SIAM J. Optim.
Vol. 21, No. 3, pp. 1082–1108

ON THE SOLUTION OF THE KKT CONDITIONS OF GENERALIZED NASH EQUILIBRIUM PROBLEMS*

AXEL DREVES†, FRANCISCO FACCHINEI‡, CHRISTIAN KANZOW†, AND SIMONE SAGRATELLA‡

Abstract. We consider the solution of generalized Nash equilibrium problems by concatenating the KKT optimality conditions of each player's optimization problem into a single KKT-like system. We then propose two approaches for solving this KKT system. The first approach is rather simple and uses a merit-function/equation-based technique for the solution of the KKT system. The second approach, partially motivated by the shortcomings of the first one, is an interior-point-based method. We show that this second approach has strong theoretical properties and, in particular, that it is possible to establish global convergence under sensible conditions, this probably being the first result of its kind in the literature. We discuss the results of an extensive numerical testing on four KKT-based solution algorithms, showing that the new interior-point method is efficient and very robust.

Key words. generalized Nash equilibrium problem, KKT conditions, merit function, interior-point method, global convergence

AMS subject classifications. 90C30, 90C51, 91A10, 65K05

DOI. 10.1137/100817000

1. Introduction.

We consider the generalized Nash equilibrium problem (GNEP for short) where player ν ($\nu = 1, \ldots, N$) controls $x^\nu \in \mathbb{R}^{n_\nu}$ and tries to solve the optimization problem

$$\min_{x^\nu} \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t. } g^\nu(x^\nu, x^{-\nu}) \le 0 \tag{1.1}$$

with given $\theta_\nu : \mathbb{R}^n \to \mathbb{R}$ and $g^\nu : \mathbb{R}^n \to \mathbb{R}^{m_\nu}$, where "s.t." stands for "subject to." Here, $n := n_1 + \cdots + n_N$ denotes the total number of variables, $m := m_1 + \cdots + m_N$ will be the total number of (inequality) constraints, and $(x^\nu, x^{-\nu})$ is a shorthand notation for the full vector $\mathbf{x} := (x^1, x^2, \ldots, x^N)$, so that $x^{-\nu}$ subsumes all the block vectors x^μ with $\mu \ne \nu$. A vector $\mathbf{x} = (x^1, \ldots, x^N)$ is called feasible for the GNEP if it satisfies the constraints $g^\nu(\mathbf{x}) \le 0$ for all players $\nu = 1, \ldots, N$. A feasible point $\bar{\mathbf{x}}$ is a solution of the GNEP if, for all players $\nu = 1, \ldots, N$, we have

$$\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \le \theta_\nu(x^\nu, \bar{x}^{-\nu}) \quad \forall\, x^\nu : g^\nu(x^\nu, \bar{x}^{-\nu}) \le 0,$$

i.e., if, for all players ν, $\bar{x}^\nu$ is the solution of the νth player's problem when the other players set their variables to $\bar{x}^{-\nu}$.

In this paper we assume that the following blanket assumptions always hold:

A1. $\theta_\nu(\cdot, x^{-\nu})$ and $g_i^\nu(\cdot, x^{-\nu})$ are convex for every $x^{-\nu}$ and for every $\nu = 1, \ldots, N$ and $i = 1, \ldots, m_\nu$;

A2. θ_ν and g^ν are C^2 functions for every $\nu = 1, \ldots, N$.

*Received by the editors December 3, 2010; accepted for publication (in revised form) May 23, 2011; published electronically September 27, 2011. This research was partially supported by the Deutsche Forschungsgemeinschaft (DFG) under grant KA 1296/17-1 and by the MIUR PRIN National Research Program 20079PLLN7 "Nonlinear Optimization, Variational Inequalities and Equilibrium Problems."

http://www.siam.org/journals/siopt/21-3/81700.html

†University of Würzburg, Institute of Mathematics, Emil-Fischer-Str. 30, 97074 Würzburg, Germany (dreves@mathematik.uni-wuerzburg.de, kanzow@mathematik.uni-wuerzburg.de).

‡Sapienza University of Rome, Department of Computer and System Sciences Antonio Ruberti, Via Ariosto 25, 00185 Rome, Italy (facchinei@dis.uniroma1.it, sagratella@dis.uniroma1.it).

This is a very general form of a GNEP, and finding a solution of such a GNEP is a very hard problem; see [15], [20] for a detailed discussion. In fact, the solution of a GNEP in this general form is still little investigated. Due to its daunting difficulty, only very few results are available for the solution of a GNEP at the level of generality described above; see [10], [14], [17], [23], [29], [30], [31] for some different approaches. Some subclasses, in particular jointly convex Nash equilibrium problems (where $g^1 = g^2 = \cdots = g^N$ are the same convex functions, defining the same joint constraints for all players) and pure Nash equilibrium problems (where g^ν depends on x^ν alone for all $\nu = 1, \ldots, N$), have been more widely investigated, and some reasonably efficient methods for the solution of these latter problems have been proposed; see [15], [21].

The main aim of this paper is to study and give convergence results based on the use of the KKT conditions of the general GNEP (1.1) (see next section). This has been done previously in [14], [29], where the authors were mainly interested in the local convergence behavior of suitable Newton-type methods. In particular, it is shown in [14] that one has to expect difficulties in solving the KKT system due to some singularity problems; hence local fast convergence cannot be obtained in many standard situations. Apart from these papers, the KKT approach is also part of the folklore in the engineering world, but in spite of this, there is still a lack of any serious analysis dealing with the solution of this peculiar KKT-like system.

Here we fill this gap and provide sound results establishing the viability of the KKT approach, both at the theoretical and numerical level, and with a special emphasis on the global behavior of the methods. In particular, we provide conditions under which the global convergence is guaranteed. These conditions are reasonable and, to the best of our knowledge, they are the first set of explicit conditions on a general GNEP under which global convergence can be established. Regarding global convergence, we are only aware of two other papers where this issue has been positively handled. One is [32], where, however, only a problem arising from a very specific telecommunication application is considered. The techniques and methods used in [32] are peculiar to the problem considered there, and their generalization to a wider class of problems seems very difficult. Global convergence results are also discussed in [18], where a penalty technique for the solution of a general GNEP is proposed. Although the results in [18] are of great interest, global convergence for genuine GNEPs can only be established under restrictive conditions. These conditions depend on the unknown value of a (penalty) parameter, and so their application appears to be problematic in practice.

In this paper, we consider two different approaches and introduce two rather distinct classes of algorithms for the solution of the GNEP KKT conditions. In the first approach, we use an equation reformulation of the KKT conditions and a corresponding merit function, while the second approach is based on interior-point ideas. Although both approaches have been proposed and successfully used for the solution of the KKT systems arising in the solution of optimization or variational inequality (VI) problems, we will see, somewhat surprisingly, that the interior-point approach seems definitely superior in our setting.

The paper is organized in the following way: We begin with the formulation of the KKT conditions of a GNEP in a compact form. Then in section 3 we consider the optimization reformulation of the KKT system and give conditions guaranteeing stationary points to be solutions of the GNEP and further show a coercivity result. In order to give a concrete algorithm for the solution of GNEPs and to get a more problem-tailored approach, we introduce in section 4 an interior-point method with its

convergence theory. In section 5 we discuss the numerical behavior of the different approaches on several test problems.

A few words regarding our notation: $\mathbb{R}^n$ denotes the n-dimensional Euclidean vector space, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$ denote the corresponding subsets consisting of all vectors whose components are nonnegative and positive, respectively. Given a differentiable mapping $H:\mathbb{R}^n \to \mathbb{R}^m$, we denote by $JH(z)$ the Jacobian of H at a given point $z \in \mathbb{R}^n$, whereas $\nabla H(z)$ is the transposed Jacobian. If the set of variables z can be split into two (or more) groups, say, $z = (x, y)$, then $J_x H(x, y)$ denotes the Jacobian of H at (x, y) with respect to x alone, and the transposed matrix is again $\nabla_x H(x, y)$. Given a nonsingular matrix $M \in \mathbb{R}^{n\times n}$, we write M^{-T} for the inverse of M^T, which is identical to the transposed of M^{-1}. Furthermore, $\operatorname{diag}(w)$ denotes the diagonal matrix of appropriate dimension with the vector w on its diagonal. A matrix $M \in \mathbb{R}^{n\times n}$ is called a P_0-matrix if $\det(M_{\alpha\alpha}) \geq 0$ for all $\alpha \subseteq \{1, 2, \ldots, n\}$. Note that the class of P_0-matrices strictly includes the positive semidefinite matrices; see [6] for more details. The symbol $\|\cdot\|$ always denotes the Euclidean vector norm or the corresponding induced matrix norm. Sometimes, we also write explicitly $\|\cdot\|_2$ for this norm in order to avoid any confusion. Finally, the symbol $\mathbb{B}(x, r)$ denotes the open (Euclidean) ball centered in x and with radius $r > 0$, whereas $\operatorname{cl}\mathbb{B}(x, r)$ is the corresponding closed ball.

2. The KKT conditions. Let $\bar{\mathbf{x}}$ be a solution of the GNEP (1.1). Assuming any standard constraint qualification holds, it is well known that the following KKT conditions will be satisfied for every player $\nu = 1, \ldots, N$:

$$\begin{aligned} &\nabla_{x^\nu}\theta_\nu(\bar{x}^\nu, \bar{x}^{-\nu}) + \sum_{i=1}^{m_\nu} \lambda_i^\nu \nabla_{x^\nu} g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0, \\ &\lambda_i^\nu \geq 0, \quad g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) \leq 0, \quad \lambda_i^\nu g_i^\nu(\bar{x}^\nu, \bar{x}^{-\nu}) = 0 \quad \forall\ i = 1, \ldots, m_\nu, \end{aligned} \tag{2.1}$$

where λ^ν is the vector of Lagrange multipliers of player ν. Vice versa, recalling that the player's problems are convex (see A1), we have that if a point $\bar{\mathbf{x}}$ satisfies, together with a suitable vector of multipliers $\boldsymbol{\lambda} := (\lambda^1, \lambda^2, \ldots, \lambda^N)$, the KKT conditions (2.1) for every ν, then $\bar{\mathbf{x}}$ is a solution of the GNEP. It then seems rather natural to try to solve the GNEP by solving the system obtained by concatenating the N systems (2.1). In order to use a more compact notation, we introduce some further definitions.

