Mathematical Programming 75 (1996) 407–439

A semismooth equation approach to the solution of nonlinear complementarity problems

Tecla De Luca [a,1], Francisco Facchinei [a,*], Christian Kanzow [b,2]

[a] *Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Via Buonarroti 12, 00185 Roma, Italy*

[b] *Institute of Applied Mathematics, University of Hamburg, Bundesstrasse 55, D-20146 Hamburg, Germany*

Received 31 January 1995; revised manuscript received 11 December 1995

Abstract

In this paper we present a new algorithm for the solution of nonlinear complementarity problems. The algorithm is based on a semismooth equation reformulation of the complementarity problem. We exploit the recent extension of Newton's method to semismooth systems of equations and the fact that the natural merit function associated to the equation reformulation is continuously differentiable to develop an algorithm whose global and quadratic convergence properties can be established under very mild assumptions. Other interesting features of the new algorithm are an extreme simplicity along with a low computational burden per iteration. We include numerical tests which show the viability of the approach.

Keywords: Nonlinear complementarity problem; Semismoothness; Smooth merit function; Global convergence; Quadratic convergence

1. Introduction

We consider the nonlinear complementarity problem, NCP(F) for short, which is to find a vector in $\mathbb{R}^n$ satisfying the conditions

$$x \geq 0, \qquad F(x) \geq 0, \qquad x^{\mathrm{T}} F(x) = 0,$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a given function which we shall always assume to be continuously differentiable.

* Corresponding author. Email: soler@peano.dis.uniroma1.it.

[1] Email: deluca@newton.dis.uniroma1.it.

[2] Email: kanzow@math.uni-hamburg.de.

Published by Elsevier Science B.V.
PII S0025-5610(96)00028-7

The fastest algorithms for the solution of NCP(F) are Newton-type methods, which, however, are in general only locally convergent. In the last years much attention has been devoted to techniques for globalizing these local algorithms. To this end there have been several different proposals, but a common scheme can be seen to underlie most of these methods:

(a) reformulate NCP(F) as a system of (possibly nonsmooth) equations $\Phi(x) = 0$;
(b) define a local Newton-type method for the solution of the system of equations;
(c) perform a linesearch to minimize a suitable merit function, usually $\|\Phi(x)\|^2$ or $\|\Phi(x)\|$, in order to globalize the local method.

There are several possibilities to redefine NCP(F) as a system of equations. The first proposal is probably due to Mangasarian, see [29], where a class of reformulations, mainly smooth ones, is described; the smooth reformulation approach has been further explored in [9,15,25–27,30,32,49]. In the last years, however, nonsmooth reformulations have attracted much more attention [5,7,8,10,14,16,18,20,32,34–36,41,46,54,55], since they allow the definition of superlinearly convergent algorithms even for degenerate problems, and the subproblems to be solved at each iteration tend to be more numerically stable. However, there is a price to pay: the globalization (point (c) above) becomes more complex since the merit function $\|\Phi(x)\|^2$ is nonsmooth. Furthermore, the subproblems which have to be solved at each iteration, usually (mixed) linear complementarity problems, are more complex than in the smooth reformulation case. We note that a common analytical property of the recent nonsmooth reformulations is that, though not F-differentiable, they are B-differentiable, so that the machinery developed in [34,43–45] can be usefully employed.

In this paper we describe a new algorithm for the solution of NCP(F) which is both globally and superlinearly convergent. We use a nonsmooth equation reformulation which is based on the simple function $\phi(a, b) := \sqrt{a^2 + b^2} - (a + b)$ which was first introduced by Fischer [11] and further employed by several authors, see [7,8,12,16,25,27,39,52]. The resulting system of equations is not F-differentiable, but nevertheless it is semismooth [31,40]. Semismoothness is a stronger analytical property than B-differentiability so that, using the recent powerful theory for the solution of semismooth equations [37,38,40], it is possible to develop a fast local algorithm for the solution of NCP(F) which only requires, at each iteration, the solution of one linear system. Furthermore, the natural merit function $\|\Phi(x)\|^2$ is, surprisingly, smooth, so that global convergence through a linesearch procedure can easily be enforced. The assumptions under which global and superlinear convergence can be proved compare favourably with those of existing algorithms and the overall algorithm seems to convey the advantages of both smooth and nonsmooth reformulations of NCP(F) in a very simple scheme.

The use of the function ϕ mentioned above to reformulate nonlinear complementarity problems as systems of nonsmooth equations seems a very promising approach. Algorithms for linear complementarity problems based on this function were considered in [12,27], while the nonlinear case was studied in [7,8,16,25]. The results of this paper, however, are an improvement on previous results because the semismooth nature of the

reformulation is here fully exploited for the first time and the overall analysis carried out is much more refined and detailed than in previous works. While we were completing this paper, we became aware of a recent report of Jiang and Qi [23] which has some similarities with this work in that the same local approach is adopted. However, the global algorithm is quite different and the analysis is restricted to the case in which F is a uniform P-function.

This paper is organized as follows, in the next section we collect some background material, in Section 3 we state the algorithm and its main properties, Sections 4 and 5 are devoted to some technical, but fundamental results, while in Section 6 we prove and further discuss the properties of the algorithm. In Section 7 we report some preliminary numerical results to show the viability of the approach adopted. Some concluding remarks are made in the last section.

Throughout this paper, the index set $\{1,\dots,n\}$ is abbreviated by the upper case letter I. For a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$, we denote the Jacobian of F at $x \in \mathbb{R}^n$ by $F'(x)$, whereas the transposed Jacobian is denoted by $\nabla F(x)$. $\|\cdot\|$ denotes the Euclidean norm and $S(\bar{x},\delta)$ is the closed Euclidean sphere of center $\bar{x}$ and radius δ, i.e. $S(x,\delta) = \{x \in \mathbb{R}^n : \|x-\bar{x}\| \le \delta\}$. If Ω is a subset of $\mathbb{R}^n$, $dist\{x \mid \Omega\} := \inf_{y\in\Omega} \|y-x\|$ denotes the (Euclidean) distance of x to Ω. If M is an $n \times n$ matrix with elements M_{jk}, $j,k = 1,\dots,n$, and J and K are index sets such that $J,K \subseteq \{1,\dots,n\}$, we denote by $M_{J,K}$ the $|J| \times |K|$ submatrix of M consisting of elements M_{jk}, $j \in J$, $k \in K$ and, assuming that $M_{J,J}$ is nonsingular, by $M/M_{J,J}$ the Schur-complement of $M_{J,J}$ in M, i.e. $M/M_{J,J} = M_{K,K} - M_{K,J}M_{J,J}^{-1}M_{J,K}$, where $K = I \setminus J$. If w is an n vector, we denote by w_J the subvector with components w_j, $j \in J$.

2. Background material

In this section we collect several definitions and results which will be used throughout the paper.

2.1. Basic definitions

A solution to the nonlinear complementarity problem NCP(F) is a vector $x^* \in \mathbb{R}^n$ such that

$$F(x^*) \ge 0, \qquad x^* \ge 0, \qquad F(x^*)^{\mathrm{T}}x^* = 0.$$

Associated to the solution x^* we define three index sets:

$$\alpha := \{i \mid x_i^* > 0\}, \qquad \beta := \{i \mid x_i^* = 0 = F_i(x^*)\}, \qquad \gamma := \{i \mid F_i(x^*) > 0\}.$$

The solution x^* is said to be nondegenerate if $\beta = \emptyset$.

Definition 1. We say that the solution x^* is *R-regular* if $\nabla F_{\alpha,\alpha}(x^*)$ is nonsingular and the Schur-complement of $\nabla F_{\alpha,\alpha}(x^*)$ in

$$\begin{pmatrix} \nabla F_{\alpha,\alpha}(x^*) & \nabla F_{\alpha,\beta}(x^*) \\ \nabla F_{\beta,\alpha}(x^*) & \nabla F_{\beta,\beta}(x^*) \end{pmatrix}$$

is a P-matrix (see below).

Note that R-regularity coincides with the notion of regularity introduced by Robinson in [47] (see also [42], where the same condition is called strong regularity) and is strictly related to similar conditions used, e.g., in [10,32,36].

We next introduce a few classes of matrices and functions and recall some related properties.

Definition 2. A matrix $M \in \mathbb{R}^{n\times n}$ is a

- P_0-matrix if each of its principal minors is non-negative;
- P-matrix if each of its principal minors is positive;
- S_0-matrix if

$$\{x \in \mathbb{R}^n \mid x \geq 0, x \neq 0, Mx \geq 0\} \neq \emptyset.$$

It is obvious that every P-matrix is also a P_0-matrix and it is known [4] that every P_0-matrix is an S_0-matrix. We shall also need the following characterization of P_0-matrices [4].

Proposition 3. *A matrix $M \in \mathbb{R}^{n\times n}$ is a P_0-matrix iff for every nonzero vector x there exists an index i such that $x_i \neq 0$ and $x_i(Mx)_i \geq 0$.*

Definition 4. A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a

- P_0-function if, for every x and y in $\mathbb{R}^n$ with $x \neq y$, there is an index i such that

$$x_i \neq y_i, \qquad (x_i - y_i)[F_i(x) - F_i(y)] \geq 0.$$

- P-function if, for every x and y in $\mathbb{R}^n$ with $x \neq y$, there is an index i such that

$$(x_i - y_i)[F_i(x) - F_i(y)] > 0.$$

- Uniform P-function if there exists a positive constant μ such that, for every x and y in $\mathbb{R}^n$, there is an index i such that

$$(x_i - y_i)[F_i(x) - F_i(y)] \geq \mu\|x - y\|^2.$$

It is obvious that every uniform P-function is a P-function and that every P-function is a P_0-function. Furthermore, it is known that the Jacobian of every continuously differentiable P_0-function is a P_0-matrix and that if the Jacobian of a continuously differentiable function is a P-matrix for every x, then the function is a P-function. If F is affine, that is if $F(x) = Mx + q$, then F is a P_0-function iff M is a P_0-matrix, while F is a (uniform) P-function iff M is a P-matrix (note that in the affine case the concept of uniform P-function and P-function coincide). We finally note that every

monotone function is a P_0-function, every strictly monotone function is a P-function and that every strongly monotone function is a uniform P-function.

2.2. *Differentiability of functions and generalized Newton's method*

Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian; by Rademacher's theorem G is differentiable almost everywhere. If we indicate by D_G the set where G is differentiable, we can define the B-subdifferential of G at x [38] as

$$\partial_B G(x) := \left\{ H : \exists \{x^k\},\ x^k \in D_G, \quad \text{with } \lim_{x^k \to x} G'(x^k) = H \right\}$$

and the Clarke subdifferential of G at x as

$$\partial G(x) := co\ \partial_B G(x),$$

where *co* denotes the convex hull of a set.

Semismooth functions were introduced in [31] and immediately shown to be relevant to optimization algorithms. Recently the concept of semismoothness has been extended to vector valued functions [40].

