method has six failures for each of the parameter values α and β (on a different set of examples). The results also verify the finite termination property of the Newton method that was mentioned at the end of Section 3.2. In fact, for most examples, the function value at the last iterate is exactly zero!

6.2. Gradient Sampling Algorithm

To test the nonsmooth unconstrained optimization reformulation using the objective function $\bar{V}_{\alpha\beta}^c$ from Section 4 the robust gradient sampling algorithm from [8] is used. The implementation is the one written by the authors of [8] which is available online, see [9].

With probability one every limiting point of a sequence generated by this method is a Clarke stationary point. The algorithm stops if the norm of the vector with the smallest Euclidian norm in the convex hull of the sampled gradients is less than 10^{-6} . Apart from using standard parameter settings, the three values $\alpha = 0.02$, $\beta = 0.05$ and $c = 10^3$ are used to define the objective function $\bar{V}_{\alpha\beta}^c$. In view of Corollary 4.9, $c = 0$ was used for jointly convex GNEPs. For every function evaluation there are optimization problems that have to be solved to obtain $\bar{y}_\alpha(x)$ and $\bar{y}_\beta(x)$. This is done by the *fmincon* solver from the Matlab Optimization Toolbox. Further if $x \notin X$ the projection onto the convex set X has to be computed, which is either done with the *quadprog* (for polyhedral X) or again with the *fmincon* solver (for nonpolyhedral X) from the Matlab Optimization Toolbox.

By the famous Rademacher Theorem a Lipschitz continuous, and hence also a PC^1 function, is differentiable with probability one at any randomly chosen point. Since the

algorithm computes gradients at randomly chosen points in a neighbourhood of the current iterate it is with probability one enough to be able to compute the gradient of $\bar{V}_{\alpha\beta}^c$ at a differentiable point. This leads to the necessity of the computation of the gradient of the projection mapping at a differentiable point. One possibility to do this is to reformulate the KKT conditions of the corresponding minimization problem as an equation system using a complementarity function and then differentiate the system using chain rules. The required gradient can be computed by solving the resulting system, see [44] for details. The way we compute the gradients $\nabla\bar{V}_\gamma$ for $\gamma = \alpha, \beta$ is by using some results from [69]. Assuming we are at a differentiable point x of all involved functions and that some assumptions like a Slater condition and uniqueness of the multipliers $\lambda_\gamma^\nu(x)$ are satisfied for the problem defining \bar{y}_γ , we obtain, for example by using the formula for the directional derivative from [69, Proposition 4.2.15],

$$\nabla\bar{V}_\gamma(x) = \sum_{\nu=1}^N [\nabla\theta_\nu(x^\nu, x^{-\nu}) - \nabla_x L_\gamma^\nu(x, \bar{y}_\gamma^\nu(x), \lambda_\gamma^\nu(x))],$$

where

$$L_\gamma^\nu(x, \bar{y}_\gamma^\nu(x), \lambda_\gamma^\nu(x)) := \theta_\nu(\bar{y}_\gamma^\nu(x), x^{-\nu}) + \frac{\gamma}{2} \|\bar{y}_\gamma^\nu(x) - x^\nu\|^2 + \sum_{i=1}^N \lambda_{i,\gamma}^\nu(x) g_i^\nu(\bar{y}_\gamma^\nu(x), P_X[x]^{-\nu})$$

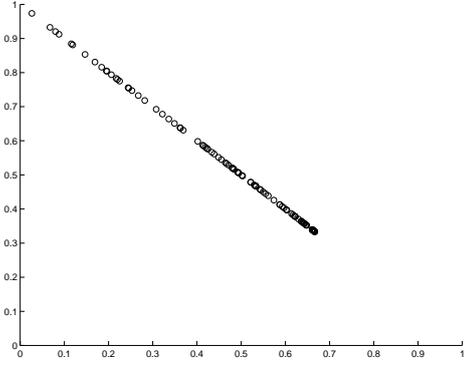
is the Lagrange function of the corresponding problem. Putting all these pieces together we can build a formula for the gradient of $\bar{V}_{\alpha\beta}^c$.

The results obtained by our method are reported in the first two Tables in Appendix A.3. Since the Matlab function *fmincon* computes a solution at a limited precision it is possible that for $\bar{y}_\alpha(x) \approx \bar{y}_\beta(x)$ the objective function values $\bar{V}_{\alpha\beta}^c(x)$ get slightly negative, although theoretically $\bar{V}_{\alpha\beta}^c(x) \geq 0$ holds for all $x \in \mathbb{R}^n$. Note that the use of an unconstrained reformulation may naturally lead to problems with examples were the cost functions are defined only on a subspace of \mathbb{R}^n and the starting vector or some iterate gets out of this subspace. In most cases, however, the solutions are sufficiently far inside the subspace and an appropriate choice of the starting vector leads to successful runs.

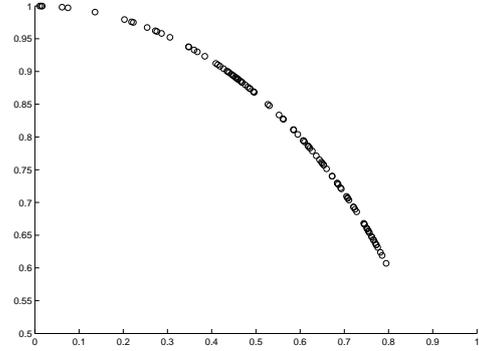
The use of a random sampling strategy in the algorithm may result in different solutions for different test runs, even if the same starting point is used. If the starting points are chosen randomly the computed solutions spread over the entire solution set. To show this effect, four small Examples 6.1 - 6.4 with 2 players, each controlling a single variable, are presented. For these examples the numerical tests were run with 100 randomly chosen starting vectors. The distribution of the solutions is shown in scatter plots in Figure 6.2, where one can see that the solutions obtained by the algorithm spread over the entire solution set that was determined in Figure 6.1 by using Theorem 2.1. For problems where the solution set can not be computed analytically (or where it is at least difficult), one can also see that different solutions are computed, see the last Table in Appendix A.3 for some results on two of the test problems with 10 test runs. Thus if a GNEP has multiple solutions and we are not searching for only one, for example a normalized, this method

with randomly chosen starting vectors can be used to compute different solutions and it gives an idea about the distribution of the solutions. For GNEPs with a unique solution this algorithm is not competitive to the algorithms described in the subsequent sections, because at each iteration $\bar{y}_\alpha(x^k)$ and $\bar{y}_\beta(x^k)$, and for iterates $x^k \notin X$ also $P_X[x^k]$, have to be computed by solving constrained optimization problems, which makes the method very expensive and slow, in particular for large examples or those with more complicated cost or constraint functions.

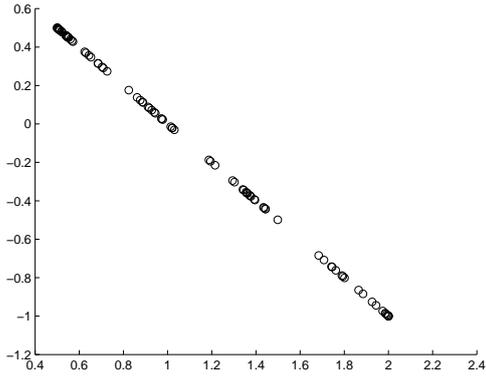
Example 6.1:
 $\{(\lambda, 1 - \lambda) \mid \lambda \in [0, \frac{2}{3}]\}$



Example 6.2:
 $\{(\lambda, \sqrt{1 - \lambda^2}) \mid \lambda \in [0, \frac{4}{5}]\}$



Example 6.3:
 $\{(\lambda, 1 - \lambda) \mid \lambda \in [\frac{1}{2}, 2]\}$



Example 6.4:
 $\{(0, 0), (-1, -1)\} \cup$
 $\{(\lambda, -\frac{2}{3} + \frac{1}{3}\lambda) \mid \lambda \in [\frac{1}{5}, 1]\} \cup$
 $\{(\lambda, 2 + 3\lambda) \mid \lambda \in [-\frac{3}{5}, -\frac{1}{3}]\}$

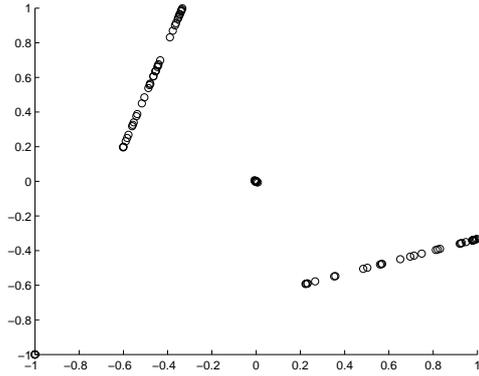


Figure 6.2.: Scatter plots of the solutions of the 2-dimensional GNEPs via $\bar{V}_{\alpha,\beta}^c$

6.3. Matlab's `fminunc` Solver

In the merit function approach to the solution of the GNEP KKT system the unconstrained optimization problem

$$\Theta(x, \lambda) := \frac{1}{2} \left\| \begin{pmatrix} F(x, \lambda) \\ \Phi(x, \lambda) \end{pmatrix} \right\|^2$$

has to be solved. In Section 5.1 the properties of the merit function were studied, but no specific algorithm was proposed. One option is to use a general purpose algorithm that does not exploit in any way the structure of the objective function Θ . This is by far the simplest choice and requires little beyond furnishing routines that calculate the objective and gradient values. In particular the Matlab function `fminunc` from the Optimization Toolbox with option `GradObj` set to 'on' is used. The Matlab option `LargeScale` was set to 'off', so that `fminunc` uses a BFGS line-search algorithm for the minimization. For the complementarity function the penalized Fischer-Burmeister function with parameter $\gamma = 0.975$ is used. Beside the function and the gradient, this routine only requires a starting point (x^0, λ^0) , but no further ingredients.