We denote by $L^\nu(\mathbf{x}, \lambda^\nu) := \theta_\nu(x^\nu, x^{-\nu}) + \sum_{i=1}^{m_\nu} \lambda_i^\nu g_i^\nu(x^\nu, x^{-\nu})$ the Lagrangian of player ν. If we set $\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) := (\nabla_{\mathbf{x}^\nu} L^\nu(\mathbf{x}, \lambda^\nu))_{\nu=1}^N$ and $\mathbf{g}(\mathbf{x}) := (g^\nu(\mathbf{x}))_{\nu=1}^N$, the concatenated KKT system can be written as

$$\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\lambda} \geq 0, \quad \mathbf{g}(\mathbf{x}) \leq 0, \quad \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) = 0. \tag{2.2}$$

There is a large amount of literature on reformulating the KKT conditions of an optimization problem or of a VI as a (constrained) system of equations or as a (constrained) optimization problem. These reformulations are the basis for many efficient algorithms for the solution of these problems; see [19]. However, probably due to the difficulty of the analysis, to date 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. The main aim of this paper is therefore to derive sound theoretical results related to system (2.2) and to define some new solution methods. More specifically, we will analyze a merit function approach and an interior-point method for the solution of the KKT system (2.2). 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 and perform extensive numerical experiments.

3. 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 lead to very useful developments (see [12], [19]) is to reduce it to a system of equations through the use of a *complementarity function*. More specifically, let $\phi:\mathbb{R}^2 \to \mathbb{R}$ be any function such that $\phi(a,b)=0$ if and only if $a \geq 0$, $b \geq 0$, and $ab=0$. Then it is immediate to see that the concatenated KKT system can be rewritten as

$$\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})=0, \qquad \boldsymbol{\Phi}(\mathbf{x},\boldsymbol{\lambda})=0,$$

where

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

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 development of local Newton methods discussed in [14]. 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. Therefore, we use only the Fischer–Burmeister function in this paper; i.e., ϕ always denotes this complementarity function.

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

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

Note that $\boldsymbol{\Phi}$ (using the Fischer–Burmeister function) is not differentiable, in general, because the Fischer–Burmeister complementarity function is obviously nondifferentiable at (0,0). However, it is by now very well known that Θ is once (though not twice) continuously differentiable. 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 (a) conditions guaranteeing that unconstrained stationary points of Θ are global solutions, and (b) conditions under which Θ can be shown to

be coercive. Once this has been done, one can safely attempt to solve the KKT system (2.2) by performing the unconstrained minimization of Θ. Unfortunately, while in the optimization and in the VI fields "reasonable" conditions guaranteeing the above-mentioned results can be identified (see [19]), the situation becomes much more involved in the case of system (2.2).

3.1. Stationarity conditions. For the sake of notational simplicity, it is useful to introduce the matrix

$$E(\mathbf{x}) := \begin{pmatrix} \nabla_{x^1} g^1(\mathbf{x}) & & 0 \\ & \ddots & \\ 0 & & \nabla_{x^N} g^N(\mathbf{x}) \end{pmatrix} \quad \text{with} \quad \nabla_{x^\nu} g^\nu(\mathbf{x}) \in \mathbb{R}^{n_\nu \times m_\nu}. \tag{3.1}$$

Using the chain rule from [5] and some standard calculations, we obtain that the gradient of Θ is given by

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

where the matrices $D_{\boldsymbol{\lambda}}$ and $D_{\mathbf{g}}$ are $m \times m$ diagonal matrices

$$D_{\boldsymbol{\lambda}}(\mathbf{x},\boldsymbol{\lambda}) := \operatorname{diag}(a^1(\mathbf{x},\lambda^1), \ldots, a^N(\mathbf{x},\lambda^N)),$$
$$D_{\mathbf{g}}(\mathbf{x},\boldsymbol{\lambda}) := \operatorname{diag}(b^1(\mathbf{x},\lambda^1), \ldots, b^N(\mathbf{x},\lambda^N)),$$

with vectors $a^\nu(\mathbf{x},\lambda^\nu), b^\nu(\mathbf{x},\lambda^\nu) \in \mathbb{R}^{m_\nu}$ whose entries are given by

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

for all $i = 1, \ldots, m_\nu$ and for all $\nu = 1, \ldots, N$. Note that, 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 canceled by corresponding zero entries in $\boldsymbol{\Phi}(\mathbf{x},\boldsymbol{\lambda})$.

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 3.1. *Let $(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) \in \mathbb{R}^n \times \mathbb{R}^m$ be a stationary point of Θ, and suppose that $J_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ is nonsingular and*

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

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

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

$$\nabla_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) - \nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}) D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = 0, \tag{3.2}$$

$$E(\bar{\mathbf{x}})^T\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) + D_{\boldsymbol{\lambda}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = 0. \tag{3.3}$$

By the nonsingularity of $\nabla_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$, we obtain from (3.2) that

$$\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = \nabla_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})^{-1}\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}})D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}), \tag{3.4}$$

and substituting this into (3.3), we get

$$\begin{aligned} &E(\bar{\mathbf{x}})^T\nabla_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})^{-1}\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}})D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) + D_{\boldsymbol{\lambda}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) \\ &\quad = [M(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})^T D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) + D_{\lambda}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})]\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = 0. \end{aligned} \tag{3.5}$$

Now let us recall that $a_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu)$, $b_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu)$ are nonpositive with $(a_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu), b_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu)) \neq (0,0)$ for all i, ν and that $a_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu) = 0$ or $b_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu) = 0$ can happen only if we have $\boldsymbol{\phi}(\bar{\lambda}_i^\nu, -\mathbf{g}_i^\nu(\bar{\mathbf{x}})) = 0$. Since, in the previous equations, both elements $a_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu)$ and $b_i^\nu(\bar{\mathbf{x}},\bar{\lambda}_i^\nu)$ are always postmultiplied by $\boldsymbol{\phi}(\bar{\lambda}_i^\nu, -\mathbf{g}_i^\nu(\bar{\mathbf{x}})) = 0$, we do not change these equations if we assume without loss of generality that both diagonal matrices $D_{\boldsymbol{\lambda}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ and $D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ are negative definite. Since $M(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ is assumed to be a P_0-matrix, it is then easy to see that $(M(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})^T D_{\mathbf{g}}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) + D_{\lambda}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}))$ is nonsingular. Hence by (3.5) it follows that $\boldsymbol{\Phi}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = 0$. This immediately implies that $\mathbf{F}(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}}) = 0$ by (3.4), and so we obtain the thesis. □

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

Example 3.2. 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

$$\begin{aligned} &\text{Player 1: } \min_{x_1} \frac{1}{2}(x_1-1)^2 - x_1x_2 \quad \text{s.t. } x_1+x_2+x_3 \le 1, \\ &\text{Player 2: } \min_{x_2} \frac{1}{2}(x_2-1)^2 + x_1x_2 \quad \text{s.t. } x_1+x_2+x_3 \le 1, \\ &\text{Player 3: } \min_{x_3} \frac{1}{2}(x_3-1)^2 \quad \text{s.t. } 0 \le x_3 \le x_1+x_2. \end{aligned}$$

Then we have

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

which is nonsingular, and we get

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

An elementary calculation shows that $\det(M(\mathbf{x},\boldsymbol{\lambda})_{\alpha\alpha}) \ge 0$ holds for all $\alpha \subseteq \{1,2,3,4\}$; hence $M(\mathbf{x},\boldsymbol{\lambda})$ is a P_0-matrix. Consequently, Theorem 3.1 can be applied and guarantees that every stationary point of $\boldsymbol{\Theta}$ is a solution of the GNEP.

This example also indicates a limitation of Theorem 3.1 if the constraints are not linear. In this case, the nonsingularity of $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ and the P_0 property of $M(\mathbf{x},\boldsymbol{\lambda})$ must hold even for negative values of $\boldsymbol{\lambda}$, and it is apparent that this won't be the case in general. In fact, $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ will contain block-diagonal terms of the type $\lambda_i^\nu \nabla^2_{x^\nu x^\nu} g_i^\nu(\mathbf{x})$, which will be negative definite if λ_i^ν is negative, and can lead to a singular matrix $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$.

Example 3.3. Consider a two-player game where each player controls a single variable, given by

$$\text{Player 1: } \min_{x_1} \frac{1}{2}x_1^2 + \frac{32}{5}x_1 \quad \text{s.t. } \frac{1}{6}x_1^2 + x_2 - \frac{5}{2} \le 0,$$
$$\text{Player 2: } \min_{x_2} \frac{1}{2}x_2^2 + x_1x_2 - \frac{4}{5}x_2 \quad \text{s.t. } x_2 \in \mathbb{R}.$$

Then we have

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

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

$$\nabla\Theta(\mathbf{x},\lambda) = \begin{bmatrix} 1+\frac{1}{3}\lambda & 1 & -\frac{1}{3}x_1 b(\mathbf{x},\lambda) \\ 0 & 1 & -b(\mathbf{x},\lambda) \\ \frac{1}{3}x_1 & 0 & a(\mathbf{x},\lambda) \end{bmatrix} \begin{pmatrix} x_1 + \frac{32}{5} + \frac{1}{3}x_1\lambda \\ x_2 + x_1 - \frac{4}{5} \\ \phi(\lambda, -\frac{1}{6}x_1^2 - x_2 + \frac{5}{2}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

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

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, i.e., by solving the problem

$$\min \Theta(\mathbf{x},\boldsymbol{\lambda}) \quad \text{s.t.} \quad \boldsymbol{\lambda} \ge 0. \tag{3.6}$$

This leads to successful results in the optimization/VI case; see [13], [19]. Unfortunately, also this approach leads to problems in our game setting. This is illustrated by the following example.

Example 3.4. Consider an apparently well-behaved game where each player controls a single variable, and the players' problems are given by

$$\text{Player 1: } \min_{x_1} x_1 \quad \text{s.t. } x_1^2 + x_2 \le 1, \qquad \text{Player 2: } \min_{x_2} \frac{1}{2}x_2^2 \quad \text{s.t. } x_2 \in \mathbb{R}.$$

It is not difficult to show that the point $(-1,0)$ together with $\lambda = \frac{1}{2}$ is the only generalized Nash equilibrium. However, it is easy to see that the point $(0,0)$ together with $\lambda = 0$ is an unconstrained stationary point of Θ and also a stationary point of the constrained problem (3.6).