Definition 5. Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian at $x \in \mathbb{R}^n$. We say that G is *semismooth* at x if

$$\lim_{\substack{H \in \partial G(x + tv') \\ v' \to v, t \downarrow 0}} Hv' \tag{1}$$

exists for all $v \in \mathbb{R}^n$.

Semismooth functions lie between Lipschitz functions and C^1 functions. Note that this class is strictly contained in the class of B-differentiable functions. Furthermore it is known that if G is semismooth at x then it is also directionally differentiable there, and its directional derivative in the direction v is given by (1).

A slightly stronger notion than semismoothness is strong semismoothness, defined below (see [38] and [40] where, however, different names are used).

Definition 6. Suppose that G is semismooth at x. We say that G is *strongly semismooth* at x if for any $H \in \partial G(x + d)$, and for any $d \to 0$,

$$Hd - G'(x; d) = O(\|d\|^2).$$

In the study of algorithms for the local solution of semismooth systems of equations, the following regularity condition plays a role similar to that of the nonsingularity of the Jacobian in the study of algorithms for smooth systems of equations.

Definition 7. We say that a semismooth function $G : \mathbb{R}^n \to \mathbb{R}^n$ is *BD-regular* at x if all the elements in $\partial_B G(x)$ are nonsingular.

We note that this condition was introduced by Qi in [38] under the name of strong BD-regularity. However, since this turned out to be a central notion, Qi himself called the same condition BD-regularity in a number of subsequent papers, see e.g. [37].

A generalized Newton method for the solution of a semismooth system of n equations $G(x) = 0$ can be defined as

$$x^{k+1} = x^k - (H^k)^{-1}G(x^k), \qquad H^k \in \partial_B G(x^k) \tag{2}$$

(H^k can be any element in $\partial_B G(x^k)$). The following result holds [38].

Theorem 8. *Suppose that x^* is a solution of the system $G(x) = 0$ and that G is semismooth and BD-regular at x^*. Then the iteration method (2) is well defined and convergent to x^* superlinearly in a neighborhood of x^*. If, in addition, G is directionally differentiable in a neighborhood of x^* and strongly semismooth at x^*, then the convergence rate of (2) is quadratic.*

We finally give the definition of a SC^1 function.

Definition 9. A function $f : \mathbb{R}^n \to \mathbb{R}$ is an SC^1 function if f is continuously differentiable and its gradient is semismooth.

SC^1 functions can be viewed as functions which lie between C^1 and C^2 functions.

2.3. *Reformulation of* NCP(F)

Our reformulation of NCP(F) as a system of equations is based on the following two variables convex function:

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

The most interesting property of this function is that, as it is easily verified,

$$\phi(a, b) = 0 \iff a \geq 0,\ b \geq 0,\ ab = 0; \tag{3}$$

note also that ϕ is continuously differentiable everywhere but in the origin. The function ϕ was introduced by Fischer [11] in 1992, since then it has attracted the attention of many researchers and it has proved to be a valuable tool in nonlinear complementarity theory [7,8,12,16,22,23,25,27,39,52]. An up-to-date review on the uses of the function ϕ can be found in [13].

Exploiting (3) it is readily seen that the nonlinear complementarity problem is equivalent to the following system of nonsmooth equations:

$$\Phi(x) := \begin{bmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_i, F_i(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{bmatrix} = 0.$$

It is then also obvious that the nonnegative function

$$\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2 = \frac{1}{2}\sum_{i=1}^{n} \phi(x_i, F_i(x))^2$$

is zero at a point x if and only if x is a solution of NCP(F), so that solving NCP(F) is equivalent to finding the unconstrained global solutions of the problem $\{\min \Psi(x)\}$. The following results have been proven in [7] or easily follow from [39, Lemma 3.3]. Part (c) has also been shown in the recent paper [22].

Theorem 10. *It holds that*

(a) Φ *is semismooth everywhere; furthermore, if every* F_i *is twice continuously differentiable with Lipschitz continuous Hessian, then* Φ *is strongly semismooth everywhere*;

(b) Ψ *is continuously differentiable; furthermore, if every* F_i *is an* SC^1 *function, then also* Ψ *is an* SC^1 *function*;

(c) *if* F *is a uniform P-function then the level sets of* Ψ *are bounded.*

3. The algorithm

In this section we present the algorithm for the solution of NCP(F). Roughly speaking this algorithm can be seen as an attempt to solve the semismooth system of equations $\Phi(x) = 0$ by using the generalized Newton's method described in the previous section (see (2)). To ensure global convergence, a linesearch is performed to minimize the smooth merit function Ψ; if the search direction generated according to (2) is not a "good" descent direction, we resort to the negative gradient of Ψ.

Global algorithm.

Data : $x^0 \in \mathbb{R}^n$, $\rho > 0$, $p > 2$, $\beta \in (0, 1/2)$, $\varepsilon \geq 0$.

Step 0: Set $k = 0$.

Step 1: (Stopping criterion). If $\|\nabla\Psi(x^k)\| \leq \varepsilon$ stop.

Step 2: (Search direction calculation). Select an element H^k in $\partial_B\Phi(x^k)$. Find the solution d^k of the system

$$H^k d = -\Phi(x^k). \tag{4}$$

If system (4) is not solvable or if the condition

$$\nabla\Psi(x^k)^{\mathrm{T}}d^k \leq -\rho\|d^k\|^p \tag{5}$$

is not satisfied, set $d^k = -\nabla\Psi(x^k)$.

Step 3: (Linesearch). Find the smallest $i^k = 0, 1, 2, \ldots$ such that

$$\Psi(x^k + 2^{-i^k}d^k) \leq \Psi(x^k) + \beta 2^{-i^k}\nabla\Psi(x^k)^{\mathrm{T}}d^k. \tag{6}$$

Set $x^{k+1} = x^k + 2^{-i^k}d^k$, $k \leftarrow k+1$ and go to Step 1.

We note that the above algorithm is virtually indistinguishable from a global algorithm for the solution of an F-differentiable system of equations. Formally, the only point which requires some care is the calculation of an element H belonging to $\partial_B\Phi(x^k)$ in Step 2. This, however, turns out to be an easy and cheap task, as it will be discussed in Section 7.

Another point which is worth of attention is that, if it exists, the direction obtained by solving (4) is always a descent direction for the function Ψ, unless $\Phi(x^k) = 0$. This is a standard property of the Newton direction for the solution of a *smooth* system of equations, but it is no longer true, in general, when the system of equations is nonsmooth. To see that this assertion is true, it is sufficient to note that, as it is stated in Theorem 10 (b), Ψ is differentiable at x^k; on the other hand, applying standard nonsmooth calculus rules (see [3]), we obtain

$$\partial\Psi(x^k) = \{\nabla\Psi(x^k)\} = V^{\mathrm{T}}\Phi(x^k) \quad \text{for every } V \in \partial\Phi(x^k);$$

so that the expression on the right hand side is independent of the element $V \in \partial\Phi(x^k)$ chosen. Hence we can take $V = H^k$ and write, taking into account (4),

$$\nabla\Psi(x^k)^{\mathrm{T}}d^k = \left(\Phi(x^k)^{\mathrm{T}}H^k\right)d^k = -\|\Phi(x^k)\|^2.$$

This fact clearly shows, once again, the similarity of our algorithm with Newton's method for the solution of a system of equations; furthermore, in our opinion, it also indicates that the direction (4) is a "good" search direction even when far from a solution of the system $\Phi(x) = 0$, which is not true for a general semismooth system.

The stopping criterion at Step 1 can be substituted by any other criterion without changing the properties of the algorithm. In what follows, as usual in analyzing the behaviour of algorithms, we shall assume that $\varepsilon = 0$ and that the algorithm produces an infinite sequence of points. The following result, summarizing the main properties of the algorithm, will be proved in Section 6, where we shall also discuss the properties of the algorithm in greater detail.

Theorem 11. *It holds that*

(a) *Each accumulation point of the sequence* $\{x^k\}$ *generated by the algorithm is a stationary point of* Ψ.

(b) *If one of the limit points of the sequence* $\{x^k\}$, *let us say* x^*, *is an isolated solution of NCP*(F), *then* $\{x^k\} \to x^*$.

(c) *If one of the limit points of the sequence* $\{x^k\}$, *let us say* x^*, *is a BD-regular solution of the system* $\Phi(x) = 0$, *and if each* F_i *is an* SC^1 *function, then* $\{x^k\} \to x^*$ *and*

(1) *eventually* d^k *is always given by the solution of system* (4) (*i.e., the negative gradient is never used eventually*),

(2) *eventually the stepsize of one is always accepted so that* $x^{k+1} = x^k + d^k$,

(3) *the convergence rate is superlinear; furthermore if each* F_i *is twice continuously differentiable with Lipschitz continuous Hessian, then the convergence rate is quadratic.*

The results stated in this theorem are somewhat "crude" and need to be completed by answering the following two questions. Under which conditions is a stationary point x^* of the function Ψ a global solution and hence a solution of NCP(F)? In the next section sufficient and necessary-and-sufficient conditions will be established in order to ensure this key property. The second question is: under which conditions is a solution of NCP(F) a BD-regular solution of the system $\Phi(x) = 0$? In fact, according to Theorem 11, a superlinear (at least) convergence rate of the algorithm can be ensured under the assumption of BD-regularity of solutions of the system $\Phi(x) = 0$. More in general, the possibility of using the (hopefully) good "second order" search direction (4) even far from a solution is closely related to this issue. Conditions guaranteeing the nonsingularity of all the elements in $\partial\Phi(x)$ will be analyzed in Section 5.

We note that, as discussed in [38], other strategies are possible to globalize the local Newton method (2) for the system $\Phi(x) = 0$. However the one we chose here, i.e., using the gradient of Ψ in "troublesome" situations, appears to be by far the simplest choice. The possibility of using the gradient, in turn, heavily depends on the surprising fact that $\|\Phi\|^2$ is continuously differentiable, which is not true, in general, for the square of a semismooth system of equations. This peculiarity of the system $\Phi(x) = 0$ paves the way for a simple extension of practically any standard globalization technique used in the solution of smooth systems of equations: Levenberg-Marquardt methods, trust region methods etc. In this paper we have chosen the simplest of these techniques since we were mainly interested in showing the potentialities of our approach and its numerical viability.

4. Regularity conditions

In this section we give some (necessary-and-) sufficient conditions for stationary points of our merit function Ψ to be solutions of NCP(F). We call these conditions regularity conditions. We first recall a result of Facchinei and Soares [7, Proposition 3.1].

Lemma 12. *For an arbitrary* $x \in \mathbb{R}^n$, *we have*

$$\partial\Phi(x)^T \subseteq D_a(x) + \nabla F(x) D_b(x),$$

where $D_a(x) = diag(a_1(x), \ldots, a_n(x))$, $D_b(x) = diag(b_1(x), \ldots, b_n(x)) \in \mathbb{R}^{n\times n}$ *are diagonal matrices whose ith diagonal element is given by*

$$a_i(x) = \frac{x_i}{\|(x_i, F_i(x))\|} - 1, \qquad b_i(x) = \frac{F_i(x)}{\|(x_i, F_i(x))\|} - 1,$$

if $(x_i, F_i(x)) \neq 0$, *and by*

$$a_i(x) = \xi_i - 1,\ b_i(x) = \rho_i - 1 \quad \textit{for every } (\xi_i, \rho_i) \in \mathbb{R}^2 \textit{ such that } \|(\xi_i, \rho_i)\| \leq 1,$$

if $(x_i, F_i(x)) = 0$.