The main stopping criteria for this method and all the subsequent ones, is a small violation of the KKT conditions (5.2), i.e., the algorithm stops, if

$$V(x, \lambda) := \left\| \begin{pmatrix} F(x, \lambda) \\ \min(\lambda, -g(x)) \end{pmatrix} \right\| \leq \sqrt{n+m} \varepsilon, \quad (6.1)$$

with $\varepsilon = 10^{-4}$. Further a maximum number of 10^3 iterations is allowed.

In addition to the main stopping criteria, `fminunc` stops if the relative change in function value is less than the parameter `TolFun` = 10^{-8} or the maximum number of function evaluations `MaxFunEvals` = 10^5 is reached. The λ -part of the starting vector was always set to $\lambda^0 = 0$, whereas details regarding the x -part are reported in the tables with the results in Appendix A.4.

Using the Matlab function `fminunc` for the unconstrained problem, immediately suggests to use the Matlab function `fmincon` for the constrained problem (5.8). The obtained results are very poor (45 % failures) and thus they are not reported. In view of Example 5.4 and the fact that also the `fminunc` solver is not very successful, this is not surprising.

6.4. Semismooth-like Minimization Algorithm

It should be noted that the general purpose minimization algorithm just described presupposes that the objective function is two times continuously differentiable, but Θ is not so, in fact $\nabla\Theta$ is only strongly semismooth, see [11, 31]. So, as an alternative method for the solution of the unconstrained optimization problem, one can use the semismooth minimization algorithm from [16, 15, 31]. The references give all the details of this globalized semismooth Newton-type method and here are only some relevant implementation details

reported. For the sake of notational simplicity set

$$T(x, \lambda) := \begin{pmatrix} F(x, \lambda) \\ \Phi(x, \lambda) \end{pmatrix},$$

where Φ is defined with the penalized Fischer-Burmeister function with parameter $\gamma = 0.975$. In order to find a search direction an element H of the B-subdifferential $\partial_B T(x, \lambda)$ is evaluated at each iteration, see [31, 64]. This can be done by the following theoretical procedure which is similar to the procedure in [15] for the Fischer-Burmeister function and in [11] for the penalized variant.

Algorithm 6.5 (Procedure to compute an element $H \in \partial_B T(x, \lambda)$)

(S.0) For a given $(x, \lambda) \in \mathbb{R}^{n+m}$ define

$$M_1 = \{(\nu, i) \mid \lambda_i^\nu = 0 = g_i^\nu(x)\} \quad \text{and} \quad M_2 = \{(\nu, i) \mid \lambda_i^\nu > 0, -g_i^\nu(x) > 0\}.$$

(S.1) For $(\nu, i) \in M_1$ set

$$a_i^\nu := -\frac{2 + \sqrt{2}}{2}\gamma \quad \text{and} \quad b_i^\nu := -\frac{2 + \sqrt{2}}{2}\gamma.$$

(S.2) For $(\nu, i) \in M_2$ set

$$\begin{aligned} a_i^\nu &:= \gamma \left(\frac{\lambda_i^\nu}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right) - (1 - \gamma)(-g_i^\nu(x)) \quad \text{and} \\ b_i^\nu &:= \gamma \left(\frac{-g_i^\nu(x)}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right) - (1 - \gamma)\lambda_i^\nu. \end{aligned}$$

(S.3) For $(\nu, i) \notin M_1 \cup M_2$ set

$$a_i^\nu := \gamma \left(\frac{\lambda_i^\nu}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right) \quad \text{and} \quad b_i^\nu := \gamma \left(\frac{-g_i^\nu(x)}{\sqrt{(\lambda_i^\nu)^2 + (g_i^\nu(x))^2}} - 1 \right).$$

(S.4) Using the definitions of Section 5.1.1, set

$$H := \begin{pmatrix} J_x F(x, \lambda) & E(x) \\ -D_g(x, \lambda) J_x g(x) & D_\lambda(x, \lambda) \end{pmatrix}.$$

Lemma 6.6 Assume that for all $\bar{x} \in \mathbb{R}^n$ there exists a sequence $\{x^k\}$ converging to \bar{x} such that $-g_i^\nu(x^k) < 0$ for all $(\nu, i) \in \{(\nu, i) \mid g_i^\nu(\bar{x}) = 0\}$ and all $k \in \mathbb{N}$. Then Algorithm 6.5 computes an element of $\partial_B T(x, \lambda)$.

Proof. Let $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^{n+m}$ be given. By the assumption we can find a sequence $\{x^k\}$ converging to \bar{x} such that

$$-g_i^\nu(x^k) \begin{cases} < 0, & \text{if } -g_i^\nu(\bar{x}) \leq 0, \\ > 0, & \text{if } -g_i^\nu(\bar{x}) > 0, \end{cases}$$

for all $\nu = 1, \dots, N, i = 1, \dots, m_\nu$ and all $k \in \mathbb{N}$. Further let us define the sequence

$$(\lambda^k)_i^\nu := \begin{cases} \bar{\lambda}_i^\nu, & \text{if } \bar{\lambda}_i^\nu \neq 0, \\ -g_i^\nu(x^k), & \text{if } \bar{\lambda}_i^\nu = 0, g_i^\nu(\bar{x}) = 0, \\ -1/k, & \text{if } \bar{\lambda}_i^\nu = 0, g_i^\nu(\bar{x}) \neq 0 \end{cases}$$

for all $\nu = 1, \dots, N, i = 1, \dots, m_\nu$ and all $k \in \mathbb{N}$. Then the sequence $\{(x^k, \lambda^k)\}$ converges to $(\bar{x}, \bar{\lambda})$ and we have $g_i^\nu(x^k) \neq 0$ and $(\lambda^k)_i^\nu \neq 0$ for all $\nu = 1, \dots, N, i = 1, \dots, m_\nu$ and all $k \in \mathbb{N}$. Thus the penalized Fischer-Burmeister function, and hence also T , is differentiable at all points (x^k, λ^k) and the statement follows with the formulas of Section 5.1.1. \square

The assumption of the previous lemma is rather weak and requires that for any $\bar{x} \in \mathbb{R}^n$ where we have active constraints, we can find a sequence $\{x^k\}$ converging to \bar{x} that simultaneously violates all these constraints. If this assumption is not satisfied, we can at least find for each active constraint one sequence that violates this constraint, and we can show analogous to the construction in the above proof that Algorithm 6.5 computes an element of the C-subdifferential of $T(x, \lambda)$, an upper estimate for $\partial_B T(x, \lambda)$. Note that the computational costs to get the matrix H are negligible.

Semismooth Newton methods for solving nonsmooth systems, usually enjoy a superlinear/ quadratic convergence rate under mild assumptions. However, as discussed in great detail in [26], the conditions under which superlinear convergence occur are often in jeopardy when solving reformulations of GNEPs. Since the focus of the algorithms given here is on the global convergence property, the local properties are not analysed and hence there is no guarantee whether the implemented semismooth Newton method enjoys locally fast convergence properties under reasonable assumptions, although, in practice, a fast local convergence was often observed.

The search direction d^k is computed at each iteration by Matlab's linear systems solver *linsolve*, that solves the $(n + m)$ -dimensional square linear system

$$H^k d = -T(x, \lambda) \tag{6.2}$$

with the matrix H^k belonging to the B-subdifferential of $T(x, \lambda)$.

If the Newton-like direction does not satisfy certain ‘‘sufficient descent’’ conditions, the line search is performed along the anti-gradient of Θ . The details are as follows: if the 1-norm condition number estimate of H^k is bigger than 10^{16} (then the linear system (6.2) is ill conditioned), or if

$$\nabla \Theta(x^k, \lambda^k)^T d^k > -10^{-8} \|d^k\|^{2.1}$$

(then d^k is rather orthogonal to $\nabla\Theta(x^k, \lambda^k)$ and thus the succeeding linesearch will generate tiny stepsizes), then d^k is taken as $-\nabla\Theta(x^k, \lambda^k)$.