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

THEOREM 3.5. *Consider a NEP. Let* $(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) \in \mathbb{R}^n \times \mathbb{R}^m$ *be a stationary point of* Θ*, suppose that* $J_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ *is positive semidefinite, and it holds that*

$$d^T J_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})d > 0 \quad \forall\ d \neq 0\colon\ E(\bar{\mathbf{x}})^T d = 0.$$

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

Proof. In a NEP we have $\nabla_x \mathbf{g}(\mathbf{x}) = E(\mathbf{x})$. Taking the two stationarity conditions (3.2) and (3.3), multiplying the first with $\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})^T$, and substituting the second one in the resulting expression, we get

$$\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})^T \nabla_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) + \boldsymbol{\Phi}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})^T D_{\lambda}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) D_{\mathbf{g}}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})\boldsymbol{\Phi}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) = 0.$$

By the positive semidefiniteness of $J_{\mathbf{x}}\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ and since we may assume, without loss of generality, that both diagonal matrices $D_{\boldsymbol{\lambda}}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ and $D_{\mathbf{g}}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ have negative entries (cf. the proof of Theorem 3.1), we get $\boldsymbol{\Phi}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) = 0$. Then (3.2) and (3.3) together with $d^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})d > 0\ \forall\ d \neq 0\colon\ E(\bar{\mathbf{x}})^T d = 0$ imply that $\mathbf{F}(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) = 0$, which completes the proof. ∎

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 $\{d | E(\bar{\mathbf{x}})^T d = 0\}$ is (in general) much smaller than the usual tangent cone. To this end, note that this tangent cone may be rewritten as $\mathcal{T}(x) = \{d = (d^1, \ldots, d^N) | \nabla g_i^{\nu}(x^{\nu})^T d^{\nu} = 0\ \forall\ i = 1, \ldots, m_{\nu}\ \forall\ \nu = 1, \ldots, N\}$; i.e., 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 3.5 is significantly weaker than the usual one.

3.2. 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 that remains to be answered, at least from a theoretical point of view, is under which assumptions a sequence $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$, 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 our merit function Θ shows that this has mainly to do with the behavior of the sequence $\{\boldsymbol{\lambda}^k\}$ which, in particular, might be unbounded. On the other hand, it is possible to show that the sequence $\{\mathbf{x}^k\}$ remains bounded under very reasonable assumptions.

To this end, consider a GNEP that is defined via the optimization problems

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

with functions $h_j^{\nu}\colon \mathbb{R}^{n_{\nu}} \to \mathbb{R}$ and $g_i^{\nu}\colon \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \ldots, p_{\nu}$, $i = p_{\nu} + 1, \ldots, m_{\nu}$, that are assumed to be convex in x^{ν}; i.e., here we distinguish, for each player $\nu = 1, \ldots, N$, 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. We then define the set $X_0 := \{x \in \mathbb{R}^n | h^{\nu}(x^{\nu}) \leq 0\ \forall\ \nu = 1, \ldots, N\}$. The set X_0 is closed and convex since the contraints h^{ν} are convex by assumption. If we assume boundedness of the set X_0, we can show boundedness of the x-iterates.

PROPOSITION 3.6. *Suppose $h_j^\nu:\mathbb{R}^{n_\nu}\to\mathbb{R}$ 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 $\Theta(x^k,\lambda^k)\le\Theta(x^0,\lambda^0)$ 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 $h_j^\nu(x^\nu)\le\gamma\ \forall\ j=1,\dots,p_\nu\ \forall\ \nu=1,\dots,N\ \Leftrightarrow\ h_{\max}(x)\le\gamma$ for any given $\gamma\in\mathbb{R}$. In particular, we can rewrite the set X_0 as $X_0=\{x\in\mathbb{R}^n|h_{\max}(x)\le 0\}$. Since $h_{\max}$ is a single convex function, it follows from our assumptions on X_0 together with [33, Corollary 8.7.1] that the level sets

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

are also bounded for any $\gamma\in\mathbb{R}$. Now assume that the sequence $\{x^k\}$ is unbounded, say, $\{\|x^k\|\}\to\infty$. 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 are 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

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

and thus we obtain $\Theta(x^{\ell(k)},\lambda^{\ell(k)})\ge\frac{1}{2}\phi^2((\lambda^{\ell(k)})_j^\nu,-h_j^\nu(x^{\ell(k)}))>\frac{1}{2}k^2$. Hence we have $\Theta(x^{\ell(k)},\lambda^{\ell(k)})\to\infty$ for $k\to_K\infty$, contradicting the assumption that $\Theta(x^k,\lambda^k)\le\Theta(x^0,\lambda^0)$ for all $k\in\mathbb{N}$. ∎

In spite of its theoretical interest, the above proposition is of limited practical use, since an unbounded multiplier typically produces a failure of any suitable method for the minimization of Θ. In the next section we will see that, when using an interior-point method, we will be able to guarantee the boundedness of all variables involved. We also note that this is, in principle, possible for the merit function approach if one takes another ϕ-function; see Remark 4.12 for more details.

4. Interior-point method. The results in the previous section are certainly valuable, but they also have their limitations which, apparently, are essentially due to the "$\boldsymbol{\lambda}$ part" of the variables. We already discussed that a straightforward treatment of the sign constraints for the multipliers is not likely to be helpful in the merit function approach. One therefore has to look for suitable alternatives, and this leads naturally to the consideration of an interior-point approach to the solution of the GNEP KKT system. Furthermore, and beside the considerations above, interior-point methods are well known to be efficient methods for solving KKT systems arising from optimization or VI problems. We therefore devote this section to the analysis of an interior-point method for the solution of the KKT system (2.2). To this end, we formulate this system as a constrained nonlinear system of equations (CE or, more precisely, CE(H,Z)) of the form

$$H(z)=0,\qquad z\in Z \tag{4.1}$$

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

We introduce slack variables $\mathbf{w} := (w^\nu)_{\nu=1}^N$, where $w^\nu \in \mathbb{R}^{m_\nu}$, and set $\boldsymbol{\lambda} \circ \mathbf{w} := (\lambda_1^1 w_1^1, \ldots, \lambda_{m_N}^N w_{m_N}^N)^T$. Then we define

$$H(z) := H(\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}) := \begin{pmatrix} \mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{x}) + \mathbf{w} \\ \boldsymbol{\lambda} \circ \mathbf{w} \end{pmatrix} \tag{4.2}$$

and

$$Z := \{z = (\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}) | \mathbf{x} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \mathbb{R}^m_+, \mathbf{w} \in \mathbb{R}^m_+\}. \tag{4.3}$$

It is immediate to verify that a point $(\mathbf{x}, \boldsymbol{\lambda})$ solves the KKT system (2.2) if and only if this point, together with a suitable $\mathbf{w}$, solves the constrained equation defined by (4.2) and (4.3).

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 $(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)$ with $\boldsymbol{\lambda}^k > 0$ and $\mathbf{w}^k > 0$ for every k.

The particular method that we base our analysis on is the potential reduction method from [26], also discussed in detail in [19]. 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 (4.2) and (4.3).

To this end, we define the set $S := \mathbb{R}^n \times \mathbb{R}^{2m}_+$ as well as a potential function on 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.

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

$$Z_I := H^{-1}(\operatorname{int} S) \cap \operatorname{int} Z \text{ by setting } \psi(z) := p(H(z)) \quad \text{for } z \in Z_I.$$

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

$$H(z^k) + JH(z^k) d^k = \sigma_k \frac{a^T H(z^k)}{\|a\|^2} a \tag{4.4}$$

(the constant vector a is defined below). Once this direction has been calculated, a line search is performed by using the potential function ψ. The version we describe and analyze below is a variant where we allow the possibility of an inaccurate solution of system (4.4).

Algorithm 4.1 (Inexact Potential Reduction Method for GNEPs).

(S.0) Choose $z^0 \in Z_I$, β, $\gamma \in (0,1)$, 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

$$\left\| H(z^k) + JH(z^k)d^k - \sigma_k \frac{a^T H(z^k)}{\|a\|^2}\, a \right\| \leq \eta_k \|H(z^k)\| \quad \text{and} \tag{4.5}$$

$$\nabla\psi(z^k)^T d^k < 0. \tag{4.6}$$

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

$$z^k + t_k d^k \in Z_I \quad \text{and} \tag{4.7}$$

$$\psi(z^k + t_k d^k) \leq \psi(z^k) + \gamma t_k \nabla\psi(z^k)^T d^k. \tag{4.8}$$

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

Remark 4.2.

(a) By construction, all iterates z^k generated by Algorithm 4.1 belong to the set Z_I; hence we have $z^k \in \operatorname{int} Z$ and $H(z^k) \in \operatorname{int} S$ for all $k \in \mathbb{N}$.

(b) If $JH(z^k)$ is a nonsingular $(n+2m) \times (n+2m)$ matrix for all k, it follows that the linear system of equations (4.4) always has an exact solution $\hat{d}^k$. In particular, this exact solution satisfies the inexactness requirement from (4.5) for an arbitrary number $\eta_k \geq 0$. Furthermore, this exact solution also satisfies the descent property $\nabla\psi(z^k)^T \hat{d}^k < 0$; see [19]. It therefore follows that one can always find a vector d^k satisfying the two requirements (4.5) and (4.6); i.e., (S.2) is well defined.

(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 (4.7) holds for all sufficiently small stepsizes t_k. Furthermore, the Armijo line search from (4.8) is eventually satisfied since d^k is a descent direction of the potential function ψ in view of the construction in (S.2); cf. (4.6). In particular, this means that (S.3) is also well defined.

The following is the main convergence result for Algorithm 4.1, where, implicitly, we assume that Algorithm 4.1 does not terminate within a finite number of iterations with a solution of the constrained nonlinear system CE(H, Z).

Theorem 4.3. *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* 4.1 *satisfy the conditions*

$$\limsup_{k\to\infty} \sigma_k < \bar{\sigma} \quad \text{and} \quad \lim_{k\to\infty} \eta_k = 0. \tag{4.9}$$

Let $\{z^k\}$ be any sequence generated by Algorithm 4.1. *Then*

(a) *The sequence $\{H(z^k)\}$ is bounded.*

(b) *Any accumulation point of $\{z^k\}$ is a solution of* (4.1).

Proof. We first note that our assumptions together with Remark 4.2(b), (c) guarantee that Algorithm 4.1 is at least well defined. Throughout this proof, we use the abbreviation $u^k := H(z^k)$ for all $k \in \mathbb{N}$.