In the following, for simplicity, we sometimes suppress the dependence on x in our notation. For example, the gradient of the differentiable function Ψ can be written as follows:

$$\nabla\Psi(x) = D_a(x)\Phi(x) + \nabla F(x) D_b(x)\Phi(x) = D_a\Phi + \nabla F D_b\Phi.$$

In the analysis of this section, the signs of the vectors $D_a\Phi \in \mathbb{R}^n$ and $D_b\Phi \in \mathbb{R}^n$ play an important role. We therefore introduce the index sets

$$\mathcal{C} := \{i \in I \mid x_i \geq 0,\ F_i(x) \geq 0,\ x_i F_i(x) = 0\} \quad (\underline{c}\text{omplementary indices}),$$
$$\mathcal{R} := I \setminus \mathcal{C} \quad (\underline{r}\text{esidual indices}),$$

and further partition the index set $\mathcal{R}$ as follows:

$$\mathcal{P} := \{i \in \mathcal{R} \mid x_i > 0,\ F_i(x) > 0\} \quad (\underline{p}\text{ositive indices}),$$
$$\mathcal{N} := \mathcal{R} \setminus \mathcal{P} \quad (\underline{n}\text{egative indices}).$$

Note that these index sets depend on x, but that the notation does not reflect this dependence. However, this should not cause any confusion because the given vector $x \in \mathbb{R}^n$ will always be clear from the context.

The names $\mathcal{P}$ and $\mathcal{N}$ of the above index sets are motivated by the following simple relations, which can be easily verified:

$$\begin{aligned} &(D_a\Phi)_i > 0 \Longleftrightarrow (D_b\Phi)_i > 0 \Longleftrightarrow i \in \mathcal{P}, \\ &(D_a\Phi)_i = 0 \Longleftrightarrow (D_b\Phi)_i = 0 \Longleftrightarrow i \in \mathcal{C}, \\ &(D_a\Phi)_i < 0 \Longleftrightarrow (D_b\Phi)_i < 0 \Longleftrightarrow i \in \mathcal{N}; \end{aligned} \tag{7}$$

in particular, the signs of the corresponding elements of $D_a\Phi$ and $D_b\Phi$ are the same.

The following definition of regular vector is motivated by some similar definitions in the recent papers of Pang and Gabriel [36, Definition 1], Moré [32, Definition 3.1] and Ferris and Ralph [10, Definition 2.4].

Definition 13. A point $x \in \mathbb{R}^n$ is called *regular* if for every vector $z \neq 0$ such that

$$z_{\mathcal{C}} = 0, \qquad z_{\mathcal{P}} > 0, \qquad z_{\mathcal{N}} < 0, \tag{8}$$

there exists a vector $y \in \mathbb{R}^n$ such that

$$y_{\mathcal{P}} \geq 0, \qquad y_{\mathcal{N}} \leq 0, \qquad y_{\mathcal{R}} \neq 0$$

and

$$y^T \nabla F(x) z \geq 0. \tag{9}$$

It is difficult to compare our definition of regular vector with the corresponding ones in [10,32,36] since we use different index sets (the definition of the index sets depends heavily on the merit function chosen). Nevertheless, we think that the following points should be remarked:

(a) As pointed out by Moré [32], the definition of regularity given by Pang and Gabriel [36] depends on the scaling of the function F, i.e., a vector $x \in \mathbb{R}^n$ could be regular for NCP(F) in the sense of Pang and Gabriel [36], but not for the equivalent problem NCP(F_s), where $F_s(x) = sF(x)$ for some positive scaling parameter s. Obviously our definition is independent of the scaling of F.

(b) In the related definition of regularity given by Moré [32] and Ferris and Ralph [10], a condition similar to (9) is employed. However, they have to assume $y^T \nabla F(x) z > 0$ instead of (9). In this respect, our definition of regularity seems to be weaker. For example, if $\nabla F(x)$ is a positive semidefinite matrix, then we can choose $y = z$ and directly obtain from Definition 13 that x is a regular vector. More in general, whenever directly comparable, our regularity conditions appear to be weaker than those introduced in [10,32,36].

The following result shows why the notion of regular vector is so important.

Theorem 14. *A vector $x \in \mathbb{R}^n$ is a solution of NCP(F) if and only if x is a regular stationary point of Ψ.*

Proof. First assume that $x \in \mathbb{R}^n$ is a solution of NCP(F). Then x is a global minimum of Ψ and hence a stationary point of Ψ by the differentiability of this function. Moreover, $\mathcal{P} = \mathcal{N} = \emptyset$ in this case, and therefore the regularity of x holds vacuously since $z = z_{\mathcal{C}}$, and there exists no nonzero vector z satisfying conditions (8).

Suppose now that x is regular and that $\nabla \Psi(x) = 0$. As mentioned at the beginning of this section, the stationary condition can be rewritten as

$$D_a \Phi + \nabla F(x) D_b \Phi = 0.$$

Consequently, we have

$$y^T (D_a \Phi) + y^T \nabla F(x) (D_b \Phi) = 0 \tag{10}$$

for any $y \in \mathbb{R}^n$. Assume that x is not a solution of NCP(F). Then $\mathcal{R} \neq \emptyset$ and hence, by (7), $z := D_b \Phi$ is a nonzero vector with

$$z_{\mathcal{C}} = 0, \qquad z_{\mathcal{P}} > 0, \qquad z_{\mathcal{N}} < 0.$$

Recalling that the components of $D_a\Phi$ and $z = D_b\Phi$ have the same signs, and taking $y \in \mathbb{R}^n$ from the definition of regular vector, we have

$$y^{\mathrm{T}}(D_a\Phi) = y_{\mathcal{C}}^{\mathrm{T}}(D_a\Phi)_{\mathcal{C}} + y_{\mathcal{P}}^{\mathrm{T}}(D_a\Phi)_{\mathcal{P}} + y_{\mathcal{N}}^{\mathrm{T}}(D_a\Phi)_{\mathcal{N}} > 0 \tag{11}$$

(since $y_{\mathcal{R}} \neq 0$) and

$$y^{\mathrm{T}}\nabla F(x)(D_b\Phi) = y^{\mathrm{T}}\nabla F(x)z \geq 0. \tag{12}$$

The inequalities (11) and (12) together, however, contradict condition (10). Hence $\mathcal{R} = \emptyset$. This means that x is a solution of NCP(F). □

From Theorem 14 and the remark (b) after Definition 13, we directly obtain that a stationary point x is a solution of NCP(F) if the Jacobian matrix $F'(x)$ is positive semidefinite. In particular, we have the result that all stationary points of Ψ are solutions of NCP(F) for monotone functions F, i.e., we reobtain in this way a recent result of Geiger and Kanzow [16, Theorem 2.5].

In the remaining part of this section we shall investigate weaker sufficient conditions which ensure regularity of a point. In light of Theorem 14 these conditions also ensure that stationary points of the merit function Ψ are global minimizers of Ψ and hence solutions of NCP(F).

Let $T \in \mathbb{R}^{|\mathcal{R}|\times|\mathcal{R}|}$ denote a diagonal matrix with diagonal entries given by

$$T_{ii} := \begin{cases} +1 & \text{if } i \in \mathcal{P}, \\ -1 & \text{if } i \in \mathcal{N} \end{cases}$$

and note that $TT = I$. Using this notation, we can prove the following result.

Theorem 15. *Let $x \in \mathbb{R}^n$ be a vector such that the matrix $TF'(x)_{\mathcal{R}\mathcal{R}}T$ is an S_0-matrix. Then x is a regular point.*

Proof. By the definition of S_0-matrix and the assumptions made, there exists a nonzero vector $\tilde{y}_{\mathcal{R}}$ such that

$$\tilde{y}_{\mathcal{R}} \geq 0 \quad \text{and} \quad TF'(x)_{\mathcal{R}\mathcal{R}}T\tilde{y}_{\mathcal{R}} \geq 0. \tag{13}$$

Let y be the unique vector such that

$$y_{\mathcal{C}} = 0, \qquad y_{\mathcal{R}} = T\tilde{y}_{\mathcal{R}}; \tag{14}$$

by the definition of T and (13) we have that $y_{\mathcal{R}} \neq 0$ and

$$y_{\mathcal{P}} \geq 0, \qquad y_{\mathcal{N}} \leq 0.$$

Then it is easy to see that, for every $z \in \mathbb{R}^n$ such that $z \neq 0$ and

$$z_{\mathcal{C}} = 0, \qquad z_{\mathcal{P}} > 0, \qquad z_{\mathcal{N}} < 0,$$

we can write, recalling (14),

$$\begin{aligned} y^{\mathrm{T}}\nabla F(x)z &= y_{\mathcal{R}}^{\mathrm{T}}\nabla F(x)_{\mathcal{R}\mathcal{R}}z_{\mathcal{R}} \\ &= y_{\mathcal{R}}^{\mathrm{T}}(TT)\nabla F(x)_{\mathcal{R}\mathcal{R}}(TT)z_{\mathcal{R}} \\ &= (y_{\mathcal{R}}^{\mathrm{T}}T)(T\nabla F(x)_{\mathcal{R}\mathcal{R}}T)(Tz_{\mathcal{R}}) \\ &= \tilde{y}_{\mathcal{R}}^{\mathrm{T}}T(F'(x)_{\mathcal{R}\mathcal{R}})^{\mathrm{T}}T(Tz_{\mathcal{R}}) \geq 0, \end{aligned} \tag{15}$$

where the last inequality follows by (13) and the fact that $(Tz_{\mathcal{R}}) > 0$. □

In the following corollary, we summarize some simple, but noteworthy consequences of Theorem 15. First recall that a vector $x \in \mathbb{R}^n$ is called feasible for NCP(F) if $x \geq 0$ and $F(x) \geq 0$.

Corollary 16. *It holds that*

(a) *If $x \in \mathbb{R}^n$ is a vector such that the submatrix $F'(x)_{\mathcal{R}\mathcal{R}}$ is a P_0-matrix, then x is a regular point.*

(b) *If $F : \mathbb{R}^n \to \mathbb{R}^n$ is a P_0-function, then all vectors $x \in \mathbb{R}^n$ are regular points.*

(c) *If $x \in \mathbb{R}^n$ is a feasible vector such that the submatrix $F'(x)_{\mathcal{R}\mathcal{R}}$ is an S_0-matrix, then x is a regular point.*

Proof.

(a) Since a square matrix M is a P_0-matrix if and only if the matrix DMD is a P_0-matrix for all nonsingular diagonal matrices D, the assertion is a direct consequence of Theorem 15 and the fact that every P_0-matrix is also an S_0-matrix.