The linesearch is an Armijo-type one, that computes $t_k = \max\{2^{-\ell} \mid \ell = 0, 1, 2, \dots\}$ such that

$$\Theta((x^k, \lambda^k) + t_k d^k) \leq \Theta(x^k, \lambda^k) + 10^{-4} t_k \nabla\Theta(x^k, \lambda^k)^T d^k.$$

$\lambda^0 = 0$ was always used as λ -part of the starting vector. Furthermore the main stopping criteria (6.1) and a maximum number of 10^3 iterations are used.

6.5. Potential Reduction Method

The used implementation is the exact version of Algorithm 5.9, because the library of test problems considered does not contain large scale problems. More in detail, at step (S.2) of Algorithm 5.9 the search direction d^k is determined by solving a reduced linear system of equations with

$$\sigma_k = \min \left\{ 0.1, 10^4 \frac{\|H(z^k)\| \min\{H_{n+i}(z^k) \mid i = 1, \dots, 2m\}}{\mu_k} \right\}.$$

This choice of σ_k guarantees that d^k is a uniform descent direction in view of Lemma 5.11, and further that Assumption (A3') from section 5.2.3 is satisfied. By this choice for σ_k we obtain slightly different results than in [19], where $\sigma_k = 0.1$ was fixed. Note that formally the method calls for the solution of a $n + 2m$ square linear system at each iteration. However, this system is very structured and some simple manipulations permit to solve it by actually solving a linear system of dimension n . More precisely, a solution $(\bar{d}^1, \bar{d}^2, \bar{d}^3)$ of the following linear system of dimension $n + 2m$

$$\begin{pmatrix} J_x F(x, \lambda) & E(x) & 0 \\ J_x g(x) & 0 & I_m \\ 0 & W & \Lambda \end{pmatrix} \begin{pmatrix} d^1 \\ d^2 \\ d^3 \end{pmatrix} = \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix} \quad (6.3)$$

has to be found, where all the quantities involved are defined in detail in Section 5.2. It is easy to verify, by substitution and by the fact that $w > 0$ in Z_I , that if \bar{d}^1 is a solution of

$$(J_x F(x, \lambda) + E(x)W^{-1}\Lambda J_x g(x)) \bar{d}^1 = b^1 + E(x)W^{-1}\Lambda b^2 - E(x)W^{-1}b^3$$

and \bar{d}^2 and \bar{d}^3 are computed by

$$\begin{aligned} \bar{d}^3 &= b^2 - J_x g(x) \bar{d}^1 && \text{and} \\ \bar{d}^2 &= W^{-1}b^3 - W^{-1}\Lambda \bar{d}^3, \end{aligned}$$

respectively, this is indeed a solution of (6.3). This shows clearly that the main computational burden in solving the linear system (6.3) is actually the solution of an $n \times n$ square linear system, which is done by Matlab's linear systems solver *linsolve*.

Similarly to the semismooth-like approach, if the Newton-like direction does not satisfy

$$\nabla\psi(x^k, \lambda^k, w^k)^T d^k \leq -10^{-5} \|d^k\|^{2.1},$$

that is if the direction d^k is almost orthogonal to $\nabla\psi(x^k, \lambda^k, w^k)$, then the anti-gradient $-\nabla\psi(x^k, \lambda^k, w^k)$ is used as search direction d^k . In view of Lemma 5.11 this can only happen if we were not able to solve the exact equation accurate enough.

For the line search, as described in step (S.3) of Algorithm 5.9, $\gamma = 10^{-2}$ and $\xi = 2m$ are used. $d^k = (d_x^k, d_\lambda^k, d_w^k)$ is preliminarily rescaled in order to stay in Z_I . First an analytically computation of a positive constant α such that $\lambda^k + \alpha d_\lambda^k$ and $w^k + \alpha d_w^k$ are greater than 10^{-14} is made to ensure that the last two blocks in $z^k + \alpha d^k$ are in the interior of \mathbb{R}_+^{2m} . Then, if necessary, a further reduction of α by successive bisections is made until $g(x^k + \alpha d_x^k) + w^k + \alpha d_w^k \geq 10^{-14}$, thus finally guaranteeing that $z^k + \alpha d^k$ belongs to Z_I . In this latter phase, an evaluation of g is needed for each bisection. At the end of this process, set $d^k \leftarrow \alpha d^k$ and proceed to perform the Armijo line-search (5.19).

Since we implemented the exact version of Algorithm 5.9 the choice of the starting vector gives some opportunities, as a consequence of the following Lemma.

Lemma 6.7 *Suppose that for indices $i, j \in \{1, \dots, m\}$ with $g_i \equiv g_j$ we have $\lambda_i^0 = \lambda_j^0$ and $w_i^0 = w_j^0$. Then we have $\lambda_i^k = \lambda_j^k$ and $w_i^k = w_j^k$ for all $k \in \mathbb{N}$.*

Proof. Let $i, j \in \{1, \dots, m\}$ with $g_i \equiv g_j$ be given. By Lemma 5.27 $w_i^0 = w_j^0$ implies $w_i^k = w_j^k$ and $d_{w,i}^k = d_{w,j}^k$ for all $k \in \mathbb{N}$. If $\lambda_i^k = \lambda_j^k$ for some $k \in \mathbb{N}$, we get from equation (5.31)

$$\begin{aligned} d_{\lambda,i}^k &= -\lambda_i^k - \lambda_i^k \frac{d_{w,i}^k}{w_i^k} + \frac{\sigma_k \mu_k}{w_i^k} \\ &= -\lambda_j^k - \lambda_j^k \frac{d_{w,j}^k}{w_j^k} + \frac{\sigma_k \mu_k}{w_j^k} = d_{\lambda,j}^k, \end{aligned}$$

and thus

$$\lambda_i^{k+1} = \lambda_i^k + t_k d_{\lambda,i}^k = \lambda_j^k + t_k d_{\lambda,j}^k = \lambda_j^{k+1},$$

which completes the proof. □

As an immediate consequence, choosing equal starting vectors for repeated constraints leads to equal multipliers at the solution. In particular for jointly convex GNEPs the algorithm then finds normalized Nash equilibria in the sense of [68].

For the (λ, w) -part of the starting vector $\lambda_\nu^0 = 10 - \frac{\nu}{N}$ (which is different for each player) and $w^0 = \max(10, 5 - g(x^0))$ are used to keep the possibility to compute non normalized solutions, and to ensure that the starting point is “well inside” Z_I , thus resulting in larger stepsizes. If x^0 is close to a solution, other choices of the (λ^0, w^0) -part which are also closer to the solution might seem better, but our numerical experiments did not show an overall improvement. Again the main stopping criteria (6.1) and a maximum number of 10^3 iterations are used.

6.6. Scaled Trust-Region Solver for Constrained Nonlinear Equations

As it was shown in Section 5.2 the KKT system can be written as a constrained equation defined by (5.11) and (5.12). This can be solved by STRSCNE (Scaled Trust-Region Solver for Constrained Nonlinear Equations), a software that is freely available, see [4], and whose detailed description can be found in [2, 3]. In order to compare this method with the others, a few details have to be mentioned. STRSCNE is essentially a suitably tailored method that minimizes $\frac{1}{2}\|H(x, \lambda, w)\|^2$ over (5.12). The method uses ellipsoidal trust-regions defined by an affine scaling. The scaling is determined by the nearness of the current iterate to the box boundary and has the effect of angling the scaled steepest descent direction away from the boundary, possibly allowing a longer step to be taken within the feasible region. At each step of the method, a dogleg strategy is used to approximately minimize a quadratic approximation of the objective function over the elliptical trust-region whose shape depends on the bounds. An important property of the proposed method is that all the iterates generated are in the strict interior of the set defined by (5.12). To maintain strict feasibility, suitable restrictions of the chosen steps are performed, if necessary. Note that although STRSCNE is not an interior-point method in the classical sense, it does generate strictly feasible iterates only, and thus comparison with the interior-point method appears particularly appropriate and meaningful.

The algorithm is globally convergent to a stationary point of $\frac{1}{2}\|H(x, \lambda, w)\|^2$ over (5.12). As usual, if the stationary point so found is a global minimizer with zero value, the point is a solution of the constrained system (5.11) and (5.12). However, conditions that guarantee that stationary points are actually solutions of the original constrained system (5.11) and (5.12) are not available at the moment, hence it is not clear how typical nonoptimal stationary points are.

The used algorithm is a slightly modified STRSCNE implementation so that the method uses the same stopping criteria employed by the other methods tested. For the (λ, w) -part of the starting vector $\lambda^0 = 10$ and $w^0 = \max(10, 5 - g(x^0))$ are used. The dogleg strategy used in order to approximately solve the trust region problem entails that, as in all other methods considered for the solution of the KKT formulation, the main computational burden per iteration is the solution of a linear system. More precisely the linear system that is solved at each iteration is exactly the same one considered in the potential reduction algorithm.

Also for the constrained optimization problem (5.8) STRSCNE can be used. However, the numerical solution of the constrained problem leads to failures in 26 % of all test runs, and hence to non competitive results, which are therefore not reported.