Suppose that $\{u^k\}$ is unbounded. Subsequencing if necessary, we may assume without loss of generality that $\lim_{k\to\infty}\|u^k\| = \infty$. Since $\{u^k\} \subseteq \mathrm{int}S$ in view of Remark 4.2(a), an elementary calculation then shows that $\lim_{k\to\infty} p(u^k) = \infty$. 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 (4.8) that $p(u^k)=p(H(z^k))=\psi(z^k)<\psi(z^{k-1})< \cdots <\psi(z^0)$, and this contradiction completes the proof of part (a).

Let z^∞ be an accumulation point of the sequence $\{z^k\}$, and let $\{z^k\}_K$ be a corresponding subsequence converging to z^∞. Since $z^k \in \mathrm{int}Z$ for all $k \in \mathbb{N}$ (cf. Remark 4.2(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 (4.9), we may assume without loss of generality that $\lim_{k\in K}\sigma_k = \sigma_\infty$ for some $\sigma_\infty \in [0, \bar{\sigma})$ and $\lim_{k\in K} u^k = u^\infty \neq 0$. 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 $\Lambda(\varepsilon, \delta) := \{u \in \mathrm{int}S | p(u) \leq \delta, \|u\| \geq \varepsilon\}$, which is a compact set. Hence we have $u^\infty = H(z^\infty) \in \Lambda(\varepsilon, \delta) \subseteq \mathrm{int}S$. Consequently, we have $z^\infty \in H^{-1}(\mathrm{int}S) \cap Z$. However, since $H^{-1}(\mathrm{int}S) \cap \mathrm{bd}(Z) \subseteq \mathrm{int}Z \cap \mathrm{bd}(Z) = \emptyset$, it therefore follows that z^∞ belongs to the set $H^{-1}(\mathrm{int}S) \cap \mathrm{int}Z = Z_I$.

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

$$r^k := H(z^k) + JH(z^k)d^k - \sigma_k \frac{a^T H(z^k)}{\|a\|^2}\, a \quad \forall\ k \in \mathbb{N}. \tag{4.10}$$

Then the inexactness requirement (4.5) can be written down as

$$\|r^k\| \leq \eta_k \|H(z^k)\| \quad \forall\ k \in \mathbb{N}. \tag{4.11}$$

Since the Jacobian $JH(z^k)$ is nonsingular at $z^k \in Z_I$, we obtain from (4.10) that

$$d^k = JH(z^k)^{-1}\left[r^k - H(z^k) + \sigma_k \frac{a^T H(z^k)}{\|a\|^2}\, a\right] \quad \forall\ k \in \mathbb{N}. \tag{4.12}$$

Since $\{z^k\}_{k\in K} \to z^\infty$, the continuity of the Jacobian implies that $\{JH(z^k)\}_{k\in K} \to JH(z^\infty)$. 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 (4.12) 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 (4.9) 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 $\eta_k \to 0$, it follows from (4.11) that $r^k \to 0$. On the other hand, using the definition of the residuum r^k and taking the limit $k \to \infty$ 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 $z^\infty \in Z_I$ and $H(z^\infty) = u^\infty \neq 0$ by assumption, we obtain that $\nabla\psi(z^\infty)^T d^\infty < 0$; cf. Remark 4.2(b). 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 (4.8) 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 (4.8) once more, we have $\psi(z^{k+1}) - \psi(z^k) \le \gamma t_k \nabla\psi(z^k)^T d^k < 0$ for all $k \in \mathbb{N}$. Since the left-hand side converges to zero, we obtain $\lim_{k\to\infty} t_k \nabla\psi(z^k)^T d^k = 0$. 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 $k \in \mathbb{N}$. Since $\lim_{k\in K} t_k = 0$, 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 (4.8) 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 $k \to \infty$ on the subset K, the continuous differentiability of the potential function ψ on the set Z_I then gives $\nabla\psi(z^\infty)^T d^\infty \ge \gamma \nabla\psi(z^\infty)^T d^\infty$. Since $\nabla\psi(z^\infty)^T d^\infty < 0$, this is only possible if $\gamma \ge 1$, a contradiction to the choice of $\gamma \in (0, 1)$. Consequently, we have $0 = u^\infty = H(z^\infty)$; i.e., z^∞ is a solution of the constrained system of nonlinear equations (4.1). □

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. [14]. Hence, also these Newton methods are not quadratically convergent, and the rate of convergence may actually slow down dramatically, which is in contrast to our method; see section 5 for a numerical comparison.

4.1. Nonsingularity conditions. The critical issue in applying Theorem 4.3 is establishing the nonsingularity of JH. This section is devoted to this issue. We will see that while the condition we will use in order to establish the nonsingularity of JH are similar to those obtained in the equation reformulation approach, they need only to be valid for positive values of λ.

The structure of $JH(z)$ is the following:

$$JH(z) := \begin{pmatrix} J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) & E(\mathbf{x}) & 0 \\ J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) & 0 & I \\ 0 & \operatorname{diag}(\mathbf{w}) & \operatorname{diag}(\boldsymbol{\lambda}) \end{pmatrix} \tag{4.13}$$

with E defined in (3.1). In order to analyze the nonsingularity of this matrix, we first introduce the following terminology; cf. [19].

DEFINITION 4.4. *A matrix* $Q \coloneqq [M_1, M_2, M_3]$ *is said to have the mixed* P_0*-property if* M_3 *has full column rank and*

$$\left.\begin{array}{c} M_1u + M_2v + M_3s = 0 \\ (u, v) \neq 0 \end{array}\right\} \Rightarrow u_iv_i \geq 0 \quad \text{for some } i \text{ such that } |u_i| + |v_i| > 0.$$

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 $[M - I]$ (with a vacuous M_3-part) has the mixed P_0-property; i.e., Definition 4.4 generalizes the standard notion of a P_0-matrix. A useful characterization of the mixed P_0-property is given in [19, Lemma 11.4.3] and is restated in the following result.

LEMMA 4.5. *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* $Q \coloneqq [M_1, M_2, M_3]$ *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*

$$M \coloneqq \begin{bmatrix} D_1 & D_2 & 0 \\ M_1 & M_2 & M_3 \end{bmatrix}$$

is nonsingular.

Note that this lemma is immediately applicable to (4.13) and gives a necessary and sufficient condition for the nonsingularity of JH when $\boldsymbol{\lambda} > 0$ and $\mathbf{w} > 0$. However, the mixed P_0-property is difficult to interprete 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 4.5; 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 4.5.

The following theorem gives a first nonsingularity result.

THEOREM 4.6. *Let* $z = (\mathbf{x}, \boldsymbol{\lambda}, \mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++}$ *be given such that* $J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})$ *is nonsingular and*

$$M(\mathbf{x}, \boldsymbol{\lambda}) \coloneqq J_{\mathbf{x}}\mathbf{g}(\mathbf{x})J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})^{-1}E(\mathbf{x}) \tag{4.14}$$

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:

$$J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})q^{(1)} + E(\mathbf{x})q^{(2)} = 0, \tag{4.15}$$

$$J_{\mathbf{x}}\mathbf{g}(\mathbf{x})q^{(1)} + q^{(3)} = 0, \tag{4.16}$$

$$\operatorname{diag}(\mathbf{w})q^{(2)} + \operatorname{diag}(\boldsymbol{\lambda})q^{(3)} = 0. \tag{4.17}$$

Since $J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})$ is nonsingular by assumption, (4.15) yields $q^{(1)} = -J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})^{-1}\ E(\mathbf{x})q^{(2)}$. Hence we obtain $q^{(3)} = -J_{\mathbf{x}}\mathbf{g}(\mathbf{x})q^{(1)} = J_{\mathbf{x}}\mathbf{g}(\mathbf{x})J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})^{-1}E(\mathbf{x})q^{(2)} = M(\mathbf{x}, \boldsymbol{\lambda})q^{(2)}$ from (4.16) and the definition of $M(\mathbf{x}, \boldsymbol{\lambda})$. Substituting this expression into (4.17) gives $[\operatorname{diag}(\mathbf{w}) + \operatorname{diag}(\boldsymbol{\lambda})M(\mathbf{x}, \boldsymbol{\lambda})]q^{(2)} = 0$. Since $M(\mathbf{x}, \boldsymbol{\lambda})$ is a P_0-matrix by assumption and $\mathbf{w}$, $\boldsymbol{\lambda} > 0$, it follows that $[\operatorname{diag}(\mathbf{w}) + \operatorname{diag}(\boldsymbol{\lambda})M(\mathbf{x}, \boldsymbol{\lambda})]$ is nonsingular, and hence $q^{(2)} = 0$. This, in turn, implies that $q^{(1)} = 0$ and $q^{(3)} = 0$. Consequently, $JH(z)$ is nonsingular. ∎

Note that this condition is identical to the assumptions for the stationarity condition in Theorem 3.1. 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 3. To illustrate this point, let us consider once again Example 3.4. It is now easy to see that this example satisfies the conditions of Theorem 4.6:

$$J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) = \begin{pmatrix} 2\lambda & 0 \\ 0 & 1 \end{pmatrix}$$

is nonsingular for all $\lambda > 0$, $M(\mathbf{x},\boldsymbol{\lambda}) = \frac{2x_1^2}{\lambda} \geq 0$ 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 4.6; nevertheless it is interesting because it gives a quantitative insight into what is necessary to guarantee the nonsingularity of JH. We use the notation $\mathrm{eig}_{\min}(A)$ for the smallest eigenvalue of a symmetric matrix A.

THEOREM 4.7. *Let $z = (\mathbf{x},\boldsymbol{\lambda},\mathbf{w}) \in \mathbb{R}^n \times \mathbb{R}^m_{++} \times \mathbb{R}^m_{++}$ be given such that $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ is nonsingular and*

$$\begin{aligned} &\mathrm{eig}_{\min}\left(\frac{1}{2}E(\mathbf{x})^T(J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} + J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-T})E(\mathbf{x})\right) \\ &\qquad \geq \|J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) - E(\mathbf{x})^T\|_2 \|J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1}\|_2 \|E(\mathbf{x})\|_2. \end{aligned}$$

Then the Jacobian $JH(z)$ is nonsingular.