(b) Since F is a continuously differentiable P_0-function, its Jacobian $F'(x)$ is a P_0-matrix for all $x \in \mathbb{R}^n$, see Moré and Rheinboldt [33, Theorem 5.8]. The second assertion is therefore a direct consequence of part (a) since, by definition, all principal submatrices of a P_0-matrix are also P_0-matrices.

(c) Since x is assumed to be a feasible vector, we have $\mathcal{N} = \emptyset$. Hence the diagonal matrix T introduced before Theorem 15 reduces to the identity matrix, and part (c) follows directly from Theorem 15. □

We finally prove a result which includes and generalizes the sufficient conditions for regularity obtained so far. To this end we introduce some further notation. Let $\tilde{\mathcal{C}}$ be such that $\emptyset \subseteq \tilde{\mathcal{C}} \subseteq \mathcal{C}$. We consider the following partitioned submatrix of the Jacobian $F'(x)$:

$$\tilde{F}'(x) = \begin{pmatrix} F'(x)_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}} & F'(x)_{\tilde{\mathcal{C}}\mathcal{P}} & F'(x)_{\tilde{\mathcal{C}}\mathcal{N}} \\ F'(x)_{\mathcal{P}\tilde{\mathcal{C}}} & F'(x)_{\mathcal{P}\mathcal{P}} & F'(x)_{\mathcal{P}\mathcal{N}} \\ F'(x)_{\mathcal{N}\tilde{\mathcal{C}}} & F'(x)_{\mathcal{N}\mathcal{P}} & F'(x)_{\mathcal{N}\mathcal{N}} \end{pmatrix}$$

and define

$$\tilde{J}(x) := \tilde{T}\tilde{F}'(x)\tilde{T},$$

where the transformation matrix $\tilde{T}$ has the diagonal structure

$$\tilde{T} := \begin{pmatrix} +I_{\tilde{\mathcal{C}}} & & 0 \\ & +I_{\mathcal{P}} & \\ 0 & & -I_{\mathcal{N}} \end{pmatrix}. \tag{16}$$

Then we have

$$\begin{aligned} \tilde{J}(x) &= \begin{pmatrix} F'(x)_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}} & F'(x)_{\tilde{\mathcal{C}}\mathcal{P}} & -F'(x)_{\tilde{\mathcal{C}}\mathcal{N}} \\ F'(x)_{\mathcal{P}\tilde{\mathcal{C}}} & F'(x)_{\mathcal{P}\mathcal{P}} & -F'(x)_{\mathcal{P}\mathcal{N}} \\ -F'(x)_{\mathcal{N}\tilde{\mathcal{C}}} & -F'(x)_{\mathcal{N}\mathcal{P}} & F'(x)_{\mathcal{N}\mathcal{N}} \end{pmatrix} \\ &= \begin{pmatrix} F'(x)_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}} & \tilde{J}(x)_{\tilde{\mathcal{C}}\mathcal{R}} \\ \tilde{J}(x)_{\mathcal{R}\tilde{\mathcal{C}}} & \tilde{J}(x)_{\mathcal{R}\mathcal{R}} \end{pmatrix}, \end{aligned} \tag{17}$$

where, we recall, $\mathcal{R} = \mathcal{P} \cup \mathcal{N}$. Using this matrix, we can prove the following result.

Theorem 17. *Let $x \in \mathbb{R}^n$ be given and assume that the submatrix $F'(x)_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}}$ is nonsingular and the Schur-complement of this matrix in $\tilde{J}(x)$ is an S_0-matrix. Then x is a regular point.*

Proof. We first show that there exist $\tilde{y}_{\mathcal{P}}, \tilde{y}_{\mathcal{N}} \geq 0$ with $\tilde{y}_{\mathcal{R}} \neq 0$ and $q_{\mathcal{P}}, q_{\mathcal{N}} \geq 0$ such that

$$\tilde{J}(x)\tilde{y} = q, \tag{18}$$

where $\tilde{y} = (\tilde{y}_{\tilde{\mathcal{C}}}, \tilde{y}_{\mathcal{P}}, \tilde{y}_{\mathcal{N}})^{\mathrm{T}}$, $q = (0_{\tilde{\mathcal{C}}}, q_{\mathcal{P}}, q_{\mathcal{N}})^{\mathrm{T}}$. In view of (17), system (18) can be rewritten as

$$F'(x)_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}}\tilde{y}_{\tilde{\mathcal{C}}} + \tilde{J}(x)_{\tilde{\mathcal{C}}\mathcal{R}}\tilde{y}_{\mathcal{R}} = 0_{\tilde{\mathcal{C}}}, \tag{19}$$

$$\tilde{J}(x)_{\mathcal{R}\tilde{\mathcal{C}}}\tilde{y}_{\tilde{\mathcal{C}}} + \tilde{J}(x)_{\mathcal{R}\mathcal{R}}\tilde{y}_{\mathcal{R}} = q_{\mathcal{R}}. \tag{20}$$

Solving the first equation for $\tilde{y}_{\tilde{\mathcal{C}}}$ yields

$$\tilde{y}_{\tilde{\mathcal{C}}} = -F'(x)^{-1}_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}}\tilde{J}(x)_{\tilde{\mathcal{C}}\mathcal{R}}\tilde{y}_{\mathcal{R}}. \tag{21}$$

Substituting this into (20) and rearranging leads to

$$\left(\tilde{J}(x)_{\mathcal{R}\mathcal{R}} - \tilde{J}(x)_{\mathcal{R}\tilde{\mathcal{C}}}F'(x)^{-1}_{\tilde{\mathcal{C}}\tilde{\mathcal{C}}}\tilde{J}(x)_{\tilde{\mathcal{C}}\mathcal{R}}\right)\tilde{y}_{\mathcal{R}} = q_{\mathcal{R}}. \tag{22}$$

Since, by assumption, the matrix of the linear system (22) is an S_0-matrix, there exists an $\tilde{y}_{\mathcal{R}} \geq 0, \tilde{y}_{\mathcal{R}} \neq 0$, and a $q_{\mathcal{R}} \geq 0$ satisfying this system. Then define $\tilde{y}_{\mathcal{C}}$ by (21) and set $y := T\tilde{y}$, where $\tilde{y} \in \mathbb{R}^n$ and the nonsingular diagonal matrix $T \in \mathbb{R}^{n\times n}$ are obtained from $\tilde{y}$ and $\tilde{T}$ respectively by setting

$$\bar{y}_i = \tilde{y}_i \quad \text{if } i \in \tilde{\mathcal{C}} \cup \mathcal{R}, \quad \bar{y}_i = 0 \text{ otherwise},$$

$$T_{ii} = \tilde{T}_{ii} \quad \text{if } i \in \tilde{\mathcal{C}} \cup \mathcal{R}, \quad T_{ii} = 1 \text{ otherwise}.$$

By definition we have $y_{\mathcal{P}} = \bar{y}_{\mathcal{P}} = \tilde{y}_{\mathcal{P}} \geq 0$, $y_{\mathcal{N}} = -\bar{y}_{\mathcal{N}} = -\tilde{y}_{\mathcal{N}} \leq 0$ and $y_{\mathcal{R}} \neq 0$, so that the vector $y \in \mathbb{R}^n$ satisfies all the conditions required in Definition 13 for a regular

vector x. Now let $z \neq 0$ be an arbitrary vector with $z_{\mathcal{C}} = 0, z_{\mathcal{P}} > 0$ and $z_{\mathcal{N}} < 0$, then $\bar{z} := Tz$ satisfies the conditions $\bar{z}_{\mathcal{C}} = 0_{\mathcal{C}}, \bar{z}_{\mathcal{P}} > 0, \bar{z}_{\mathcal{N}} > 0$. Therefore, define $\tilde{z} \in \mathbb{R}^{|\tilde{\mathcal{C}} \cup \mathcal{R}|}$ by setting $\tilde{z}_i = \bar{z}_i$ for $i \in \tilde{\mathcal{C}} \cup \mathcal{R}$. Then, since $\bar{z}_i = \bar{y}_i = 0$ for $i \in \mathcal{C} \setminus \tilde{\mathcal{C}}$, and since $T = T^{-1}$, we obtain from (18):

$$\begin{aligned} y^{\mathrm{T}} \nabla F(x) z &= z^{\mathrm{T}} F'(x) y \\ &= z^{\mathrm{T}} TTF'(x) TTy \\ &= \bar{z}^{\mathrm{T}} (TF'(x)T) \bar{y} \\ &= \tilde{z}^{\mathrm{T}} \tilde{T} \tilde{F}'(x) \tilde{T} \tilde{y} \\ &= \tilde{z}^{\mathrm{T}} \tilde{J}(x) \tilde{y} \\ &= \tilde{z}^{\mathrm{T}} q \\ &= \tilde{z}_{\mathcal{P}}^{\mathrm{T}} q_{\mathcal{P}} + \tilde{z}_{\mathcal{N}}^{\mathrm{T}} q_{\mathcal{N}} \\ &\geq 0, \end{aligned}$$

i.e., x is regular. □

If $\tilde{\mathcal{C}} = \emptyset$, then Theorem 17 reduces to Theorem 15. If $\tilde{\mathcal{C}} = \mathcal{C}$ we obtain a result similar to the ones in [10,32,36]. Note however that in all these papers a certain Schur-complement is assumed to be an S-matrix, which, again, is a stronger assumption than the one used in our Theorem 17.

In order to illustrate our theory, consider NCP(F) with $F : \mathbb{R}^4 \to \mathbb{R}^4$ being defined by

$$F(x) := \begin{pmatrix} 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\ 2x_1^2 + x_1 + x_2^2 + 3x_3 + 2x_4 - 2 \\ 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1 \\ x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3 \end{pmatrix}.$$

This small example is due to Kojima [28] and is a standard test problem for nonlinear complementarity codes. Harker and Xiao [20] report a failure of their method for this example. Their method converges to the point

$$\hat{x} = (1.0551, 1.3347, 1.2681, 0)^{\mathrm{T}},$$

where

$$F(\hat{x}) = (4.9873, 6.8673, 9.8472, 5.9939)^{\mathrm{T}},$$

which is obviously not a solution of NCP(F). The corresponding Jacobian matrix is

$$F'(\hat{x}) = \begin{pmatrix} 9.0001 & 7.4491 & 1 & 3 \\ 5.2204 & 2.6695 & 3 & 2 \\ 7.6653 & 6.3940 & 2 & 3 \\ 2.1102 & 8.0084 & 2 & 3 \end{pmatrix}.$$

Obviously, $F'(\hat{x})$ is an S_0-matrix. Since $\hat{x}$ is also feasible, Corollary 16 (c) guarantees that our method will not terminate at $\hat{x}$.