6.7. Comparison of the Algorithms Solving the KKT System

In order to compare the algorithms from Sections 6.3 to 6.6, they are all run on the same set of examples using, in some cases, several starting points. This results in 57 runs for each method. At each iteration of the algorithms considered, the most costly operation is the solution of a square linear system. These systems have dimension $n + m$ for the first two algorithms and $n + 2m$ for the last two. However, as discussed in Section 6.5, the system solved by the interior point method can be easily reduced to the solution of a square system of dimension n and this is also possible for the STRSCNE method. It might seem that similar manipulations could be performed also for the system arising in the semismooth method. In fact, the matrix of the linear system is

$$\begin{pmatrix} J_x F(x, \lambda) & E(x) \\ -D_g(x, \lambda) J_x g(x) & D_\lambda(x, \lambda) \end{pmatrix}.$$

The peculiarity of this matrix is that the bottom right block is diagonal. So one could think that, similarly to what is done for the interior point method linear system, one could express the λ variables in function of x and then solve a square n -dimensional system. However, in general the bottom right diagonal block could easily have zero or very small entries. In particular, suppose that $(\bar{x}, \bar{\lambda})$ is a solution of the KKT conditions of the game. If, for example, we have $g_1^1(\bar{x}) = 0$ and $\bar{\lambda}_1^1 > 0$, that is if the first constraint of the first player is active and has a positive multiplier (a common case indeed), one can see that the corresponding element $[D_\lambda(\bar{x}, \bar{\lambda})]_{1,1}$ is 0. So, in a neighbourhood of this point this entry will be either 0 or very small, and one cannot directly exploit the diagonal structure of this block in order to reduce the dimension of the linear system. It is clear that there will be situations (especially in early iterations, probably) where the diagonal elements of $D_\lambda(x, \lambda)$ are all negative, but for the reasons exposed the detection and handling of this diagonal block was left to the linear system solver. Note that, here, the interior point method has an advantage, since the diagonal blocks present in its linear system are always guaranteed to have positive diagonal elements, because the iterates are in Z_I .

The detailed numerical results are reported in the Tables in Appendix A.4. It is obvious that the unconstrained minimization of Θ through the general purpose code *fminunc* is not competitive with the other three approaches and there is not much difference in the results whether the standard Fischer-Burmeister function is used, as in [19], or the penalized variant as used here. This approach leads to very many failures (21) and the iteration numbers are constantly higher than those for the other algorithms. Table 6.1 reports the total number of failures for the semismooth-like algorithm, STRSCNE, and the interior point method, along with the cumulative counts of iterations obtained by considering only runs that are solved by all three algorithms (for a total of 48 runs).

Table 6.1 shows that the interior point method seems more reliable, in that it solves all problems except one. The cumulative iteration numbers are a bit smaller for the interior point algorithm than for the semismooth-like. STRSCNE has a small number of failures

Algorithm	Failures	It.
Semismooth-like	7	846
STRSCNE	3	1633
Interior Point	1	816

Table 6.1.: Cumulative counts of iterations for the semismooth-like algorithm, STRSCNE, and the interior point method for 48 test runs.

but needs about twice as much iterations than the other algorithms. A closer look at the Tables in Appendix A.4 shows that actually the semismooth-like algorithm performs best on a good part of the problems, but it can not solve any of the transportation problems Tr1, Tr2 or Tr3. The large cumulative iteration numbers for the semismooth-like method are mainly caused by Example A8, where many anti-gradient steps are used.

To get a better picture of the behavior of the algorithms, performance profiles with the number of iterations as performance measure are used and the results are shown in Figure 6.3. The left figure shows the performance profiles for small τ , and hence describes the speed of the algorithms, whereas the robustness of the methods can be seen for large τ and thus at the right end of the right figure.

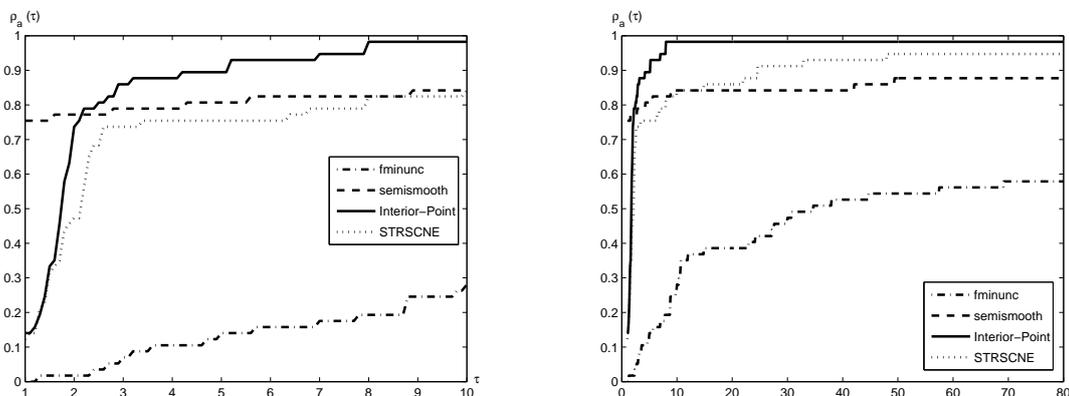


Figure 6.3.: Performance Profiles

These profiles confirm and make the impressions described above more precise. The fminunc algorithm is not competitive to any of the other methods. For $\tau = 1$ the semismooth-like algorithm performs best. However, in comparing the number of iterations one should keep in mind that the dimensions of the linear system solved by the interior point method are, in general, smaller, as discussed at the beginning of this section. As soon as τ is greater than 2,5 (more or less) the interior point method takes the lead, thus showing that the overall performance of this method is not too distant from that of the semismooth-like

and more reliable. The performance of the STRSCNE method is for $\tau < 1,5$ about the same as for the interior point method, but for larger τ the latter one is superior.

There is probably only one other method for which convergence properties are known and a relatively extensive numerical testing has been performed. This is the penalty approach proposed in [28] which is the preprint version with complete numerical results of the paper [29]. It is not totally straightforward to compare the results reported there and those reported here. For one thing, the test set used for the penalty method is a subset of the problems considered in this paper and the stopping criterion is different. Nevertheless, each minor iteration in the penalty method requires the solution of a linear system and, from the linear algebra point of view, this is still the main computational effort of the algorithm. A comparison of the results shows that the solution of the KKT conditions is by far more efficient than the penalty approach.

7. Final Remarks

Let us summarize the results of this thesis. At the beginning we showed two theorems for the analytical solution of small GNEPs. They can be used for the computation of the solution sets of many simple examples, also if no constraint qualification holds.

The main part started with a globalization of a known local Newton method for the computation of normalized Nash equilibria in the case of jointly convex GNEPs. The presented algorithm uses two characterizations of the normalized solutions, a nonsmooth fixed point formulation for the computation of the Newton direction and an optimization formulation with a differentiable merit function as a basis of a linesearch procedure that guarantees global convergence.

Next we considered optimization reformulations of the GNEP using the Nikaido-Isoda function. While the constrained reformulation was known before, we gave a new unconstrained reformulation for a large class of GNEPs including the jointly convex ones. The merit functions of the constrained and unconstrained reformulations were shown to be PC^1 functions, which allows us to use a solver for nonsmooth problems that is able to find Clarke stationary points. Unfortunately we were not able to find conditions guaranteeing that Clarke stationary points are solutions. It was indicated why this is a very challenging future research topic since probably new estimates for the Clarke subdifferential have to be developed.

Moreover, this thesis presented theoretical results on the solution of GNEPs via their KKT conditions. Two different approaches were discussed. The first simply uses a merit function for an equation formulation equivalent to the KKT conditions and the second one develops an interior point algorithm based on a potential function to solve a constrained equation. The conditions guaranteeing that stationary points of the merit function are solutions, and the solvability of the subproblem of the interior point method, respectively, are very similar and require the nonsingularity and P_0 property for certain matrices. However, these properties have to hold only at the feasible sets, and therefore are less restrictive for the constrained equation solver. Furthermore, boundedness of the generated sequences, implying the existence of an accumulation point, was shown under similar mild conditions for both approaches. For the interior point method a finite termination result, describing the relation between the achieved accuracy at the solution and the number of iterations, was proved. It is a further research topic if the result can be improved, such that the step-size bound can be estimated by problem data, thus allowing a more general complexity result.

At the end, we described practical implemented versions for the theoretical approaches and we reported the obtained numerical results, showing benefits of the methods. Finally, the new methods were compared to each other and to existing algorithms and the interior

point method turns out to be very competitive in its overall performance. Moreover the semismooth-like algorithm is quite fast at many of the test problems, but its robustness should be improved.

All the methods were developed with the focus on global convergence. For the subclass of jointly convex GNEPs and the aim of finding a normalized Nash equilibrium the stated globalized Newton method is a robust, global convergent, and also fast local convergent algorithm. This comfortable situation has not been achieved yet by any algorithm dealing with the most general case of GNEPs. The most promising approach is probably the solution of the KKT system. The necessary results for global convergence were presented here. Combining the semismooth-like and the interior point approaches with each other or for example with a Levenberg-Marquardt algorithm, as described in [26], might result in a global and local fast convergent algorithm.