Proof. For all $u \in \mathbb{R}^m$ we have

$$\begin{aligned} u^T E(\mathbf{x})^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} E(\mathbf{x}) u &= \frac{1}{2} u^T (E(\mathbf{x})^T (J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} + J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-T}) E(\mathbf{x})) u \\ &\geq \mathrm{eig}_{\min}\left(\frac{1}{2}E(\mathbf{x})^T(J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} + J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-T})E(\mathbf{x})\right)\|u\|_2^2 \\ &\geq \|J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) - E(\mathbf{x})^T\|_2 \|J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1}\|_2 \|E(\mathbf{x})\|_2 \|u\|_2^2 \\ &\geq |u^T (J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) - E(\mathbf{x})^T) J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} E(\mathbf{x}) u| \\ &\geq -u^T (J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) - E(\mathbf{x})^T) J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} E(\mathbf{x}) u. \end{aligned}$$

Using the matrix $M(\mathbf{x},\boldsymbol{\lambda})$ from (4.14), this implies that $u^T M(\mathbf{x},\boldsymbol{\lambda}) u = u^T J_{\mathbf{x}}\mathbf{g}(\mathbf{x})$ $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})^{-1} E(\mathbf{x}) u \geq 0$ for all $u \in \mathbb{R}^m$. Therefore $M(\mathbf{x},\boldsymbol{\lambda})$ is positive semidefinite; hence a P_0-matrix, and Theorem 4.6 guarantees nonsingularity of $JH(z)$. ∎

In the case of a NEP, if $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ is positive definite, matrix (4.14) is automatically P_0 (actually, positive semidefinite) since $J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) = E(\mathbf{x})^T$ 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 $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ and still get the nonsingularity of $JH(z)$. In fact, we have the following counterpart of the stationary point condition from Theorem 3.5.

THEOREM 4.8. *Consider a NEP, and let $z = (\mathbf{x},\boldsymbol{\lambda},w) \in \mathbb{R}^n \times \mathbb{R}^m_{++}$ be given such that $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ is positive semidefinite, and it holds that*

$$d^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) d > 0 \quad \forall\ d \neq 0\colon E(\bar{\mathbf{x}})^T d = 0.$$

Then the Jacobian $JH(z)$ is nonsingular.

Proof. Consider once again the homogeneous linear system $JH(z)q = 0$ so that (4.15)–(4.17) hold with $J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) = E(\mathbf{x})^T$, since we are in the NEP case. Since $\boldsymbol{\lambda} \in \mathbb{R}^m_{++}$, (4.17) can be solved for $q^{(3)}$ and we obtain

$$\begin{aligned} 0 &\overset{(4.15)}{=} (q^{(1)})^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) q^{(1)} + (q^{(1)})^T E(\mathbf{x}) q^{(2)} \\ &\overset{(4.16)}{=} (q^{(1)})^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) q^{(1)} - (q^{(3)})^T q^{(2)} \\ &\overset{(4.17)}{=} (q^{(1)})^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) q^{(1)} + (q^{(2)})^T \mathrm{diag}(\mathbf{w} \circ \boldsymbol{\lambda}^{-1}) q^{(2)}. \end{aligned}$$

Positive semidefiniteness of $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ together with $\mathbf{w} > 0$, $\boldsymbol{\lambda} > 0$ implies $q^{(2)} = 0$ and thus also $q^{(3)} = 0$ by (4.17). Then we have from (4.15) and (4.16) that $(q^{(1)})^T J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ $q^{(1)} = 0$ and $E(\mathbf{x})^T q^{(1)} = 0$, and the assumptions show $q^{(1)} = 0$, hence nonsingularity of $JH(z)$. ∎

In spite of the result above, it should be pointed out that in general, in Theorem 4.6, we do not need the matrix $J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda})$ to be positive (semi) definite. This is illustrated by the following example.

Example 4.9. Consider a GNEP with two players, each controlling a single variable. The problem is given by

$$\begin{aligned} &\text{Player 1: } \min_{x_1} \frac{1}{2}x_1^2 - 2x_1 \qquad \text{s.t. } x_1^2 + x_2 \le 0, \\ &\text{Player 2: } \min_{x_2} \frac{1}{2}x_2^2 + (2 - x_1^2)x_2 \quad \text{s.t. } x_2 \in \mathbb{R}. \end{aligned}$$

It is easy to see that

$$J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\lambda) = \begin{pmatrix} 1 + 2\lambda & 0 \\ -2x_1 & 1 \end{pmatrix}$$

is nonsingular for all $x \in \mathbb{R}^2$ and all $\lambda > 0$, but it is not positive semidefinite everywhere. However, since a simple calculation shows that $J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\lambda)^{-1} E(\mathbf{x}) = 8x_1^2/(1+2\lambda) \ge 0$, it follows that the conditions from Theorem 4.6 are satisfied.

4.2. Boundedness. Note that Theorem 4.3 does not guarantee the existence of an accumulation point of the sequence generated by Algorithm 4.1. The following result therefore considers precisely this question and provides conditions under which the entire sequence generated by our algorithm remains bounded.

THEOREM 4.10. *Assume that*

(a) $JH(z)$ *is nonsingular for all* $z \in Z_I$*;*
(b) $\lim_{\|\mathbf{x}\|\to\infty} \|\mathbf{g}_+(\mathbf{x})\| = +\infty$*, where* $\mathbf{g}_+(\mathbf{x}) := \max\{0, \mathbf{g}(\mathbf{x})\}$*;*
(c) *the extended Mangasarian–Fromovitz constraint qualification (EMFCQ) holds for each player; i.e., for all* $\nu = 1, \ldots, N$ *and for all* $\mathbf{x} \in \mathbb{R}^n$*,*

$$\exists d^\nu \in \mathbb{R}^{n_\nu} : \nabla_{\mathbf{x}^\nu} g_i^\nu(\mathbf{x})^T d^\nu < 0 \quad \forall\, i \in I_{\ge}^\nu(\mathbf{x}), \tag{4.18}$$

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

Then any sequence $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\}$ *generated by Algorithm* 4.1 *remains bounded.*

Proof. Assume the existence of a sequence $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\} \subseteq Z_I$ such that $\lim_{k\to\infty}$ $\|(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| = \infty$. We will show that this implies $\|H(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| \to \infty$ for $k \to \infty$, contradicting part (a) of Theorem 4.3. We consider two cases.

Case 1*:* $\|(\mathbf{x}^k, \mathbf{w}^k)\| \to \infty$. Then either $\{\mathbf{x}^k\}$ is bounded, or $\|\mathbf{x}^k\| \to \infty$. If $\{\mathbf{x}^k\}$ is bounded, then $\|\mathbf{w}^k\| \to \infty$, and there exists $\nu \in \{1, \ldots, N\}$ such that $\|(\mathbf{w}^k)^\nu\| \to \infty$. Since $\{g^\nu(\mathbf{x}^k)\}$ is bounded due to the continuity of g^ν, we therefore obtain $\|g^\nu(\mathbf{x}^k) + (\mathbf{w}^k)^\nu\| \to \infty$. This, in turn, implies $\|H(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| \to \infty$. On the other hand, if $\|\mathbf{x}^k\| \to \infty$, it follows from assumption (b) that $\|g^\nu_+(\mathbf{x}^k)\| \to \infty$ for some player $\nu \in \{1, \ldots, N\}$. Moreover, since all components of the vector $\mathbf{w}^k$ are positive, this also implies $\|g^\nu(\mathbf{x}^k) + (\mathbf{w}^k)^\nu\| \to \infty$, and it follows once again that $\|H(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| \to \infty$ also in this (sub) case.

Case 2*:* $\|(\mathbf{x}^k, \mathbf{w}^k)\|$ is bounded. Then we have $\|\boldsymbol{\lambda}^k\| \to \infty$. Let ν be a player such that $\|(\boldsymbol{\lambda}^k)^\nu\| \to \infty$, and let J be the set of indices such that $(\boldsymbol{\lambda}^k)^\nu_j \to \infty$, whereas, subsequencing if necessary, we can assume that the remaining components stay bounded. Without loss of generality, we may also assume that $\mathbf{x}^k \to \bar{\mathbf{x}}$ and $\mathbf{w}^k \to \bar{\mathbf{w}}$. If, for some $j \in J$, we have $\bar{w}^\nu_j > 0$, it follows that $(\lambda^k)^\nu_j (w^k)^\nu_j \to +\infty$, and therefore $\|H(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| \to \infty$. Hence it remains to consider the case where $\bar{w}^\nu_j = 0$ for all $j \in J$. Since $(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)$ belongs to Z_I, we have $g^\nu_j(\mathbf{x}^k) + (w^k)^\nu_j > 0$, and, therefore, $g^\nu_j(\bar{\mathbf{x}}) \geq 0$ for all $j \in J$. Using EMFCQ from (4.18), there exists a vector d^ν such that $\nabla_{\mathbf{x}^\nu} g^\nu_j(\bar{\mathbf{x}})^T d^\nu < 0 \;\forall\; j \in J$. This implies

$$\begin{aligned}\lim_{k\to\infty} \nabla_{\mathbf{x}^\nu} L^\nu(\mathbf{x}^k, (\boldsymbol{\lambda}^k)^\nu)^T d^\nu &= \lim_{k\to\infty} \Big(\nabla_{\mathbf{x}^\nu} \theta_\nu(\mathbf{x}^k) + \sum_{j \notin J} (\lambda^k)^\nu_j \nabla_{\mathbf{x}^\nu} g^\nu_j(\mathbf{x}^k) \Big)^T d^\nu \\ &\quad + \lim_{k\to\infty} \Big(\sum\nolimits_{j \in J} (\lambda^k)^\nu_j \nabla_{\mathbf{x}^\nu} g^\nu_j(\mathbf{x}^k) \Big)^T d^\nu = -\infty\end{aligned}$$

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

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

for $k \to \infty$. Since d^ν is a fixed vector, this implies $\|\nabla_{\mathbf{x}^\nu} L^\nu(\mathbf{x}^k, (\boldsymbol{\lambda}^k)^\nu)\| \to +\infty$ which, in turn, implies $\|H(\mathbf{x}^k, \boldsymbol{\lambda}^k, \mathbf{w}^k)\| \to \infty$ for $k \to \infty$, also in this case. ∎

Note that condition (b) in the theorem above is a mild boundedness assumption on the feasible sets of the players. In particular, (b) holds in the setting of Proposition 3.6. Also condition (c) is rather mild and common in an optimization context.

Remark 4.11. As we have seen in the previous sections, nonsingularity of $J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})$ and the P_0-condition on the matrix $M(\mathbf{x}, \boldsymbol{\lambda})$ 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 semidefiniteness assumptions on $J_{\mathbf{x}}\mathbf{F}(\mathbf{x}, \boldsymbol{\lambda})$. Let us recall that in the context of the interior-point approach, all conditions only have to hold for positive $\boldsymbol{\lambda}$ and, therefore, are less restrictive than in the merit function context.