We note that our conditions for a stationary point to be a solution of NCP(F) are very weak. Similar results, which are almost as strong as those given in this section, have been shown only by Moré [32], who considers the bound constrained reformulation

$$\min \frac{1}{2}\|h(x,y)\|^2 \quad \text{s.t. } x \geq 0, y \geq 0 \tag{23}$$

of NCP(F), where

$$h(x,y) := \begin{pmatrix} F(x) - y \\ Yx \end{pmatrix}$$

and $Y := diag(y_1, \ldots, y_n) \in \mathbb{R}^{n\times n}$. Moré proves that if $z^* := (x^*, y^*)$ is a stationary (i.e., Karush-Kuhn-Tucker) point of (23) such that a certain submatrix of the Jacobian $F'(x^*)$ is a *nonsingular* P_0-matrix, then x^* is a solution of NCP(F). The following example shows that this result is not true without this nonsingularity condition: let $n = 2$ and define $F : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(x) := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} -5/4 \\ 1 \end{pmatrix} = \begin{pmatrix} x_2 - 5/4 \\ 1 \end{pmatrix}.$$

Obviously, $F'(x)$, along with all its principal submatrices, is a singular P_0-matrix for all $x \in \mathbb{R}^2$, and it is easy to verify that the vector $z^* = (x^*, y^*) := (0, 1, 0, 1/2)$ is a stationary point of (23). However, x^* is not a solution of NCP(F) since $x_2^* F_2(x^*) = 1$. On the other hand, by Corollary 16 (b) x^* is a regular point and therefore cannot be a stationary point of Ψ.

5. Nonsingularity conditions

In this section we study conditions which guarantee that all elements in $\partial\Phi(x)$ are nonsingular.

We begin with three lemmas that will be needed later in this section.

Lemma 18. *If $M \in \mathbb{R}^{n\times n}$ is a P_0-matrix, then every matrix of the form*

$$D_a + D_b M$$

is nonsingular for all positive definite (negative definite) diagonal matrices $D_a, D_b \in \mathbb{R}^{n\times n}$.

Proof. We only consider the case in which the two matrices are positive definite, the other case is analogous. Assume that M is a P_0-matrix and $D_a := diag(a_1, \ldots, a_n)$, $D_b := diag(b_1, \ldots, b_n)$ are positive definite diagonal matrices. Then $(D_a + D_b M)q = 0$ for some $q \in \mathbb{R}^n$ implies $D_a q = -D_b M q$ and therefore $q_i = -(b_i/a_i)(Mq)_i$ $(i \in I)$. Since

M is a P_0-matrix, we get from Proposition 3 that $q = 0$, i.e., the matrix $D_a + D_bM$ is nonsingular. □

We note that it is also possible to prove the converse direction in the above lemma, i.e., the matrix $D_a + D_bM$ is nonsingular for all positive (negative) definite diagonal matrices $D_a, D_b \in \mathbb{R}^{n\times n}$ if and only if M is a P_0-matrix. In the following, however, we only need the direction stated in Lemma 18.

Lemma 19. *Let M be any matrix and consider the following partition*

$$M = \begin{pmatrix} M_{L,L} & M_{L,J} \\ M_{J,L} & M_{J,J} \end{pmatrix}.$$

Assume that

(a) *$M_{L,L}$ is nonsingular,*

(b) *$M/M_{L,L}$ is a P-matrix (P_0-matrix),*

then, for every index set K such that $L \subseteq K \subseteq L \cup J$, $M_{K,K}/M_{L,L}$ is a P-matrix (P_0-matrix).

Proof. The assertion easily follows by the fact that $M_{K,K}/M_{L,L}$ is a principal submatrix of $M/M_{L,L}$ which, in turn, is a P-matrix (P_0-matrix). □

We now introduce some more index sets:

$$\alpha := \{i \in I \mid x_i > 0, F_i(x) = 0\},$$
$$\beta := \{i \in I \mid x_i = 0, F_i(x) = 0\},$$
$$\gamma := \{i \in I \mid x_i = 0, F_i(x) > 0\},$$
$$\delta := I \setminus \{\alpha \cup \beta \cup \gamma\}$$

(once again, we suppress in our notation the dependence on x of the above index sets). If the point x under consideration is a solution of NCP(F) then $\delta = \emptyset$ and the sets α, β and γ coincide with those already introduced in Section 2. Using the notation of the previous section, the index sets α, β and γ form a partition of the set C, whereas δ corresponds to what has been called $\mathcal{R}$ in Section 4 (here we changed the name of this set just for uniformity of notation). Note that β denotes the set of degenerate indices at which Φ is not differentiable. Further note that at an arbitrary point $x \in \mathbb{R}^n$, the index sets α, γ and, in particular, β are usually very small (and quite often empty).

The last lemma is introduced in order to simplify the proof of the main theorem (Theorem 21). It gives conditions which guarantee that all elements in $\partial\Phi(x)$ are nonsingular, assuming $\beta = \emptyset$. This result will then be used to prove the general case $\beta \neq \emptyset$.

Lemma 20. *Let $x \in \mathbb{R}^n$ be given and set $M := F'(x)$. Assume that*

(a) $\beta(x) = \emptyset$,

(b) $M_{\alpha,\alpha}$ *is nonsingular,*

(c) *the Schur-complement* $M_{\alpha\cup\delta,\alpha\cup\delta}/M_{\alpha,\alpha}$ *is a* P_0*-matrix.*

Then the function Φ *is differentiable at* x *and the Jacobian matrix* $\Phi'(x)$ *is nonsingular.*

Proof. Since $\beta(x) = \emptyset$, the function Φ is differentiable at x with $\Phi'(x) = D_a + D_bM$ where D_a and D_b are the diagonal matrices introduced in Lemma 12. Let $q \in \mathbb{R}^n$ be an arbitrary vector with $(D_a + D_bM)q = 0$. Writing

$$D_a = \begin{pmatrix} D_{a,\alpha} & & \\ & D_{a,\gamma} & \\ & & D_{a,\delta} \end{pmatrix},$$

$$D_b = \begin{pmatrix} D_{b,\alpha} & & \\ & D_{b,\gamma} & \\ & & D_{b,\delta} \end{pmatrix},$$

$$M = \begin{pmatrix} M_{\alpha,\alpha} & M_{\alpha,\gamma} & M_{\alpha,\delta} \\ M_{\gamma,\alpha} & M_{\gamma,\gamma} & M_{\gamma,\delta} \\ M_{\delta,\alpha} & M_{\delta,\gamma} & M_{\delta,\delta} \end{pmatrix},$$

$$q = (q_\alpha, q_\gamma, q_\delta)^T,$$

where $D_{a,\alpha}, D_{a,\gamma}, \ldots$ are abbreviations for the diagonal matrices $(D_a)_{\alpha,\alpha}, (D_a)_{\gamma,\gamma}, \ldots$, the equation $(D_a + D_bM)q = 0$ can be rewritten as

$$M_{\alpha,\alpha}q_\alpha + M_{\alpha,\gamma}q_\gamma + M_{\alpha,\delta}q_\delta = 0_\alpha, \tag{24}$$

$$D_{a,\gamma}q_\gamma = 0_\gamma, \tag{25}$$

$$D_{a,\delta}q_\delta + D_{b,\delta}M_{\delta,\alpha}q_\alpha + D_{b,\delta}M_{\delta,\gamma}q_\gamma + D_{b,\delta}M_{\delta,\delta}q_\delta = 0_\delta; \tag{26}$$

where we have taken into account that, by Lemma 12, $D_{a,\alpha} = 0_{\alpha,\alpha}, D_{b,\gamma} = 0_{\gamma,\gamma}$ and that $D_{b,\alpha}$ is nonsingular. Since the diagonal matrix $D_{a,\gamma}$ is also nonsingular, we directly obtain

$$q_\gamma = 0_\gamma \tag{27}$$

from (25). Hence Eqs. (24) and (26) reduce to

$$M_{\alpha,\alpha}q_\alpha + M_{\alpha,\delta}q_\delta = 0_\alpha, \tag{28}$$

$$D_{a,\delta}q_\delta + D_{b,\delta}M_{\delta,\alpha}q_\alpha + D_{b,\delta}M_{\delta,\delta}q_\delta = 0_\delta. \tag{29}$$

Due to the nonsingularity of the submatrix $M_{\alpha,\alpha}$, we directly obtain from (28)

$$q_\alpha = -M_{\alpha,\alpha}^{-1}M_{\alpha,\delta}q_\delta. \tag{30}$$

Substituting this into (29) and rearranging terms yields

$$\left(D_{a,\delta} + D_{b,\delta}\left[M_{\delta,\delta} - M_{\delta,\alpha}M_{\alpha,\alpha}^{-1}M_{\alpha,\delta}\right]\right)q_\delta = 0_\delta. \tag{31}$$

According to the negative definiteness of the diagonal matrices $D_{a,\delta}$ and $D_{b,\delta}$, we obtain from assumption (c), Eq. (31) and Lemma 18 that $q_\delta = 0_\delta$. This, in turn, implies that

$q_\alpha = 0_\alpha$ by (30). In view of (27), we therefore have $q = 0$, which proves the desired result. □

Theorem 21. *Let $x \in \mathbb{R}^n$ be given, and set $M := F'(x)$. Assume that*

(a) *the submatrices $M_{\tilde{\alpha},\tilde{\alpha}}$ are nonsingular for all $\alpha \subseteq \tilde{\alpha} \subseteq (\alpha \cup \beta)$,*

(b) *the Schur-complement $M_{\alpha\cup\beta\cup\delta,\alpha\cup\beta\cup\delta}/M_{\alpha,\alpha}$ is a P_0-matrix.*

Then all elements $G \in \partial\Phi(x)$ are nonsingular.

Proof. We first note that by Lemma 19 and the assumptions made, the Schur-complement $M_{\alpha\cup\hat{\delta},\alpha\cup\hat{\delta}}/M_{\alpha,\alpha}$ is a P_0-matrix for every $\delta \subseteq \hat{\delta} \subseteq (\delta \cup \beta)$.

Let us now consider $\alpha \subseteq \tilde{\alpha} \subseteq (\alpha \cup \beta)$. We prove that the Schur-complement

$$M_{\tilde{\alpha}\cup\tilde{\delta},\tilde{\alpha}\cup\tilde{\delta}}/M_{\tilde{\alpha},\tilde{\alpha}}$$

is a P_0-matrix for all $\delta \subseteq \tilde{\delta} \subseteq \delta \cup [(\beta \cup \alpha) \setminus \tilde{\alpha}]$. By assumption (a) and the quotient formula for Schur-complements (see Cottle, Pang and Stone [4, Proposition 2.3.6]), we have that $M_{\tilde{\alpha},\tilde{\alpha}}/M_{\alpha,\alpha}$ is nonsingular and

$$M_{\tilde{\alpha}\cup\tilde{\delta},\tilde{\alpha}\cup\tilde{\delta}}/M_{\tilde{\alpha},\tilde{\alpha}} = \left(M_{\tilde{\alpha}\cup\tilde{\delta},\tilde{\alpha}\cup\tilde{\delta}}/M_{\alpha,\alpha}\right) / \left(M_{\tilde{\alpha},\tilde{\alpha}}/M_{\alpha,\alpha}\right). \tag{32}$$

By the first part of the proof, the matrix $M_{\tilde{\alpha}\cup\tilde{\delta},\tilde{\alpha}\cup\tilde{\delta}}/M_{\alpha,\alpha}$ is a P_0-matrix. Hence, recalling that Lemma 2.3 of Chen and Harker [2] states that the Schur-complement of a nonsingular principal minor of a P_0-matrix is also a P_0-matrix, we have that the right-hand side of (32) is a P_0-matrix. Therefore the matrix on the left-hand side of (32) is also a P_0-matrix.