Due to its computational effort the presented algorithm based on the optimization reformulations using the Nikaido-Isoda function is not competitive in convergence speed. However, using the gradient sampling algorithm we have seen that the computed solutions spread over the solution set. Thus the algorithm can be used if we are searching for all (or as many as possible) solutions of a GNEP and not only one, in particular since the number of alternative algorithms dealing with this question is very limited.

Bibliography

- [1] K.J. ARROW AND G. DEBREU: *Existence of an equilibrium for a competitive economy*. *Econometrica* 22, 1954, pp. 265–290.
- [2] S. BELLAVIA, M. MACCONI, AND B. MORINI: *An affine scaling trust-region method approach to bound-constrained nonlinear systems*. *Applied Numerical Mathematics* 44, 2003, pp. 257–280.
- [3] S. BELLAVIA, M. MACCONI, AND B. MORINI: *STRSCNE: A scaled trust-region solver for constrained nonlinear equations*. *Computational Optimization and Applications* 28, 2004, pp. 31–50.
- [4] S. BELLAVIA, M. MACCONI, AND B. MORINI: *Matlab implementation of STRSCNE*. <http://strscne.de.unifi.it>
- [5] A. BENSOUSSAN: *Points de Nash dans le cas de fonctionnelles quadratiques et jeux différentiels linéaires à N personnes*. *SIAM Journal on Control* 12, 1974, pp. 460–499.
- [6] C. BERGE: *Topological Spaces*. Oliver and Boyd, Edinburgh/London, 1963.
- [7] M. BRÜCKNER AND T. SCHEFFER: *Nash equilibria of static prediction games*. *Advances in Neural Information Processing Systems*, MIT Press, 2009, pp. 171–179.
- [8] J.V. BURKE, A.S. LEWIS, AND M.L. OVERTON: *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization*. *SIAM Journal on Optimization* 15, 2005, pp. 751–779.
- [9] J.V. BURKE, A.S. LEWIS, AND M.L. OVERTON: *Matlab implementation of the robust gradient sampling algorithm for nonsmooth, nonconvex optimization*. <http://www.cs.nyu.edu/overton/papers/gradsamp>
- [10] R.W. CHANEY: *Piecewise C^k functions in nonsmooth analysis*. *Nonlinear Analysis: Theory, Methods & Applications* 15, 1990, pp. 649–660.
- [11] B. CHEN, X. CHEN, AND C. KANZOW: *A penalized Fischer-Burmeister NCP-function*. *Mathematical Programming* 88, 2000, pp. 211–216.
- [12] F.H. CLARKE: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia, 1990.
- [13] R.W. COTTLE, J.-S. PANG, AND R.E. STONE: *The Linear Complementarity Problem*. Academic Press, Boston, 1992.

- [14] A.-A. COURNOT: *Recherches sur les principes mathématiques de la théorie des richesses*. Hachette, Paris, 1838.
- [15] T. DE LUCA, F. FACCHINEI, AND C. KANZOW: *A semismooth equation approach to the solution of nonlinear complementarity problems*. *Mathematical Programming* 75, 1996, pp. 407–439.
- [16] T. DE LUCA, F. FACCHINEI, AND C. KANZOW: *A theoretical and numerical comparison of some semismooth algorithms for complementarity problems*. *Computational Optimization and Applications* 16, 2000, pp. 173–205.
- [17] E.D. DOLAN AND J.J. MORÉ: *Benchmarking optimization software with performance profiles*. *Mathematical Programming Series A*, Vol. 91, 2002, pp. 201–213.
- [18] D. DORSCH, H.TH. JONGEN, AND V. SHIKHMAN: *On structure and computation of generalized Nash equilibria*. Preprint 143, Department of Mathematics C, RWTH Aachen University, Germany, 2011.
- [19] A. DREVES, F. FACCHINEI, C. KANZOW, AND S. SAGRATELLA: *On the solution of the KKT conditions of generalized Nash equilibrium problems*. *SIAM Journal on Optimization* 21, 2011, pp. 1082–1108.
- [20] A. DREVES, A. VON HEUSINGER, C. KANZOW, AND M. FUKUSHIMA: *A globalized Newton method for the computation of normalized Nash equilibria*. *Journal of Global Optimization*, to appear, DOI 10.1007/s10898-011-9824-9.
- [21] A. DREVES AND C. KANZOW: *Nonsmooth optimization reformulations characterizing all solutions of jointly convex generalized Nash equilibrium problems*. *Computational Optimization and Applications* 50, 2011, pp. 23–48.
- [22] A. DREVES, C. KANZOW, AND O. STEIN: *Nonsmooth optimization reformulations of player convex generalized Nash equilibrium problems*. *Journal of Global Optimization*, to appear, DOI: 10.1007/s10898-011-9727-9.
- [23] F. FACCHINEI, A. FISCHER, AND C. KANZOW: *Regularity properties of semismooth reformulations of variational inequalities*. *SIAM Journal on Optimization* 8, 1998, pp. 850–869.
- [24] F. FACCHINEI, A. FISCHER, C. KANZOW, AND J.-M. PENG: *A simply constrained optimization reformulation of KKT systems arising from variational inequalities*. *Applied Mathematics and Optimization* 40, 1999, pp. 19–37.
- [25] F. FACCHINEI, A. FISCHER, AND V. PICCIALLI: *On generalized Nash games and variational inequalities*. *Operations Research Letters* 35, 2007, pp. 159–164.
- [26] F. FACCHINEI, A. FISCHER, AND V. PICCIALLI: *Generalized Nash equilibrium problems and Newton methods*. *Mathematical Programming* 117, 2009, pp. 163–194.

-
- [27] F. FACCHINEI AND C. KANZOW: *Generalized Nash equilibrium problems*. Annals of Operations Research 175, 2010, pp. 177–211.
- [28] F. FACCHINEI AND C. KANZOW: *Penalty methods for the solution of generalized Nash equilibrium problems (with complete test problems)*. Technical Report, Institute of Mathematics, University of Würzburg, Germany, 2009.
- [29] F. FACCHINEI AND C. KANZOW: *Penalty methods for the solution of generalized Nash equilibrium problems*. SIAM Journal on Optimization 20, 2010, pp. 2228–2253.
- [30] F. FACCHINEI AND L. LAMPARIELLO: *Partial penalization for the solution of generalized Nash equilibrium problems*. Journal of Global Optimization, Vol. 50, No. 1, 2011, pp. 39–57.
- [31] F. FACCHINEI AND J.-S. PANG: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research, Springer Verlag, New York, 2003.
- [32] F. FACCHINEI AND J.-S. PANG: *Exact penalty functions for generalized Nash problems*. In G. DI PILLO AND M. ROMA (eds): *Large Scale Nonlinear Optimization*. Springer, Berlin, 2006, pp. 115–126.
- [33] F. FACCHINEI AND J.-S. PANG: *Nash equilibria: the variational approach*. In D.P. PALOMAR AND Y.C. ELДАР (eds): *Convex Optimization in Signal Processing and Communications*. Cambridge University Press, 2010, pp. 443–493.
- [34] F. FACCHINEI, V. PICCIALI, AND M. SCIANDRONE: *Decomposition algorithms for generalized potential games*. Computational Optimization and Applications, Vol. 50, No. 2, 2011, pp. 237–262.
- [35] F. FACCHINEI, A. SGALAMBRO AND S. TEOBALDO: *On some Generalized Nash equilibrium problems in a freight distribution environment: properties and existence conditions*. in preparation.
- [36] F. FACCHINEI AND J. SOARES: *A new merit function for nonlinear complementarity problems and a related algorithm*. SIAM Journal on Optimization 7, 1997, pp. 225–247.
- [37] S.D. FLÅM AND R. RUSZCZYŃSKI: *Noncooperative convex games: Computing equilibrium by partial regularization*. IIASA Working Paper 94–42, Laxenburg, Austria, 1994.
- [38] M. FUKUSHIMA: *Restricted generalized Nash equilibria and controlled penalty algorithm*. Computational Management Science 8, No. 3, 2011, pp. 201–218.
- [39] M. FUKUSHIMA: *A class of gap functions for quasi-variational inequality problems*. Journal of Industrial and Management Optimization 3, 2007, pp. 165–171.