Remark 4.12. Consider once again the merit function approach from section 3, but with the Fischer–Burmeister function ϕ being replaced by the penalized Fischer–Burmeister function $\phi(a, b) := \sqrt{a^2 + b^2} - a - b - a_+ b_+$ from [4], where $a_+ := \max\{0, a\}$ and b_+ is defined similarly. Consider any sequence $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ such that $\Theta(\mathbf{x}^k, \boldsymbol{\lambda}^k) \leq \Theta(\mathbf{x}^0, \boldsymbol{\lambda}^0)$ and assume that the assumptions from Theorem 4.10 hold. Then one can show that $\{(\mathbf{x}^k, \boldsymbol{\lambda}^k)\}$ stays bounded (which is a stronger statement than Proposition 3.6). In fact, if $\{\|\mathbf{x}^k\|\} \to \infty$, then $\|\mathbf{g}_+(\mathbf{x}^k)\| \to \infty$, and similar to the proof

of Case 1 of Theorem 4.10, this would imply $\Theta(\mathbf{x}^k, \boldsymbol{\lambda}^k) \to \infty$. Hence $\{\mathbf{x}^k\}$ is bounded. Therefore, without loss of generality, let $\{\mathbf{x}^k\} \to \bar{\mathbf{x}}$. If $\|\boldsymbol{\lambda}^k\| \to \infty$, then there exist ν, j such that $(\lambda^k)_j^\nu \to +\infty$ (the definition of the penalized Fischer–Burmeister function implies that $(\lambda^k)_j^\nu \to -\infty$ cannot occur). Due to the penalty term in the penalized Fischer–Burmeister function, we obtain $g_j^\nu(\bar{\mathbf{x}}) \geq 0$. The statement then follows as in the proof of Case 2 of Theorem 4.10.

5. Numerical results. In this section we compare numerically the approaches proposed in the previous two sections. We should point out that in reality in section 4 we actually proposed an algorithm, while in section 3 we studied only the properties of the merit function Θ, so that for each choice of a specific minimization routine we will have a different method. In this section we will consider two different ways to minimize Θ, therefore giving rise to two different algorithms. A third method that solves a nonlinear equation system with box constraints is considered, and all three algorithms are compared to the potential reduction method.

5.1. Test problems and stopping criterion. We solve several test problems, mostly taken from the extensive numerical test library in [17]. We also consider some further test problems from the literature, namely, Harker's problem (Harker) described in [24], an electricity market problem (Heu) from [25], two small problems (NTF1, NTF2) from [27], a transportation problem from [22] in different dimensions (Tr1a, Tr1b, Tr1c), a Spam-filtering problem (Spam) which is a multiplayer version of the two-player game described in [3], and a model for a lobbying process (Lob); see [34]. We report in Table 5.1 the number of players and the total number of variables and constraints of each problem. Some of the test problems were run more than one time using different starting points: the number of starting points used is reported in the column #s.p. For a detailed description of the problems we refer the reader to the references above; here we report only some general characteristics. Problems from A.1 to Tr1c are general GNEPs, while problems from A.11 to Spam are jointly convex GNEPs. Actually, our test problem set includes four pure NEPs: A.12, A.15, Lob, and Spam. The objective functions of each player's problems are, for fixed $\mathbf{x}^{-\nu}$, as follows:

- A.9a, A.9b : linear;
- A.3, A.4, A.5, A.6, A.7, A.8, A.10a, A.10c, A.11, A.12, A.13, A.15, A.17, A.18, Harker, NTF1, NTF2 : quadratic;
- A.1, A.2, A.10b, A.10d, A.10e, A.12, A.14, A.16 (all), Heu, Lob, Spam, Tr1 (all): nonlinear.

The constraints of each player's problem are, for fixed $\mathbf{x}^{-\nu}$, as follows:

- A.1, A.2, A.3, A.4, A.5, A.7, A.8, A.11, A.12, A.13, A.14, A.15, A.16 (all), A.17, A.18, Harker, Heu, NTF1, Lob, Spam, Tr1 (all): linear;
- A.6, A.9a, A.9b, A.10 (all), NTF2: nonlinear.

Problems A.3–A.8, A.11, A.12, A.17, Harker, NTF1, and NTF2 are purely academic problems, while the remaining problems correspond to some kind of engineering or economic models.

The methods discussed below use the same stopping criterion. We stopped the iterations when the violation $V(\mathbf{x}, \boldsymbol{\lambda})$ of the KKT conditions (2.2) is small, i.e., we set

$$V(\mathbf{x}, \boldsymbol{\lambda}) = \left\| \begin{pmatrix} \mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) \\ \min(\boldsymbol{\lambda}, -\mathbf{g}(\mathbf{x})) \end{pmatrix} \right\|_2$$

TABLE 5.1
Data on test problems.

Example	N	n	m	#s.p.
A.1	10	10	20	3
A.2	10	10	24	3
A.3	3	7	18	3
A.4	3	7	18	3
A.5	3	7	18	3
A.6	3	7	21	3
A.7	4	20	44	3
A.8	3	3	8	3
A.9a	7	56	63	1
A.9b	7	112	119	1
A.10a	8	24	33	1
A.10b	25	125	151	1
A.10c	37	222	260	1
A.10d	37	370	408	1
A.10e	48	576	625	1
Tr1a	6	18	72	2
Tr1b	6	60	228	2
Tr1c	7	80	304	2
A.11	2	2	2	1
A.12	2	2	4	1
A.13	3	3	9	1
A.14	10	10	20	1
A.15	3	6	12	1
A.16a	5	5	10	1
A.16b	5	5	10	1
A.16c	5	5	10	1
A.16d	5	5	10	1
A.17	2	3	7	1
A.18	2	12	28	3
Harker	2	2	6	1
Heu	2	10	22	2
NTF1	2	2	4	1
NTF2	2	2	4	1
Lob	50	50	50	1
Spam	101	2020	4040	1

and stopped iterations when $V(\mathbf{x}^k, \lambda^k) \leq \sqrt{n+m}10^{-4}$. In addition, we terminated the iterations when the maximum number of 10^3 iterations was reached.

5.2. Merit function approach. In the merit function approach we must solve the unconstrained optimization problem

$$\min \boldsymbol{\Theta}(\mathbf{x}, \boldsymbol{\lambda}).$$

In order to do so we tried two different, somewhat extreme, approaches.

General purpose minimization algorithm. As a first option, we used a general purpose algorithm that does not exploit in any way the structure of the objective function $\boldsymbol{\Theta}$. This is by far the simplest choice and requires little beyond furnishing routines that calculate the objective and gradient values. In particular we used the MATLAB function *fminunc* from the optimization toolbox with option *GradObj* set to "on". Besides the function and the gradient, this routine requires only a starting point $(\mathbf{x}^0, \boldsymbol{\lambda}^0)$, but no further ingredients.

In addition to the main stopping criteria described above, the *fminunc* algorithm stops if the relative change in function value is less than the parameter *TolFun* or the maximum number of function evaluations *MaxFunEvals* is reached. We set $TolFun = 10^{-8}$ and $MaxFunEvals = 10^3$. For the $\boldsymbol{\lambda}$-part of the starting vector, we always used $\boldsymbol{\lambda}^0 = 0$, whereas details regarding the $\mathbf{x}$-part are given in [9]. We set the MATLAB option $LargeScale$ "off", so that *fminunc* uses a BFGS line-search algorithm for the minimization.

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 $\boldsymbol{\Theta}$ is not so; in fact, $\nabla\boldsymbol{\Theta}$ is only strongly semismooth; see [19]. So, as an alternative method, we implemented in MATLAB the semismooth minimization algorithm from [7], [8], [19]. This is a globalized semismooth Newton-type method which has fast local convergence. We refer the reader to the references above for the details and here report only some relevant implementation details. For the sake of notational simplicity we set

$$T(\mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} \mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) \\ \boldsymbol{\Phi}(\mathbf{x}, \boldsymbol{\lambda}) \end{pmatrix}.$$

In each iteration of this method, in order to find a search direction an element of the B-subdifferential $\partial_B T(\mathbf{x}, \boldsymbol{\lambda})$ is evaluated; see [19], [28]. The following theoretical procedure evaluates an element H belonging to $\partial_B T(\mathbf{x}, \boldsymbol{\lambda})$. This procedure is analogous to that reported in [7], and it can be proved, along lines similar to those in [7], that it actually provides an element in the B-subdifferential. We gloss over the detailed proofs, since it is just an extension of known techniques.

Step 1: Set $\beta = \{(\nu, i) : \lambda_i^\nu = 0 = g_i^\nu(\mathbf{x})\}$.

Step 2: For each $(\nu, i) \notin \beta$ set

$$a_i^\nu = \left(\frac{\lambda_i^\nu}{\|\lambda_i^\nu, -g_i^\nu(\mathbf{x})\|_2} - 1\right) \quad \text{and} \quad b_i^\nu = \left(\frac{-g_i^\nu(\mathbf{x})}{\|\lambda_i^\nu, -g_i^\nu(\mathbf{x})\|_2} - 1\right).$$

Step 3: For each $(\nu, i) \in \beta$ set $a_i^\nu = -1$ and $b_i^\nu = -1$.

Step 4: Using definitions (3.1)–(3.2), set

$$H = \begin{pmatrix} J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) & E(\mathbf{x}) \\ -D_{\mathbf{g}}(\mathbf{x},\boldsymbol{\lambda})J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) & D_{\lambda}(\mathbf{x},\boldsymbol{\lambda}) \end{pmatrix}.$$

Remark 5.1. Note that at points where $\beta = \varnothing$ (which are called nondegenerate points) $T(\mathbf{x},\boldsymbol{\lambda})$ is differentiable and then the above procedure gives the Jacobian of $T(\mathbf{x},\boldsymbol{\lambda})$. Moreover, we stress that the overhead for the computation of the matrix H is 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 [14], the conditions under which superlinear convergence occur are often in jeopardy when solving reformulations of GNEPs. In this paper we did not discuss the local convergence properties of any of the methods analyzed, so we cannot 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 solving an $n+m$ square linear system

$$H^k d = -T(\mathbf{x}^k,\boldsymbol{\lambda}^k). \tag{5.1}$$

In order to perform the linear algebra involved we used MATLAB's linear systems solver *linsolve*. In a few cases, if the Newton-like direction does not satisfy certain "sufficient descent" conditions, the line search is performed along the antigradient of Θ. The details are as follows: if the 1-norm condition number estimate of H^k is bigger than 10^{16} (that is, the linear system (5.1) is ill conditioned) or if $\nabla\Theta(\mathbf{x}^k,\boldsymbol{\lambda}^k)^T d^k > -10^{-8}\|d^k\|_2^{2.1}$ (that is, d^k is rather orthogonal to $\Theta(\mathbf{x}^k,\boldsymbol{\lambda}^k)$ and then the succeeding line search will generate tiny stepsizes), then d^k is taken as $-\nabla\Theta(\mathbf{x}^k,\boldsymbol{\lambda}^k)$.