Now, recalling that β denotes the set of degenerate indices at which $(x_i, F_i(x)) = 0$, and considering $(\xi_i, \rho_i) \in \mathbb{R}^2$ such that $\|(\xi_i, \rho_i)\| \leq 1$ as in Lemma 12, let us partition the index set β into the following three subsets:

$$\beta_1 := \{i \in \beta \mid \xi_i > 0, \rho_i = 0\},$$
$$\beta_2 := \{i \in \beta \mid \xi_i = 0, \rho_i > 0\},$$
$$\beta_3 := \beta \setminus (\beta_1 \cup \beta_2),$$

and define $\tilde{\alpha} := \alpha \cup \beta_1$, $\tilde{\gamma} := \gamma \cup \beta_2$ and $\tilde{\delta} := \delta \cup \beta_3$. It is now very easy to see that the nonsingularity of any element of $\partial\Phi(x)$ can be proved following exactly the lines of the proof of Lemma 20 by replacing the index sets α, γ and δ by $\tilde{\alpha}, \tilde{\gamma}$ and $\tilde{\delta}$, respectively, and taking into account assumptions (a) and (b). □

In view of Lemma 19 the assumptions (c) of Lemma 20 and (b) of Theorem 21 are satisfied, e.g., if the Schur-complement $M/M_{\alpha,\alpha}$ is a P_0-matrix. Due to Chen and Harker [2, Lemma 2.3], Schur-complements of nonsingular submatrices of a P_0-matrix are again P_0-matrices, hence assumption (b) of Theorem 21 is satisfied in particular if $F'(x)$ is a P_0-matrix. In turn it is known, see [33, Theorem 5.8], that $F'(x)$ is a P_0-matrix if F itself is a P_0-function.

In the next two corollaries we point out two simple situations in which the assumptions of the previous Theorem are satisfied. The first of the two corollaries was already proved by Facchinei and Soares [8, Proposition 3.2].

Corollary 22. *Let $x^* \in \mathbb{R}^n$ be an R-regular solution of NCP(F). Then all the elements of the generalized Jacobian $\partial\Phi(x^*)$ are nonsingular.*

Proof. We prove that assumptions (a) and (b) of Theorem 21 are satisfied. From the definition of R-regularity we have that $M_{\alpha,\alpha}$ is nonsingular and that the Schur-complement of $M_{\alpha,\alpha}$ in $M_{\alpha\cup\beta,\alpha\cup\beta}$ is a P-matrix. By [35, Lemma 1], these conditions directly imply assumption (a). Since x^* is a solution of NCP(F), we also have $\delta(x^*) = \emptyset$, so that assumption (b) is obvious by the R-regularity. □

Due to the upper semi-continuity of the generalized Jacobian (see Clarke [3, Proposition 2.6.2 (c)]), Corollary 22 remains true for all x in a sufficiently small neighbourhood of an R-regular solution x^* of NCP(F).

Corollary 23. *Suppose that $F'(x)$ is a P-matrix. Then all the elements of the generalized Jacobian $\partial\Phi(x)$ are nonsingular.*

Proof. Since every principal submatrix of a P-matrix is nonsingular and the Schur-complement of every principal submatrix of a P-matrix is a P-matrix, we obviously have that the assumptions of Theorem 21 are satisfied. □

Jiang and Qi [23, Proposition 1] recently proved that all elements of the generalized Jacobian $\partial\Phi(x)$ are nonsingular if F is a uniform P-function. Since it is not difficult to see that the Jacobian matrices of a uniform P-function are P-matrices, we reobtain their result as a direct consequence of Corollary 23.

6. Convergence of the algorithm

The main aim of this section is to prove and discuss Theorem 11.

Proof of point (a) of Theorem 11. Suppose, renumbering if necessary, that $\{x^k\} \to x^*$. We first note that if, for an infinite set of indices K, we have $d^k = -\nabla\Psi(x^k)$ for all $k \in K$, then x^* is a stationary point of Ψ by well known results (see, e.g., Proposition 1.16 in [1]). Hence, without loss of generality, to prove the theorem we only need to consider the case in which the direction is always given by (4). Suppose, by contradiction, that $\nabla\Psi(x^*) \neq 0$. Since the direction d^k always satisfies (4), we can write

$$\|\Phi(x^k)\| = \|H^k d^k\| \leq \|H^k\| \, \|d^k\|, \tag{33}$$

from which we get

$$\|d^k\| \geq \frac{\|\Phi(x^k)\|}{\|H^k\|} \tag{34}$$

(recall the $\|H^k\|$ cannot be 0, otherwise (33) would imply $\Phi(x^k) = 0$, so that x^k would be a stationary point and the algorithm would have stopped).

We now note that

$$0 < m \leq \|d^k\| \leq M, \tag{35}$$

for some positive m and M. In fact if, for some subsequence K, $\{\|d^k\|\}_K \to 0$ we have from (34), that $\{\|\Phi(x^k)\|\}_K \to 0$ because H^k is bounded on the bounded sequence $\{x^k\}$ by known properties of the generalized Jacobian. But then, by continuity, $\Phi(x^*) = 0$ so that x^* is a solution of the nonlinear complementarity problem, thus contradicting the assumption $\nabla\Psi(x^*) \neq 0$. On the other hand $\|d^k\|$ cannot be unbounded because, taking into account that $\nabla\Psi(x^k)$ is bounded and $p > 2$, this would contradict (5).

Then, since (6) holds at each iteration and Ψ is bounded from below on the bounded sequence $\{x^k\}$ we have that $\{\Psi(x^{k+1}) - \Psi(x^k)\} \to 0$ which implies, by the linesearch test,

$$\{2^{-i^k}\nabla\Psi(x^k)^{\mathrm{T}}d^k\} \to 0. \tag{36}$$

We want to show that 2^{-i^k} is bounded away from 0. Suppose the contrary. Then, subsequencing if necessary, we have that $\{2^{-i^k}\} \to 0$ so that at each iteration the stepsize is reduced at least once and (6) gives

$$\frac{\Psi(x^k + 2^{-(i^k-1)}d^k) - \Psi(x^k)}{2^{-(i^k-1)}} > \beta\nabla\Psi(x^k)^{\mathrm{T}}d^k. \tag{37}$$

By (35) we can assume, subsequencing if necessary, that $\{d^k\} \to \bar{d} \neq 0$, so that, passing to the limit in (37), we get

$$\nabla\Psi(x^*)^{\mathrm{T}}\bar{d} \geq \beta\nabla\Psi(x^*)^{\mathrm{T}}\bar{d}. \tag{38}$$

On the other hand we also have, by (5), that $\nabla\Psi(x^*)^{\mathrm{T}}\bar{d} \leq -\rho\|\bar{d}\|^p < 0$, which contradicts (38); hence 2^{-i^k} is bounded away from 0. But then (36) and (5) imply that $\{d^k\} \to 0$, thus contradicting (35), so that $\nabla\Psi(x^*) = 0$. □

Proof of point (b) of Theorem 11. Let $\{x^k\}$ be the sequence of points generated by the algorithm and let be x^* the locally unique limit point in the statement of the theorem; then x^* is an isolated global minimum point of Ψ. Denote by Ω the set of limit points of the sequence $\{x^k\}$; we have that x^* belongs to Ω which is therefore a nonempty set. Let δ be the distance of x^* to $\Omega \setminus x^*$, if x^* is not the only limit point of $\{x^k\}$, 1 otherwise, i.e.

$$\delta = \begin{cases} \inf\limits_{y \in \Omega \setminus x^*} \{\|y - x^*\|\} & \text{if } \Omega \setminus x^* \neq \emptyset, \\ 1 & \text{otherwise;} \end{cases}$$

since x^* is an isolated solution, we have $\delta > 0$. Let us now indicate by Ω_1 and Ω_2 the following sets,

$$\Omega_1 = \{y \in \mathbb{R}^n : dist\{y \mid \Omega\} \leq \delta/4\}, \qquad \Omega_2 = \{y \in \mathbb{R}^n : \|y\| \geq \|x^*\| + \delta\}.$$

We have that for k sufficiently large, let us say for $k \geq \bar{k}$, x^k belongs at least to one of the two sets Ω_1, Ω_2. Let now K be the subsequence of all k for which $\|x^k - x^*\| \leq \delta/4$ (this set is obviously nonempty because x^* is a limit point of the sequence). Since all points of the subsequence $\{x^k\}_K$ are contained in the compact set $S(x^*, \delta/4)$ and every limit point of this sequence is also a limit point of $\{x^k\}$, we have that all the subsequence $\{x^k\}_K$ converges to x^*, the unique limit point of $\{x^k\}$ in $S(x^*, \delta/4)$. Furthermore, since x^* is a solution of NCP(F) we have that $\{\|\nabla\Psi(x^k)\|\}_K \to 0$ which in turn, by (5), implies that $\{d^k\}$ tends to 0. So we can find $\tilde{k} \geq \bar{k}$ such that $\|d^k\| \leq \delta/4$ if $k \in K$ and $k \geq \tilde{k}$. Let now $\hat{k}$ be any fixed $k \geq \tilde{k}$ belonging to K; we can write:

$$dist\{x^{\hat{k}+1} \mid \Omega \setminus x^*\} \geq \inf_{y \in \Omega \setminus x^*} \{\|y - x^*\|\} - (\|x^* - x^{\hat{k}}\| + \|x^{\hat{k}} - x^{\hat{k}+1}\|)$$

$$\geq \delta - \delta/4 - \delta/4 = \delta/2. \tag{39}$$

This implies that $x^{\hat{k}+1}$ cannot belong to $\Omega_1 \setminus S(x^*; \delta/4)$; on the other hand, since $x^{\hat{k}+1} = x^{\hat{k}} + \alpha^{\hat{k}} d^{\hat{k}}$ for some $\alpha^{\hat{k}} \in (0, 1]$, we have

$$\|x^{\hat{k}+1}\| \leq \|x^{\hat{k}}\| + \|\alpha^{\hat{k}} d^{\hat{k}}\| \leq \|x^* + (x^{\hat{k}} - x^*)\| + \|d^{\hat{k}}\|$$

$$\leq \|x^*\| + \|x^{\hat{k}} - x^*\| + \|d^{\hat{k}}\| \leq \|x^*\| + \delta/4 + \delta/4,$$

so that $x^{\hat{k}+1}$ does not belong to Ω_2. Hence we get that $x^{\hat{k}+1}$ belongs to $S(x^*; \delta/4)$. But then, by definition, we have that $\hat{k} + 1 \in K$, so by induction (recall that $\hat{k} + 1 > \tilde{k}$ also, so that $\|d^{\hat{k}+1}\| \leq \delta/4$) we have that every $k > \tilde{k}$ belongs to K and the whole sequence converges to x^*. □

Remark 24. We explicitly point out that in the proof of point (b) we have shown that if the sequence of points generated by the algorithm is converging to a locally unique solution of the nonlinear complementarity problem, then $\{\|d^k\|\} \to 0$. This fact will be used also in the proof of point (c).