- [40] M. FUKUSHIMA AND C. KANZOW: *Solving box constrained variational inequalities by using the natural residual with D-gap function globalization*. Operations Research Letters 23, 1998, pp. 45–51.
- [41] G. GÜRKAN AND J.-S. PANG: *Approximations of Nash equilibria*. Mathematical Programming 117, 2009, pp. 223–253.
- [42] P.T. HARKER: *Generalized Nash games and quasi-variational inequalities*. European Journal of Operational Research 54, 1991, pp. 81–94.
- [43] A. VON HEUSINGER: *Numerical methods for the solution of the generalized Nash equilibrium problem*. PhD Thesis, Institute of Mathematics, University of Würzburg, Germany, 2009.
- [44] A. VON HEUSINGER AND C. KANZOW: *SC^1 -optimization reformulations of the generalized Nash equilibrium problem*. Optimization Methods and Software 23, 2008, pp. 953–973.
- [45] A. VON HEUSINGER AND C. KANZOW: *Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions*. Computational Optimization and Applications 43, 2009, pp. 353–377.
- [46] A. VON HEUSINGER AND C. KANZOW: *Relaxation methods for generalized Nash equilibrium problems with inexact line search*. Journal of Optimization Theory and Applications 143, 2009, pp. 159–183.
- [47] A. VON HEUSINGER, C. KANZOW, AND M. FUKUSHIMA: *Newton’s method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation*. Mathematical Programming, to appear, DOI: 10.1007/s10107-010-0386-2.
- [48] W.W. HOGAN: *Point-to-set maps in mathematical programming*. SIAM Review 15, 1973, pp. 591–603.
- [49] R. JANIN: *Directional derivative of the marginal function in nonlinear programming*. Mathematical Programming Study 21, 1984, pp. 110–126.
- [50] K. KUBOTA AND M. FUKUSHIMA: *Gap function approach to the generalized Nash equilibrium problem*. Journal of Optimization Theory and Applications 144, 2010, pp. 511–531.
- [51] J.B. KRAWCZYK AND S. URYASEV: *Relaxation algorithms to find Nash equilibria with economic applications*. Environmental Modeling and Assessment 5, 2000, pp. 63–73.

-
- [52] B. LUDERER, L. MINCHENKO, AND T. SATSURA: *Multivalued Analysis and Non-linear Programming Problems with Perturbations*. Kluwer Academic Publishers, Dordrecht, Netherlands, 2002.
- [53] L.C. MATIOLI, W. SOSA, AND J. YUAN: *A numerical algorithm for finding solutions of a generalized Nash equilibrium problem*. Computational Optimization and Applications, to appear, DOI: 10.1007/s10589-011-9407-1.
- [54] R.D.C. MONTEIRO AND J.-S. PANG: *A potential reduction Newton method for constrained equations*. SIAM Journal on Optimization 9, 1999, pp. 729–754.
- [55] J.J. MORÉ AND D.C. SORENSEN: *Computing a trust region step*. SIAM Journal on Scientific and Statistical Computing 4, 1983, pp. 553–572.
- [56] K. NABETANI, P. TSENG, AND M. FUKUSHIMA: *Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints*. Computational Optimization and Applications 48, 2011, pp. 423–452.
- [57] J.F. NASH: *Equilibrium points in n -person games*. Proceedings of the National Academy of Sciences of the USA, Vol. 36, No. 1, 1950, pp. 48–49.
- [58] J.F. NASH: *Non-cooperative games*. Annals of Mathematics 54, 1951, pp. 286–295.
- [59] H. NIKAIDO AND K. ISODA: *Note on noncooperative convex games*. Pacific Journal of Mathematics 5, 1955, pp. 807–815.
- [60] J.-S. PANG: *Computing generalized Nash equilibria*. Technical Report, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD, 2002.
- [61] 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.
- [62] J.-S. PANG, G. SCUTARI, F. FACCHINEL, AND C. WANG: *Distributed power allocation with rate constraints in Gaussian parallel interference channels*. IEEE Transactions on Information Theory 54, 2008, pp. 3471–3489.
- [63] J.-S. PANG AND D. RALPH: *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*. Mathematics of Operations Research 21, 1996, pp. 401–426.
- [64] L. QI AND J. SUN: *A nonsmooth version of Newton’s method*. Mathematical Programming Series A, Vol. 58, 1993, pp. 353–368.
- [65] D. RALPH AND S. DEMPE: *Directional derivatives of the solution of a parametric nonlinear program*. Mathematical Programming 70, 1995, pp. 159–172.
- [66] R.T. ROCKAFELLAR: *Convex Analysis*. Princeton University Press, New Jersey, 1970.

- [67] R.T ROCKAFELLAR AND R.J.-B. WETS: *Variational Analysis*. A Series of Comprehensive Studies in Mathematics, Vol. 317, Springer, Berlin, Heidelberg, 1998.
- [68] J.B. ROSEN: *Existence and uniqueness of equilibrium points for concave N -person games*. *Econometrica* 33, 1965, pp. 520–534.
- [69] O. STEIN: *Bi-level Strategies in Semi-Infinite Programming*. Kluwer, Boston, 2003.
- [70] K. TAJI: *On gap functions for quasi-variational inequalities*. *Abstract and Applied Analysis* 2008, DOI 10.1155/2008/531361.
- [71] G. TULLOCK: *Efficient rent seeking*. In J. BUCHANAN, R. TOLLISON, AND G. TULLOCK (eds): *Towards a Theory of the Rent-Seeking Society*. A & M University Press, 1980, pp. 97–112.
- [72] S. URYASEV AND R.Y. RUBINSTEIN: *On relaxation algorithms in computation of noncooperative equilibria*. *IEEE Transactions on Automatic Control* 39, 1994, pp. 1263–1267.
- [73] J.Z. ZHANG, B. QU, AND N.H. XIU: *Some projection-like methods for the generalized Nash equilibria*. *Computational Optimization and Applications* 45, 2010, pp. 89–109.

A. Tables for Numerical Results

The appendix contains a couple of tables with some more details regarding the test problems and the numerical results presented in Section 6. More precisely,

- Appendix A.1 contains an overview of test problems that were used. For every example the number of players N and the total number of variables n and constraints m are reported. Some of the test problems were run more than once, using different starting points. The number of runs is given in column “runs”. For a better overview the Table also shows to which “class” the problems belong, a pure NEP, a jointly convex problem or a general GNEP. Further the cost functions $\theta_\nu(\cdot, x^{-\nu})$ are for fixed $x^{-\nu}$ either linear or quadratic or non linear and similar the constraint functions $g^\nu(\cdot, x^{-\nu})$ are either linear or nonlinear. For a detailed description see the references given in Section 6.
- Appendix A.2 shows the results for the local and the globalized Newton methods. The name of the example and the chosen starting point x^0 are listed in the first two columns. Most of the starting points x^0 are chosen so that all of their components are identical. For simplicity of notation, $x^0 = 1$ is written instead of $x^0 = (1, 1, \dots, 1)^T$. The next three columns of the table show the number of iterations “It.”, the number of gradient steps “grad”, and the function value “ $\|F_\beta(x)\|$ ” at the computed solution for the globalized Newton method. Since the method is a globalization of the local Newton method from [47], there are also numerical results for the local method reported, using both the parameter α in columns 6 and 7 and the parameter β in columns 8 and 9. If the maximum number of iterations is reached or if the Newton direction can not be computed accurate enough an “F” is reported instead of the function value at the solution, thus indicating a failure of the method.
- the tables in Appendix A.3 show numerical results for the gradient sampling algorithm for general GNEPs, jointly convex GNEPs and two problems with randomly chosen starting vectors, respectively. As before the column “Example” contains the name of the test problem, the x^0 -column gives details on the starting vector and “It.” lists the number of iterations until convergence was obtained. The column “ x^* ” contains the computed solution and the function value at the solution is in the last column “ $\bar{V}_{\alpha\beta}^c(x^*)$ ”. The first two Tables in Appendix A.3 only report results for lower dimensional examples, where the solution vector can be printed in a few lines. For the two problems reported in the last Table the 10 starting vectors are chosen randomly in $(0, 10)^n$, hence there is no extra column “ x^0 ” given there.

- the Tables in Appendix A.4 report the results for the four algorithms used to solve the KKT conditions of the GNEPs. For each method the column “It.” gives the iteration number or an “F” for a failure, and the column “merit” contains the value of the merit function from (6.1), when the algorithm stops. The “Example” and “ x^0 ” columns are as before. For the Examples A10a–A10e the starting vectors from [28] are used, where most components are 0 and some of the last components are chosen equal and sum up to 1. This is indicated by an “0,=” in the column “ x^0 ”. For the other parts of the starting vectors standard values are used as described in the corresponding sections. The smallest number of iterations from any of the four algorithms is printed in boldface for each run.