Then we used an Armijo-type line search that finds the smallest $i^k = 0, 1, 2, \ldots$ such that

$$\Theta((\mathbf{x}^k,\boldsymbol{\lambda}^k) + 2^{-i^k}d^k) \le \Theta(\mathbf{x}^k,\boldsymbol{\lambda}^k) + 10^{-4}2^{-i^k}\nabla\Theta(\mathbf{x}^k,\boldsymbol{\lambda}^k)^T d^k.$$

For the λ-part of the starting vector, we always used $\boldsymbol{\lambda}^0 = 0$.

5.3. Interior-point method. We have implemented only the exact version of Algorithm 4.1, because the library of test problems considered does not contain large-scale problems.

More in detail, at step $(S.2)$ of Algorithm 4.1 we find the search direction d^k by solving a reduced linear system of equations with $\sigma^k = 0.1$. Note that formally this method calls for the solution of an $n+2m$ 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, we must find a solution $(\bar{d}_1, \bar{d}_2, \bar{d}_3)$ of the following linear system of dimension $n+2m$:

$$\begin{pmatrix} J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) & E(\mathbf{x}) & 0 \\ J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) & 0 & I \\ 0 & \operatorname{diag}(\mathbf{w}) & \operatorname{diag}(\boldsymbol{\lambda}) \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \tag{5.2}$$

where all the quantities involved are defined in detail in section 4. It is easy to verify, by substitution and by the fact that $\mathbf{w} > 0$ in Z_I, that if we compute $\bar{d}_1$ as solution of

$$(J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) + E(\mathbf{x})\mathrm{diag}(\mathbf{w})^{-1}\mathrm{diag}(\boldsymbol{\lambda})J_{\mathbf{x}}\mathbf{g}(\mathbf{x}))d_1$$
$$= b_1 - E(\mathbf{x})\mathrm{diag}(\mathbf{w})^{-1}b_3 + E(\mathbf{x})\mathrm{diag}(\mathbf{w})^{-1}\mathrm{diag}(\boldsymbol{\lambda})b_2$$

and $\bar{d}_2$ and $\bar{d}_3$ by $d_3 = b_2 - J_{\mathbf{x}}\mathbf{g}(\mathbf{x})d_1$ and $d_2 = \mathrm{diag}(\mathbf{w})^{-1}b_3 - \mathrm{diag}(\mathbf{w})^{-1}\mathrm{diag}(\boldsymbol{\lambda})d_3$, respectively, this is indeed a solution of (5.2). This shows clearly that the main computational burden in solving the linear system (5.2) is actually the solution of an $n \times n$ square linear system. In order to perform the linear algebra involved we used MATLAB's linear systems solver *linsolve*.

Similarly to what is done in the semismooth-like approach, if the Newton-like direction does not satisfy $\nabla\psi(\mathbf{x}^k,\boldsymbol{\lambda}^k,\mathbf{w}^k)^T d^k \le -10^{-8}\|d^k\|_2^{2.1}$, that is, if the direction d^k is almost orthogonal to $\nabla\psi(\mathbf{x}^k,\boldsymbol{\lambda}^k,\mathbf{w}^k)$, then we use the antigradient $-\nabla\psi(\mathbf{x}^k,\boldsymbol{\lambda}^k,\mathbf{w}^k)$ as a search direction d^k.

The line search used is described in step (S.3) of Algorithm 4.1, with $\gamma = 10^{-3}$ and $\xi = 2m$. In order to stay in Z_I we preliminarily rescale $d^k = (d_x^k, d_\lambda^k, d_w^k)$. First we analytically compute a positive constant α such that $\boldsymbol{\lambda}^k + \alpha d_\lambda^k$ and $\mathbf{w}^k + \alpha d_w^k$ are greater than 10^{-10}. This ensures that the last two blocks in $z^k + \alpha d^k$ are in the interior of $\mathbb{R}^{2m}_+$. Then, if necessary, we further reduce this α by successive bisections until $\mathbf{g}(\mathbf{x}^k + \alpha d_x^k) + \mathbf{w}^k + \alpha d_w^k \ge 10^{-10}$, thus finally guaranteeing that $z^k + \alpha d^k$ belongs to Z_I. In this latter phase, an evaluation of $\mathbf{g}$ is needed for each bisection. At the end of this process, we set $d^k \leftarrow \alpha d^k$ and then proceed to perform the Armijo line search (4.8).

For the $(\boldsymbol{\lambda},\mathbf{w})$-part of the starting vector, we used $\boldsymbol{\lambda}^0 = 10$ and $\mathbf{w}^0 = \max(10, 5 - \mathbf{g}(\mathbf{x}^0))$, so that we are sure that the starting point is "well inside" Z_I.

5.4. The scaled trust-region solver for constrained nonlinear equations. We also considered a trust-region method for solving the constrained equation defined by (4.2) and (4.3). To this end, we used the scaled trust-region solver for constrained nonlinear equations (STRSCNE) , a software freely available at http://strscne.de.unifi.it and whose detailed description can be found in [1], [2]. Here we give a few details to make a comparison with the other methods we tested possible. STRSCNE is essentially a suitably tailored method that minimizes $\frac{1}{2}\|H(\mathbf{x},\boldsymbol{\lambda},\mathbf{w})\|^2$ over (4.3). 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 to 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 (4.3). 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 our interior-point method appears particularly appropriate and meaningful.

The algorithm is globally convergent to a stationary point of $\frac{1}{2}\|H(\mathbf{x},\boldsymbol{\lambda},\mathbf{w})\|^2$ over (4.3). As usual, if the stationary point so found is a global minimizer with zero value, the point is a solution of the constrained system (4.2) and (4.3). However, we remark that conditions that guarantee that stationary points are actually solutions of the original constrained system (4.2) and (4.3) are not available at the moment; hence it is not clear how typical nonoptimal stationary points are.

We slightly modified the STRSCNE implementation so that the method uses the same stopping criteria employed by the other methods we tested. We underline that the

dogleg strategy used in order to approximately solve the trust-region problem entails that, as in all other methods we considered, 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 our interior-point method.

5.5. Comparison of the algorithms. In order to evaluate the algorithms we ran each algorithm on all test problems using, in some cases, several starting points (see Table 5.1). This resulted in 57 runs for each method. The parameters that we took into account are as follows: the number of iterations (It.), the number of constraint evaluations (meaning the number of times $\mathbf{g}$ is evaluated) (g), the number of times the partial gradients $\nabla_{x^\nu}\theta_\nu$ are evaluated (Pg) (note that each time this counter is incremented by one this means that the partial gradient of all players have been evaluated), the number of times $J\mathbf{g}$ is evaluated (Jg), and the number of times $J\mathbf{F}$ is evaluated (JF). These performance criteria give a fairly detailed picture of the computation costs of each algorithm. Note, in particular, that at each iteration of the algorithms considered, the most costly operation is the solution of a square linear system. These systems have dimension $n+m$, $n+m$, $n+2m$, and $n+2m$, respectively. However, we already discussed that 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 could 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

$$\begin{pmatrix} J_{\mathbf{x}}\mathbf{F}(\mathbf{x},\boldsymbol{\lambda}) & E(\mathbf{x}) \\ -D_{\mathbf{g}}(\mathbf{x},\boldsymbol{\lambda})J_{\mathbf{x}}\mathbf{g}(\mathbf{x}) & D_{\boldsymbol{\lambda}}(\mathbf{x},\boldsymbol{\lambda}) \end{pmatrix}.$$

The peculiarity of this matrix is that the bottom-right block is diagonal. So one could think that, similar to what is done for the interior-point method linear system, one could express the $\boldsymbol{\lambda}$ variables in function of $\mathbf{x}$ and then solve a square n system. However, in general the bottom-right diagonal block could easily have zero or very small entries. In particular, suppose that $(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})$ is a solution of the KKT conditions of the game. If, for example, we have $g_1^1(\bar{\mathbf{x}})=0$ and $\bar{\lambda}_1^1>0$, i.e., if the first constraint of the first player is active and has a positive multiplier (a common case indeed), we see that the corresponding element $[D_\lambda(\bar{\mathbf{x}},\bar{\boldsymbol{\lambda}})]_{1,1}$ is 0. So, in a neighborhood of this point this entry will be either 0 or very small, and we 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 $D_{\boldsymbol{\lambda}}(\mathbf{x},\boldsymbol{\lambda})$ are all positive, but for the reasons exposed above we preferred to leave the detection and handling of this diagonal block to the linear system solver. Note that, here, the interior-point method has an advantage, since the diagonal blocks present in the linear system are always guaranteed to have positive diagonal elements, exactly because we keep iterations in Z_I.

The detailed description of our tests are reported in [9]; for lack of space, here we only report some summary results. The first consideration we can make is that the unconstrained minimization of Θ through the general purpose code *fminunc* is not competitive with the other three approaches. This approach leads to very many failures (19), and the numbers for the iterations and the other performance criteria considered are consistently higher than those for the other algorithms. In Table 5.2 we report the total number of failures for the semismooth-like algorithm, STRSCNE, and the

TABLE 5.2
Cumulative results for the semismooth-like algorithm, STRSCNE, and the interior-point method.

Algorithm	Failures	It.	g	Pg + Jg	JF
Semismooth-like	8	1217	13018	24772	1264
STRSCNE	3	2158	2257	6625	2205
Interior-point	1	857	2103	3243	857

interior-point method, along with the cumulative counts obtained by considering only runs that are solved by all three algorithms (for a total of 47 runs).