Proof of point (c) of Theorem 11. The fact that $\{x^k\} \to x^*$ follows by part (b) noting that the BD-regularity assumption implies, by [37, Proposition 3], that x^* is an isolated solution of the system $\Phi(x) = 0$ and hence also of NCP(F). Then, we first prove that locally the direction is always the solution of system (4) and then that eventually the stepsize of one satisfies the linesearch test (6), so that the algorithm eventually reduces to the local algorithm (2) and the assertions on the convergence rate readily follow from Theorems 8 and 10.

Since $\{x^k\}$ converges to a BD-regular solution of the system $\Phi(x) = 0$, we have, by Lemma 2.6 [38], that the determinant of H^k is bounded away from 0 for k sufficiently

large. Hence system (4) always admits a solution for k sufficiently large. We want to show that this solution satisfies, for some positive ρ_1, the condition

$$\nabla\Psi(x^k)^{\mathrm{T}}d^k \leq -\rho_1\|d^k\|^2. \tag{40}$$

We first note that, by (4),

$$\|d^k\| \leq \|(H^k)^{-1}\|\,\|\Phi(x^k)\|,$$

so that, recalling that $\nabla\Psi(x^k)$ can be written as $H^k\Phi(x^k)$,

$$\nabla\Psi(x^k)^{\mathrm{T}}d^k = -\|\Phi(x^k)\|^2 \leq -\frac{\|d^k\|^2}{M^2}, \tag{41}$$

where M is an upper bound on $\|(H^k)^{-1}\|$ (note that this upper bound exists because the determinant of H^k is bounded away from 0). Then (40) follows by (41) by taking $\rho_1 \leq 1/M^2$. But now it is easy to see, noting that $\{\|d^k\|\} \to 0$, that (40) eventually implies (5) for any $p > 2$ and any positive ρ.

Now to complete the proof of the theorem it only remains to show that eventually the stepsize determined by the Armijo test (6) is 1, that is, that eventually $i^k = 0$. But this immediately follows by Theorem 3.2 in [6], taking into account (40) and the fact that Ψ is SC^1 by Theorem 10. □

We can now use Theorem 11 and the results of Sections 4 and 5 to easily obtain the following two theorems.

Theorem 25. *Let x^* be a limit point of the sequence generated by the algorithm. Then x^* is a solution of NCP(F) iff it is a regular point according to Definition 13. In particular x^* is a solution of NCP(F) if it satisfies any of the conditions of Theorem 15, Corollary 16, Theorem 17, Theorem 21, Corollary 22, or Corollary 23.*

Proof. The proof is obvious except for the fact that the conditions of Theorem 21, Corollary 22, or Corollary 23 are sufficient to guarantee that x^* is a solution of NCP(F). But to prove this it suffices to note that, by Theorem 11 (a), $\nabla\Psi(x^*) = 0$. But employing elementary rules on the calculation of subgradients, it can be shown that $\nabla\Psi(x^*) = \partial\Phi(x^*)^{\mathrm{T}}\Phi(x^*)$ (see discussion in Section 3). If x^* is not a solution of NCP(F), then $\Phi(x^*) \neq 0$, so that, since Theorem 21, Corollary 22, and Corollary 23 guarantee the nonsingularity of all the elements of $\partial\Phi(x^*)$, it cannot be $\nabla\Psi(x^*) = 0$ (note that to prove the result it would be sufficient to show that there exists at least one nonsingular element in $\partial\Phi(x^*)$). □

The next result is an immediate consequence of the analysis carried out in Section 5.

Theorem 26. *Let x^* be a limit point of the sequence generated by the algorithm, and suppose that x^* satisfies any of the conditions of Theorem 21, Corollary 22, or Corollary 23. Then x^* is a BD-regular solution of system $\Phi(x) = 0$ so that all the assertions of point (c) of Theorem 11 hold true.*

It is interesting to note that Theorem 21, Corollary 22, or Corollary 23 guarantee the nonsingularity of *all* the elements of $\partial\Phi(x^*)$ while, according to Theorem 10, BD-regularity, i.e. nonsingularity of all the elements in $\partial_B\Phi(x^*)$, is sufficient to have superlinear/quadratic convergence. Hence a fast convergence rate could take place also when the conditions of Theorem 21, Corollary 22, or Corollary 23 are not met. We illustrate this with a simple example. Consider the unidimensional complementarity problem with $F(x) = -x$; obviously the unique solution of this problem is $x^* = 0$. For this problem we have $\Phi(x) = \sqrt{x^2 + x^2} - x + x = \sqrt{2}|x|$. Hence $\partial_B\Phi(0) = \{-\sqrt{2}, \sqrt{2}\}$, so that $x^* = 0$ is a BD-regular solution of $\Phi(x) = 0$. On the other hand none of the conditions of Theorem 21, Corollary 22, or Corollary 23 can be satisfied since $\partial\Phi(0) = [-\sqrt{2}, \sqrt{2}]$ contains the singular element 0.

To complete the discussion of the properties of the algorithm we note that if the function F is a uniform P-function then the level sets of Ψ are bounded by Theorem 10 (c) so that at least one limit point of the sequence produced by the algorithm exists. By Theorem 11 (a), each limit point is a stationary point of Ψ, and by Corollary 16 (b) every stationary point is a solution of NCP(F). Since a complementarity problem with a uniform P-function has a unique solution, we obtain the result that any sequence $\{x^k\}$ generated by our algorithm converges to this unique solution. We note that the case of a uniform P-function is, basically, the case considered in [23].

7. Numerical results

In this section we perform some numerical tests in order to show the viability of the approach proposed.

In Section 3 we left unanswered the problem of calculating an element of $\partial_B\Phi(x)$, which is needed in Step 2 of the algorithm. Our first task is therefore to show how this can be accomplished.

Procedure to evaluate an element H belonging to $\partial_B\Phi(x)$.

Step 1: Set $\beta = \{i : x_i = 0 = F_i(x)\}$.

Step 2: Choose $z \in \mathbb{R}^n$ such that $z_i \neq 0$ for all i belonging to β.

Step 3: For each $i \notin \beta$ set the ith row of H equal to

$$\left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2}} - 1\right) e_i^{\mathrm{T}} + \left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2}} - 1\right) \nabla F_i(x)^{\mathrm{T}}. \tag{42}$$

Step 4: For each $i \in \beta$ set the ith row of H equal to

$$\left(\frac{z_i}{\sqrt{z_i^2 + (\nabla F_i(x)^{\mathrm{T}} z)^2}} - 1\right) e_i^{\mathrm{T}} + \left(\frac{\nabla F_i(x)^{\mathrm{T}} z}{\sqrt{z_i^2 + (\nabla F_i(x)^{\mathrm{T}} z)^2}} - 1\right) \nabla F_i(x)^{\mathrm{T}}. \tag{43}$$

Theorem 27. *The element H calculated by the above procedure is an element of $\partial_B \Phi(x)$.*

Proof. We shall build a sequence of points $\{y^k\}$ where $\Phi(x)$ is differentiable and such that $\nabla\Phi(y^k)^T$ tends to H; the theorem will then follow by the very definition of B-subdifferential.

Let $y^k = x + \varepsilon^k z$, where z is the vector of Step 2 and $\{\varepsilon^k\}$ is a sequence of positive numbers converging to 0. Since, if $i \notin \beta$, either $x_i \neq 0$ or $F_i(x) \neq 0$, and $z_i \neq 0$ for all $i \in \beta$, we can assume, by continuity, that ε^k is small enough so that, for each i, either $y_i^k \neq 0$ or $F_i(y^k) \neq 0$, and Φ is therefore differentiable at y^k.

Now, if i does not belong to β, it is obvious, by continuity, that the ith row of $\nabla\Phi(y^k)^T$ tends to the ith row of H; so the only case of concern is when i belongs to β. We recall that, according to Lemma 12, the ith row of $\nabla\Phi(y^k)^T$ is given by

$$(a_i(y^k) - 1)e_i^T + (b_i(y^k) - 1)\nabla F_i(y^k)^T, \tag{44}$$

where

$$a_i(y^k) = \frac{\varepsilon^k z_i}{\sqrt{(\varepsilon^k)^2 z_i^2 + F_i(y^k)^2}}, \qquad b_i(y^k) = \frac{F_i(y^k)}{\sqrt{(\varepsilon^k)^2 z_i^2 + F_i(y^k)^2}}.$$

We note that by the Taylor-expansion we can write, for each $i \in \beta$,

$$F_i(y^k) = F_i(x) + \varepsilon^k \nabla F(\zeta^k)^T z = \varepsilon^k \nabla F(\zeta^k)^T z, \quad \text{with } \zeta^k \to x. \tag{45}$$

Substituting (45) in (44) and passing to the limit, we obtain, taking into account the continuity of ∇F that also the rows of $\nabla\Phi(y^k)^T$ relative to indices in β tend to the corresponding rows of H defined in Step 4. □

Changing z we will obtain a different element of $\partial_B \Phi(x)$. In our code we chose to set $z_i = 0$ if $i \notin \beta$ and $z_i = 1$ if $i \in \beta$. We note that the overhead to evaluate an element of $\partial_B \Phi(x)$ when Φ is not differentiable is negligible with respect to the case in which Φ is differentiable. This, again, is a favourable characteristic of our reformulation, which is not true, in general, for semismooth systems of equations.

The implemented version of the algorithm differs from the one described in Section 3 in the use of a *nonmonotone* linesearch, which can be viewed as an extension of (6). To motivate this variant we first recall that it has been often observed in the field of nonlinear complementarity algorithms that the linesearch test used to enforce global convergence can lead to very small stepsizes; in turn this can bring to very slow convergence and even to a numerical failure of the algorithm. To circumvent this problem many *heuristics* have been used, see, e.g., [20,25,36]. Here, instead, we propose to substitute the linesearch test (6) by the following *nonmonotone* linesearch:

$$\Psi(x^k + 2^{-i}d^k) \leq \mathcal{W} + \beta 2^{-i}\nabla\Psi(x^k)^T d^k, \tag{46}$$

where $\mathcal{W}$ is any value satisfying

$$\Psi(x^k) \leq \mathcal{W} \leq \max_{j=0,1,\ldots,M^k} \Psi(x^{k-j}) \tag{47}$$

and M^k are nonnegative integers bounded above for any k (we suppose for simplicity that k is large enough so that no negative superscripts can occur in (47)). This kind of linesearch has been first proposed in [17] and since then it has proved very useful in the unconstrained minimization of smooth functions. Adopting the same proof techniques used in [17] it is easy to see that all the results described in the previous section can be extended if we substitute the linesearch (46) in place of (6) in the algorithm of Section 3. If at each iteration we take $\mathcal{W} = \Psi(x^k)$ we obtain exactly the algorithm of Section 3, however much more freedom is allowed by (46). In particular, according to the most recent experiences in nonmonotone minimization algorithms [51], we chose to keep $\mathcal{W}$ fixed as long as the algorithm seems to make progress and to vary it only if for a certain prefixed number of consecutive steps the function values $\Psi(x^k)$ increase; the precise scheme is the following, where we have indicated by $\Psi^k_{\min}$ the smallest value among $\Psi(x^0), \Psi(x^1), \ldots, \Psi(x^k)$.