A.1. Data of the Test Problems

Example	N	n	m	runs	class	θ_ν	g^ν
A1	10	10	20	3	general GNEP	non linear	linear
A2	10	10	24	3	general GNEP	non linear	linear
A3	3	7	18	3	general GNEP	quadratic	linear
A4	3	7	18	3	general GNEP	quadratic	linear
A5	3	7	18	3	general GNEP	quadratic	linear
A6	3	7	21	3	general GNEP	quadratic	non linear
A7	4	20	44	3	general GNEP	quadratic	linear
A8	3	3	8	3	general GNEP	quadratic	linear
A9a	7	56	63	1	general GNEP	linear	non linear
A9b	7	112	119	1	general GNEP	linear	non linear
A10a	8	24	33	1	general GNEP	quadratic	non linear
A10b	25	125	151	1	general GNEP	non linear	non linear
A10c	37	222	260	1	general GNEP	quadratic	non linear
A10d	37	370	408	1	general GNEP	non linear	non linear
A10e	48	576	625	1	general GNEP	non linear	non linear
Tr1	6	18	72	2	general GNEP	non linear	linear
Tr2	6	60	228	2	general GNEP	non linear	linear
Tr3	7	80	304	2	general GNEP	non linear	linear
A11	2	2	2	1	jointly convex	quadratic	linear
A12	2	2	4	1	NEP	quadratic	linear
A13	3	3	9	1	jointly convex	quadratic	linear
A14	10	10	20	1	jointly convex	non linear	linear
A15	3	6	12	1	NEP	quadratic	linear
A16a	5	5	10	1	jointly convex	non linear	linear
A16b	5	5	10	1	jointly convex	non linear	linear
A16c	5	5	10	1	jointly convex	non linear	linear
A16d	5	5	10	1	jointly convex	non linear	linear
A17	2	3	7	1	jointly convex	quadratic	linear
A18	2	12	28	3	jointly convex	quadratic	linear
Harker	2	2	6	1	jointly convex	quadratic	linear
Heu	2	10	22	2	jointly convex	non linear	linear
NTF1	2	2	4	1	jointly convex	quadratic	linear
NTF2	2	2	4	1	jointly convex	quadratic	non linear
Lob	50	50	50	1	NEP	non linear	linear
Spam	101	2020	4040	1	NEP	non linear	linear

A.2. The Globalized and Local Newton Methods

Example	x^0	globalized Newton method			local method with α		local method with β	
		It.	grad	$\ F_\beta(x)\ $	It.	$\ F_\alpha(x)\ $	It.	$\ F_\beta(x)\ $
A11	0	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
A11	1	1	0	0.0e+00	1	0.0e+00	1	0.0e+00
A11	100	1	0	0.0e+00	1	0.0e+00	1	0.0e+00
A12	0	1	0	0.0e+00	1	0.0e+00	1	0.0e+00
A12	1	1	0	0.0e+00	1	0.0e+00	1	0.0e+00
A12	100	1	0	0.0e+00	3	0.0e+00	1	0.0e+00
A13	0	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
A13	1	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
A13	100	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
A14	0.01	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A14	1	3	0	0.0e+00	4	0.0e+00	3	0.0e+00
A14	100	4	1	0.0e+00	3	0.0e+00	3	0.0e+00
A15	0	1	0	0.0e+00	2	0.0e+00	1	0.0e+00
A15	1	1	0	0.0e+00	2	0.0e+00	1	0.0e+00
A15	100	2	0	0.0e+00	3	0.0e+00	2	0.0e+00
A16a	10	3	0	0.0e+00	2	0.0e+00	3	0.0e+00
A16a	100	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16a	1000	3	0	0.0e+00	1	F	3	0.0e+00
A16b	10	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16b	100	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16b	1000	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16c	10	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16c	100	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16c	1000	3	0	0.0e+00	2	F	3	0.0e+00
A16d	10	4	0	0.0e+00	2	0.0e+00	4	0.0e+00
A16d	100	3	0	0.0e+00	3	0.0e+00	3	0.0e+00
A16d	1000	3	0	0.0e+00	2	F	3	0.0e+00
A17	0	2	0	0.0e+00	3	0.0e+00	2	0.0e+00
A17	1	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
A17	100	2	0	0.0e+00	4	0.0e+00	2	0.0e+00
A18	0	17	17	2.9e-07	4	F	9	F
A18	1	17	17	2.9e-07	2	0.0e+00	8	F
A18	100	14	14	3.2e-07	7	F	1	F
Harker	0	1	0	0.0e+00	1	0.0e+00	1	0.0e+00
Heu	100	33	30	0.0e+00	4	0.0e+00	1	F
Heu	500	17	13	0.0e+00	6	0.0e+00	1	F
Heu	1000	19	14	0.0e+00	100	F	1	F
NTF1	0	2	0	0.0e+00	1	0.0e+00	2	0.0e+00
NTF2	0	5	0	0.0e+00	5	0.0e+00	5	0.0e+00
Spam	1	3	0	0.0e+00	3	0.0e+00	3	0.0e+00

A.3. The Gradient Sampling Algorithm

Example	x^0	It.	x^*	$V_{\alpha\beta}^c(x^*)$
A1	0.01	36	(0.3002, 0.0696, 0.0693, 0.0694, 0.0694, 0.0692, 0.0690, 0.0695, 0.0697, 0.0696)	8.4e-09
A1	0.1	27	(0.3000, 0.0693, 0.0693, 0.0692, 0.0692, 0.0696, 0.0695, 0.0693, 0.0697, 0.0695)	-4.8e-07
A1	1	67	(0.3002, 0.0694, 0.0695, 0.0694, 0.0694, 0.0693, 0.0692, 0.0696, 0.0693, 0.0694)	3.0e-09
A2	0.01	66	(0.3022, 0.0323, 0.0336, 0.0326, 0.2031, 0.2916, 0.0240, 0.0237, 0.0255, 0.0216)	3.0e-06
A2	0.1	36	(0.3001, 0.0174, 0.0258, 0.0279, 0.2759, 0.3008, 0.0105, 0.0103, 0.0105, 0.0108)	2.5e-08
A2	1	F		-2.3e+13
A3	0	53	(-0.3850, -0.1233, -1.0066, 0.3947, 1.1770, 0.0514, 0.0187)	7.5e-08
A3	1	82	(-0.3749, -0.1272, -0.9857, 0.3833, 1.1495, 0.0482, 0.0187)	-3.8e-07
A3	10	153	(-0.3787, -0.1207, -0.9843, 0.3881, 1.1513, 0.0513, 0.0187)	-4.2e-07
A4	0	31	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	3.5e-12
A4	1	6	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	0.0e+00
A4	10	51	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	9.7e-12
A5	0	87	(0.0004, 0.2021, 0.0002, 0.0002, 0.0705, 0.0261, 0.0004)	-4.8e-07
A5	1	42	(0.0002, 0.2030, 0.0002, 0.0000, 0.0724, 0.0253, 0.0001)	-3.6e-07
A5	10	95	(0.0024, 0.2010, 0.0027, 0.0008, 0.0661, 0.0248, 0.0027)	-1.1e-10
A6	0	35	(1.0000, 1.0000, 1.0000, 1.4167, 1.0000, 1.0000, 1.0000)	2.7e-11
A6	1	53	(1.0000, 1.0000, 1.0000, 1.4167, 1.0000, 1.0000, 1.0000)	-4.2e-07
A6	10	81	(1.0000, 1.0000, 1.0000, 1.4167, 1.0000, 1.0000, 1.0000)	3.1e-11
A7	0	133	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.8427, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	1.3e-11
A7	1	123	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.8427, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	1.8e-11
A7	10	136	(1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.8427, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000)	1.0e-11
A8	0	29	(0.4991, 0.5007, 0.7487)	-5.0e-07
A8	1	29	(0.4993, 0.4988, 0.7490)	-5.0e-07
A8	10	43	(0.4995, 0.5003, 0.7489)	1.3e-08

A. Tables for Numerical Results

Example	x^0	It.	x^*	$\bar{V}_{\alpha\beta}^c(x^*)$
A11	0	11	(0.6933, 0.3067)	2.6e-11
A12	0	24	(5.3337, 5.3332)	0.0e+00
A13	0	34	(10.6017, 8.3580, 13.3568)	4.4e-13
A14	0.01	38	(0.090913, 0.091040, 0.089130, 0.089906, 0.089606, 0.089806, 0.089658, 0.089857, 0.088901, 0.087955)	-8.6e-08
A15	0	111	(46.6617, 32.1544, 15.0025, 22.1071, 12.3397, 12.3394)	0.0e+00
A16a	10	46	(14.1175, 14.7200, 15.2233, 15.6173, 15.3269)	7.5e-08
A16b	10	111	(18.4665, 19.5989, 20.4153, 20.6567, 20.8715)	2.4e-07
A16c	10	90	(27.3867, 30.2220, 31.5729, 31.1880, 29.6344)	4.7e-08
A16d	10	79	(35.7361, 40.4337, 42.7623, 42.1098, 38.9585)	5.5e-10
A17	0	37	(1.6724, 9.3276, 6.3276)	5.4e-12
A18	0	117	(43.5366, 28.1381, 28.3253, 26.8698, 11.4714, 11.6588, 43.5364, 28.1381, 28.3256, 26.8696, 11.4713, 11.6591)	1.2e-11
Harker	0	30	(9.2669, 5.7331)	2.2e-13
Heu	100	506	(117.6409, 356.6286, 176.1078, 328.4737, 121.6064, 101.6896, 300.3188, 178.3936, 328.4737, 178.3936)	-4.7e-12
NTF1	0	13	(0.1838, 0.8162)	1.9e-13
NTF2	0	7	(0.2922, 0.9564)	0.0e+00