This table shows that the interior-point method seems more reliable in that it solves all problems except one. The cumulative results also seem to favor the interior-point method. An analysis of the detailed results in [9] shows that actually the semismooth-like algorithm performs marginally better on a good part of the problems, but for some problems its behavior deteriorates greatly, which increases very much the cumulative results, and it cannot solve any of the transportation problems Tr1a, Tr1b, or Tr1c.

To get a better picture of the behavior of the algorithms, we also present performance profiles [11]. We briefly recall how these profiles are defined. We 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). We compare the performance on problem p by algorithm a with the best performance by any algorithm on this problem using the following performance ratio:

$$r_{p,a} = \frac{m_{p,a}}{\min\{m_{p,a} : a \in A\}}.$$

Then we obtain an overall assessment of the performance of the algorithm by defining the following value $\rho_a(\tau) = \text{size}\{p \in P : r_{p,a} \leq \tau\}/n_p$, which represents the probability for algorithm $a \in A$ that the performance ratio $r_{p,a}$ is within a factor $\tau \in R$ 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.

In our comparison we take as performance measures the number of iterations that the four methods take to reach a solution of the GNEP (number of linear systems solved), the number of times that each algorithm evaluates the constraints of the GNEP (use of 0th order information), the number of times that each algorithm evaluates the partial gradient of the objective functions of each player plus the number of times that each algorithm evaluates the Jacobian of the constraints (use of first order information), and the number of times that $J\mathbf{F}$ is evaluated (use of second order information). The results are shown in Figure 5.1.

These profiles confirm and make the impressions described above more precise. For $\tau = 1$ we see that the semismooth-like algorithm performs best with respect to all criteria except g (even if the detailed results indicate that often the advantage is very slight). 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 3 (more

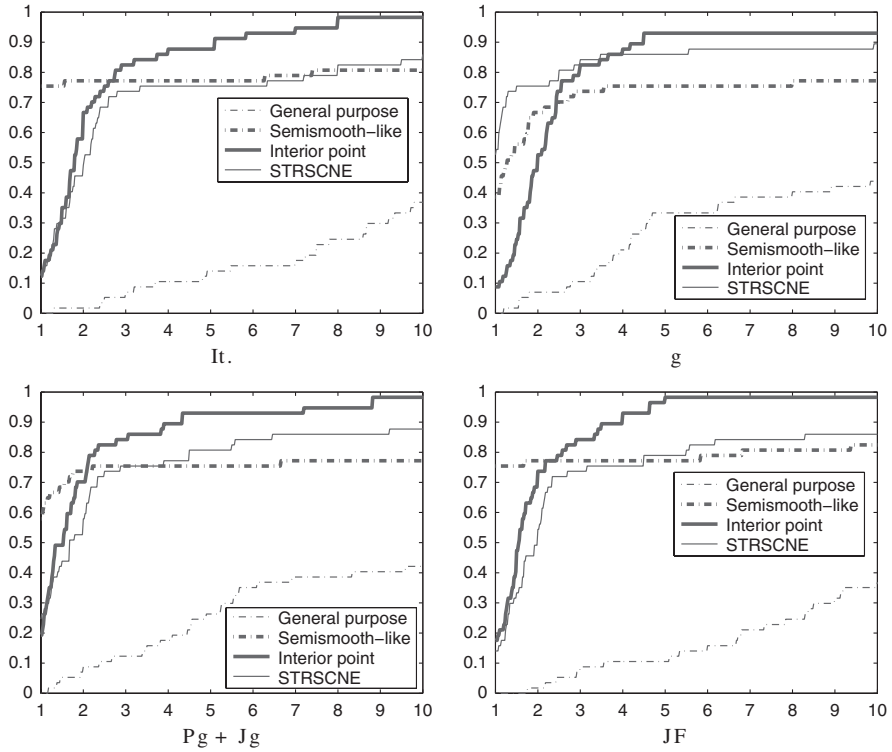

FIG. 5.1. *Performance profiles.*

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 algorithm and is more reliable. The performance of the STRSCNE method is for $\tau < 2$, about the same as for the interior-point method, but for larger τ the latter one is superior. Only for the g evaluations is the STRSCNE method superior to all other methods for $\tau < 4$.

Our implementations of the semismooth and interior-point methods are certainly not very sophisticated, but the results seem to indicate that these two methods are worthy of further investigation and could be the basis for an efficient solution method for GNEPs. We remark that we are aware of only one other method for which relatively extensive numerical testing has been performed: the solution of a general GNEP with guaranteed convergence properties; this is the penalty approach proposed in [16]. It is not totally straightforward to compare the results reported in [16] and those reported here. For one thing, the test set in [16] is a subset of the problems considered in this paper, and the stopping criterion is different. Nevertheless, each minor iteration in [16] requires the solution of a linear system, and, from the linear algebra point of view, this is still the main computational effort of the penalty algorithm. A comparison of the results in this paper and of those in [16] seems to indicate that the solution of the KKT conditions is by far more efficient than the penalty approach.

Acknowledgments. The authors would like to thank the referees for their careful reading of the original manuscript and for several interesting comments. In particular, Remark 4.12 was suggested by a referee.

REFERENCES

[1] S. Bellavia, M. Macconi, and B. Morini, *An affine scaling trust-region method approach to bound-constrained nonlinear systems*, Appl. Numer. Math., 44 (2003), pp. 257–280.

[2] S. Bellavia, M. Macconi, and B. Morini, *STRSCNE: A scaled trust-region solver for constrained nonlinear equations*, Comput. Optim. Appl., 28 (2004), pp. 31–50.

[3] M. Brückner and T. Scheffer, *Nash equilibria of static prediction games*, Adv. Neural Inf. Process. Syst., 22 (2009), pp. 171–179.

[4] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer–Burmeister NCP-function*, Math. Program., 88 (2000), pp. 211–216.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[6] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem*, Academic Press, New York, 1992.

[7] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75 (1996), pp. 407–439.

[8] T. De Luca, F. Facchinei, and C. Kanzow, *A theoretical and numerical comparison of some semismooth algorithms for complementarity problems*, Comput. Optim. Appl., 16 (2000), pp. 173–205.

[9] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella, *On the Solution of the KKT Conditions of Generalized Nash Equilibrium Problems (with Complete Numerical Results)*, preprint 302, Institute of Mathematics, University of Würzburg, Germany, 2010; also available online from http://www.mathematik.uni-wuerzburg.de/~kanzow/index.html.

[10] A. Dreves, C. Kanzow, and O. Stein, *Nonsmooth optimization reformulations for player convex generalized Nash equilibrium problems*, J. Global Optim., to appear, DOI 10.1007/s10898-011-9727-9.

[11] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. Ser. A, 91 (2002), pp. 201–213.

[12] F. Facchinei, A. Fischer, and C. Kanzow, *Regularity properties of semismooth reformulations of variational inequalities*, SIAM J. Optim., 8 (1998), pp. 850–869.

[13] F. Facchinei, A. Fischer, C. Kanzow, and J.-M. Peng, *A simply constrained optimization reformulation of KKT systems arising from variational inequalities*, Appl. Math. Optim., 40 (1999), pp. 19–37.

[14] F. Facchinei, A. Fischer, and V. Piccialli, *Generalized Nash equilibrium problems and Newton methods*, Math. Program., 117 (2009), pp. 163–194.

[15] F. Facchinei and C. Kanzow, *Generalized Nash equilibrium problems*, 4OR, 5 (2007), pp. 173–210.

[16] F. Facchinei and C. Kanzow, *Penalty methods for the solution of generalized Nash equilibrium problems*, SIAM J. Optim., 20 (2010), pp. 2228–2253.

[17] 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.

[18] F. Facchinei and L. Lampariello, *Partial penalization for the solution of generalized Nash equilibrium problems*, J. Global Optim. 50 (2010), DOI 10.1007/s10898-010-9579-8.

[19] F. Facchinei and J.-S. Pang, *Finite–Dimensional Variational Inequalities and Complementarity Problems,* Vol. II, Springer Ser. Oper. Res., Springer–Verlag, New York, 2003.

[20] F. Facchinei and J.-S. Pang, *Nash equilibria: the variational approach*, in Convex Optimization in Signal Processing and Communications, D. P. Palomar and Y. C. Eldar, eds., Cambridge University Press, Cambridge, 2010, pp. 443–493.

[21] F. Facchinei and S. Sagratella, *On the computation of all solutions of jointly convex generalized Nash equilibrium problems*, Optim. Lett., (2010), DOI 10.1007/s11590-010-0218-6.

[22] F. Facchinei, A. Sgalambro, and S. Teobaldo, *On some generalized Nash equilibrium problems in a freight distribution environment: Properties and existence conditions*, in preparation.

[23] M. Fukushima, *Restricted generalized Nash equilibria and controlled penalty algorithm*, Comput. Manag. Sci. (2010), DOI 10.1007/s10287-009-0097-4.

[24] P. T. Harker, *Generalized Nash games and quasi-variational inequalities*, Eur. J. Oper. Res., 54 (1991), pp. 81–94.

[25] A. von Heusinger, *Numerical Methods for the Solution of the Generalized Nash Equilibrium Problem*, Ph.D. thesis, Institute of Mathematics, University of Würzburg, Würzburg, Germany, 2009.

[26] R. D. C. Monteiro and J.-S. Pang, *A potential reduction Newton method for constrained equations*, SIAM J. Optim., 9 (1999), pp. 729–754.

[27] K. Nabetani, P. Tseng, and M. Fukushima, *Parametrized variational inequality approaches to generalized Nash equilibrium problems with shared constraints*, Comput. Optim. Appl., 48 (2011), pp. 423–452.

[28] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program. Ser. A, 58 (1993), pp. 353–368.

[29] J.-S. Pang, *Computing Generalized Nash Equilibria*, Technical report, Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD, 2002.

[30] J.-S. Pang and M. Fukushima, *Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Comput. Manag. Sci., 2 (2005), pp. 21–56.

[31] J.-S. Pang and M. Fukushima, *Erratum: Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games*, Comput. Manag. Sci., 6 (2009), pp. 373–375.

[32] J.-S. Pang, G. Scutari, F. Facchinei, and C. Wang, *Distributed power allocation with rate constraints in Gaussian parallel interference channels*, IEEE Trans. Inform. Theory, 54 (2008), pp. 3471–3489.

[33] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[34] G. Tullock, *Efficient rent seeking*, in Towards a Theory of the Rent–Seeking Society, J. Buchanan, R. Tollison, and G. Tullock, eds., Texas A&M University Press, College Station, TX, 1980, pp. 97–112.