Example	It.	x^*	$\bar{V}_{\alpha\beta}^c(x^*)$
A13	23	(13.8502, 5.8946, 11.5439)	1.2e-11
	21	(9.2381, 12.6505, 13.1304)	1.1e-11
	22	(9.9375, 4.7432, 14.9756)	1.4e-11
	24	(9.1918, 10.2697, 13.8884)	7.4e-12
	27	(14.8561, 12.1414, 8.8584)	1.3e-12
	21	(7.9660, 12.8074, 14.0852)	3.7e-12
	27	(11.8854, 14.8102, 10.3902)	5.7e-12
	28	(11.8203, 14.4521, 10.5500)	2.2e-13
	32	(10.3767, 9.0941, 13.3111)	1.5e-12
	28	(13.3028, 8.6643, 11.1359)	5.7e-12
A18	110	(47.7517, 25.4226, 26.8256, 25.2728, 14.1868, 10.5404, 43.2123, 26.0796, 30.7082, 24.5756, 13.5299, 11.8944)	2.2e-12
	105	(47.1941, 27.8550, 24.9509, 23.9604, 11.7548, 14.2848, 43.1392, 28.2309, 28.6299, 26.5184, 11.3792, 12.1024)	7.7e-13
	106	(41.9968, 28.6816, 29.3217, 28.1979, 10.9279, 10.8743, 40.0750, 30.5576, 29.3675, 30.5428, 9.0520, 10.4052)	7.0e-12
	105	(45.4940, 26.4997, 28.0063, 27.2335, 13.1097, 9.6568, 42.8491, 26.0692, 31.0817, 25.2358, 13.5404, 11.2239)	4.6e-12
	108	(43.3011, 28.9710, 27.7279, 27.6139, 10.6384, 11.7477, 41.5365, 29.1595, 29.3039, 28.3609, 10.4499, 11.1892)	1.2e-11
	104	(42.5858, 28.3888, 29.0254, 27.7189, 11.2206, 11.0605, 43.6225, 26.9470, 29.4305, 26.8852, 12.6625, 10.4523)	9.0e-12
	103	(44.4937, 28.1075, 27.3988, 26.7454, 11.5020, 11.7527, 42.2235, 29.0464, 28.7301, 27.3499, 10.5630, 12.0871)	1.8e-12
	115	(44.3022, 29.3935, 26.3043, 26.9583, 10.2161, 12.8256, 42.9913, 26.8983, 30.1103, 26.5605, 12.7111, 10.7284)	8.4e-12
	106	(43.2772, 29.5307, 27.1922, 29.7289, 10.0788, 10.1923, 44.1572, 27.3817, 28.4611, 23.6492, 12.2278, 14.1231)	8.2e-12
	99	(46.1020, 28.1984, 25.6996, 26.3866, 11.4112, 12.2022, 42.6535, 26.1116, 31.2349, 25.6702, 13.4978, 10.8319)	4.5e-12

A.4. Results for Solving the KKT System

Example	x^0	fminunc		semismooth		Interior Point		STRSCNE	
		It.	merit	It.	merit	It.	merit	It.	merit
A1	0.01	68	7.7e-05	7	1.5e-07	13	4.8e-05	9	8.0e-06
A1	0.1	28	8.9e-05	4	3.1e-06	10	6.5e-05	F	3.5e-01
A1	1	44	9.9e-05	5	1.5e-07	10	8.1e-05	9	2.8e-05
A2	0.01	217	4.8e-04	8	5.5e-06	23	1.0e-04	17	1.6e-05
A2	0.1	138	4.6e-04	5	2.6e-05	21	9.8e-05	34	7.6e-05
A2	1	217	4.4e-04	9	4.3e-06	46	8.6e-05	432	8.9e-05
A3	0	118	9.5e-05	1	1.5e-15	8	1.6e-05	10	5.6e-05
A3	1	118	6.2e-05	1	6.0e-15	8	1.6e-05	8	7.0e-07
A3	10	119	3.3e-04	8	6.5e-07	11	1.4e-05	10	7.0e-06
A4	0	F	3.9e-01	7	8.5e-10	36	1.6e-05	230	8.5e-06
A4	1	F	4.4e-01	9	1.3e-07	24	7.3e-05	16	5.6e-07
A4	10	F	3.2e+00	10	1.3e-05	17	8.3e-05	15	7.5e-05
A5	0	228	8.6e-05	6	2.0e-07	10	5.7e-05	13	1.2e-06
A5	1	268	8.2e-05	6	5.4e-07	10	5.5e-05	13	2.2e-07
A5	10	277	8.1e-05	8	6.7e-07	11	3.3e-05	14	4.9e-06
A6	0	F	4.2e-01	12	7.0e-06	17	1.1e-05	17	1.8e-06
A6	1	F	4.2e-01	11	3.3e-06	15	9.4e-05	16	2.1e-07
A6	10	F	4.1e+02	F	2.4e+00	30	3.1e-05	21	1.2e-06
A7	0	F	7.4e-02	10	3.3e-05	21	4.3e-05	17	2.4e-06
A7	1	F	7.5e-02	13	6.0e-05	20	5.7e-05	17	7.9e-06
A7	10	F	2.1e-01	10	9.5e-05	20	4.5e-05	20	2.2e-07
A8	0	30	5.6e-05	89	9.9e-05	18	6.3e-05	10	8.6e-06
A8	1	36	7.6e-05	494	2.3e-07	18	6.3e-05	10	4.1e-06
A8	10	51	8.6e-05	463	9.9e-05	18	8.0e-05	11	3.9e-05
A9a	0	345	1.0e-04	6	6.6e-05	12	2.7e-05	14	7.6e-05
A9b	0	597	8.9e-05	7	3.3e-05	13	6.0e-05	16	4.2e-05
A10a	0,=	F	3.1e-02	12	1.4e-06	21	2.3e-05	15	2.8e-07
A10b	0,=	F	6.9e-03	57	5.4e-05	21	2.0e-05	F	4.7e-02
A10c	0,=	F	3.8e-01	27	2.6e-05	53	6.4e-05	664	5.4e-05
A10d	0,=	F	1.5e-02	123	5.5e-05	28	7.9e-05	22	6.1e-05
A10e	0,=	F	1.3e-02	119	9.6e-05	28	3.3e-05	48	3.3e-05
Tr1	1	F	8.4e-03	F	3.0e-03	33	6.4e-05	61	4.6e-05
Tr1	10	F	8.4e-03	F	7.0e-03	29	7.2e-05	30	3.0e-05
Tr2	1	F	4.6e-01	F	3.3e-03	43	4.7e-05	943	1.0e-04
Tr2	10	F	5.2e-01	F	2.9e-02	49	6.0e-05	109	9.9e-05
Tr3	1	F	8.5e-01	F	4.8e-02	59	4.5e-05	F	1.3e+01
Tr3	10	F	8.1e-01	F	1.9e-03	F	1.3e-02	33	9.3e-05

Example	x^0	fminunc		semismooth		Interior Point		STRSCNE	
		It.	merit	It.	merit	It.	merit	It.	merit
A11	0	12	5.9e-05	5	1.1e-06	9	1.1e-05	9	3.1e-06
A12	0	23	3.9e-05	1	8.1e-16	7	5.2e-05	8	9.7e-07
A13	0	51	8.2e-05	14	8.7e-07	9	8.7e-05	11	1.9e-05
A14	0.01	9	5.6e-04	7	1.7e-07	10	1.0e-05	7	2.7e-05
A15	0	154	5.7e-05	5	4.3e-08	9	1.6e-05	13	1.3e-06
A16a	10	60	6.8e-05	5	5.1e-05	10	8.9e-05	123	4.9e-07
A16b	10	60	9.9e-05	6	3.2e-05	11	7.7e-05	89	6.6e-06
A16c	10	64	5.0e-05	6	1.3e-05	12	4.1e-05	38	1.1e-05
A16d	10	79	9.8e-05	9	8.2e-07	11	1.7e-05	20	5.9e-06
A17	0	39	3.3e-05	5	4.3e-06	16	1.9e-05	12	1.6e-07
A18	0	95	4.7e-05	9	4.0e-05	15	1.1e-05	20	1.2e-07
A18	1	96	6.3e-05	9	2.0e-07	15	1.1e-05	20	1.5e-07
A18	10	83	6.4e-05	8	3.4e-08	14	1.4e-05	17	5.8e-08
Harker	0	44	1.4e-05	5	1.0e-08	11	9.2e-05	11	2.9e-06
Heu	1	F	4.8e-01	15	7.2e-08	43	2.1e-06	39	4.8e-05
Heu	10	830	9.7e-05	12	7.2e-08	18	1.3e-05	40	4.8e-05
NTF1	0	16	4.1e-05	5	1.6e-06	9	1.3e-05	8	3.5e-06
NTF2	0	16	4.5e-05	6	2.4e-06	9	1.5e-05	9	6.8e-07
Lob	0.1	50	9.8e-05	10	5.9e-05	22	9.6e-06	15	4.6e-05
Spam	1	150	9.8e-05	5	3.5e-06	6	4.0e-05	13	3.5e-06