- Set $\mathcal{W} = \Psi(x^0)$ at the beginning of the algorithm.
- Keep the value of $\mathcal{W}$ fixed as long as

$$\min_{j=0,1,\ldots,5} \Psi(x^{k-j}) = \Psi^k_{\min}. \tag{48}$$

- If (48) is not satisfied at iteration k, set $\mathcal{W} = \Psi(x^k)$.
- In any case, at most every 30 iterations reset $\mathcal{W}$ to $\max_{j=0,1,\ldots,5} \Psi(x^{k-j})$.

We implemented this version of the algorithm in Fortran and ran it on a RISC/6000 workstation. The main stopping criterion was

$$\| \min\{x^k, F(x^k)\} \| \leq 10^{-6}.$$

To take into account the possibility of convergence to a stationary point of Ψ which is not a solution of the complementarity problem, we also considered the stopping criterion $\|\nabla\Psi(x^k)\| \leq 10^{-9}$. However, in our tests the algorithm never stopped because of this latter criterion. We also set $p = 2.1$, $\rho = 10^{-8}$ and $\beta = 10^{-4}$. The results obtained on a set of test problems are reported in Table 1, where we give for each problem the dimension (Dimen.), the number of iterations (it.) and the number of evaluations of F needed (F); in some cases we used more than one starting point, this is indicated in the column Start. point. We recall that the number of iterations is also equal to the number of Jacobian evaluations and linear systems solved. Below we report, for each problem, some relevant data, the starting points and some comments. The source reported for the problem is not necessarily the original one.

Kojima-Shindo problem. See [36]. $F(x)$ is not a P_0-function. This problem has two solutions: $x^1 = (0.5\sqrt{6}, 0, 0, 0.5)$ and $x^2 = (1, 0, 3, 0)$; x^1 fails to be R-regular and is degenerate. We report the point to which convergence occured in parenthesis, after the number of iterations. The linearized complementarity problem at 0 has no solution, so

Table 1
Results for the nonmonotone Newton method

Problem	Dimen.	Start. point	it.	F
Kojima-Shindo	4	a	13 (x^1)	14
	4	b	7 (x^2)	12
Spatial Eq.	42	a	20	23
	42	b	22	23
Traffic Eq.	50	a	13	14
Nash-Cournot	10	a	7	8
	10	b	10	11
Mod. Mathiesen	4	a	4	5
	4	b	5	6
Walrasian	14	a	25	51
	14	b	24	49
HS34	8	a	14	36
HS35	4	a	5	6
HS66	4	a	6	8
HS76	8	a	6	7
Watson 3	10	a	7	12
Watson 4	5	a	21	22
Kojima-Josephy	4	a	20	26
	4	b	7	11

that the classical Robinson-Josephy Newton-type scheme fails when started at this point.
Starting points: (a) $(0,0,0,0)$, (b) $(1,1,1,1)$.

Spatial price equilibrium problem. See [36,50]. This is a problem arising from a spatial equilibrium model. F is a P-function and the unique solution is therefore R-regular.
Starting points: (a) $(0,0,\ldots,0)$, (b) $(1,1,\ldots,1)$.

Traffic equilibrium problem. See [36]. This is a traffic equilibrium problem with elastic demand.
Starting point: (a) All the components are 0 except $x_1, x_2, x_3, x_{10}, x_{11}, x_{20}, x_{21}, x_{22}$, $x_{29}, x_{30}, x_{40}, x_{45}$ which are 1, $x_{39}, x_{42}, x_{43}, x_{46}$ which are 7, $x_{41}, x_{47}, x_{48}, x_{50}$ which are 6, and x_{44} and x_{49} which are 10.

Nash-Cournot production problem. See [19,36]. F is not twice continuously differentiable. F is a P-function on the strictly positive orthant, and since the solution obtained has all the components strictly positive, it is R-regular.
Starting points: (a) $(10,10,\ldots,10)$, (b) $(1,1,\ldots,1)$.

Modified Mathiesen equilibrium problem. This is a slight modification of the Mathiesen example of a Walrasian equilibrium model as suggested in [25]. F is not defined everywhere and does not belong to any known class of functions. There are infinitely many solutions: $(\lambda,0,0,0)$, $\lambda \in [0,3]$. For $\lambda = 0,3$ the solutions are degenerate, for

$\lambda \in (0,3)$ nondegenerate; in any case being nonisolated, all the solutions fail to be R-regular. When we started from starting point (a) convergence occurred to $(0,0,0,0)$, when the starting point was (b), instead, convergence occurred to $(3,0,0,0)$.
Starting points: (a) $(1,1,1,1)$, (b) $(10,10,10,10)$.

Walrasian equilibrium problem. See [36,48]. This is a Walrasian equilibrium model, which is in general difficult for Newton-type methods since most of the standard regularity conditions are not generally satisfied at the solution.
Starting points: (a) $(1,1,1,1,1,1,0.4,1,0,4,0,0,1,0)$, (b) $(0.5,0.5,0.5,0.5,0.5,0.5,0.4,1,0,4,0,0,1,0)$.

HS34 problem. This problem represents the KKT conditions for the 34th problem in [21]. The resulting F is monotone on the positive orthant but not even P_0 on $\mathbb{R}^n$.
Starting point: (a) $(0,1.05,2.9,0,0,0,0,0)$.

HS35 problem. This problem represents the KKT conditions for the 35th problem in [21]. The resulting F is linear and monotone but not strictly monotone.
Starting point: (a) $(0.5,0.5,0.5,0)$.

HS66 problem. This problem represents the KKT conditions for the 66th problem in [21]. The resulting F is monotone on the positive orthant but not even P_0 on $\mathbb{R}^n$.
Starting point: (a) $(0,1.05,2.9,0,0,0,0,0)$.

HS76 problem. This problem represents the KKT conditions for the 76th problem in [21]. The resulting F is linear and monotone but not strictly monotone.
Starting point: (a) $(0.5,0.5,0.5,0.5,0,0,0)$.

Watson third problem. See [53]. This is a linear complementarity problem with $F(x) = Mx + q$. M is not even semimonotone and none of the standard algebraic techniques can solve it. With the choice $q = (-1,0,\ldots,)$, which we have adopted, also the continuation method of [53] fails on this problem.
Starting point: (a) $(0,0,\ldots,0)$.

Watson fourth problem. See [53]. This problem represents the KKT conditions for a convex programming problem involving exponentials. The resulting F is monotone on the positive orthant but not even P_0 on $\mathbb{R}^n$.
Starting point: (a) $(0,0,\ldots,0)$.

Kojima-Josephy problem. See [28,24]. $F(x)$ is not a P_0-function. The problem has a unique solution which is not R-regular.
Starting points: (a) $(0,0,0,0)$, (b) $(1,1,1,1)$.

We see that the algorithm was capable of solving all the test problems, some of which are known to be quite troublesome, with a fairly low number of iterations and

Table 2
Results for pure Newton and monotone versions of the algorithm

Problem	Dimen.	Start. point	Pure Newton		Monotone	
			it.	F	it.	F
Kojima-Shindo	4	a	13 (x^1)	14	10 (x^2)	22
	4	b	10 (x^1)	11	7 (x^2)	15
Spatial Eq.	42	a	21	22	20	23
	42	b	22	23	21	23
Traffic Eq.	50	a	13	14	19	97
Nash-Cournot	10	a	7	8	7	8
	10	b	10	11	10	11
Mod. Mathiesen	4	a	4	5	4	5
	4	b	5	6	5	6
Walrasian	14	a	F(s)	-	25	51
	14	b	F(s)	-	24	49
HS34	8	a	F(s)	-	26	113
HS35	4	a	5	6	5	6
HS66	4	a	F(s)	-	6	10
HS76	8	a	6	7	6	7
Watson 3	10	a	7	8	7	13
Watson 4	5	b	21	22	21	22
Kojima-Josephy	4	a	F(m)	-	7	10
	4	b	F(m)	-	7	12

function evaluations. The algorithm uses the gradient as a search direction extremely seldom, and in all cases a fast convergence rate was observed in the last iterations. To better understand the characteristics of the algorithm we report in Table 2 the results obtained for two other "versions" of the algorithm. Under the heading "Pure Newton" we reported the results for the pure Newton, local algorithm (2). In other words at each iteration we calculate d^k by (4) and set $x^{k+1} = x^k + d^k$. This algorithm can fail for two reasons: either because the system (4) is not solvable or simply because the iterates do not converge (we set a limit of 500 iterations). The first occurrence is indicated by F(s) in Table 2, while the latter is indicated by F(m). In Table 2 we also reported the results for the algorithm described in Section 3, without the modified nonmonotone linesearch technique describe above. This algorithm is called Monotone because it corresponds to the standard Armijo linesearch, where the function values are forced to decrease monotonically.

The comparison of the results of Table 1 and Table 2 is instructive. It is well known that, generally, when Newton's method "works", it works quite well, but it can fail if the starting point is not sufficiently close to the solution. The aim of any stabilization technique is to force global convergence from far away starting points. However, ideally, we would like that the stabilization technique does not perturb the pure Newton method when it "works", and, on the set of test problems used here, this is exactly what our (nonmonotone) algorithm does. We see, comparing Tables 1 and 2, that whenever the pure Newton method works, our nonmonotone strategy does not perturb it, while, at the same time, it is able to force convergence also in the case in which the pure Newton

method fails. On the other hand we see that the more standard monotone stabilization technique can alter (and make worst) the pure Newton method also when this would not be necessary; more in general the monotone version tends to need more iterations than its nonmonotone counterpart (see, for example, the traffic equilibrium problem and the HS34 problem).

To conclude it may be interesting to compare some features of our algorithm with those of the global algorithm proposed by Jiang and Qi in [23]. They try to solve the generalized Newton equation (4) in order to get a search direction d^k; if

$$\Psi(x^k + d^k) \leq \sigma\Psi(x^k), \tag{49}$$

for some fixed constant $\sigma \in (0, 1)$, they take $x^{k+1} := x^k + d^k$ as the next iterate, otherwise they set $d^k := -\nabla\Psi(x^k)$ and perform a monotone Armijo linesearch. The test (49) is motivated by the fact that $\Psi(x) = 0$ at a solution $x = x^*$ of NCP(F), and it is not difficult to see that all our convergence results remain true if this acceptability test is added to our global algorithm. It is our feeling, however, that condition (49) will not be satisfied very often as long as we are far away from a solution of NCP(F), so that the steepest descent direction has probably to be taken in many iterations and this seems unfavourable from a numerical point of view. In our approach, instead, as we already remarked, the steepest descent direction is used only as a safeguard and most of the times the Newton direction is employed. To this end the observation that d^k, given by (4), is always a descent direction for the merit function Ψ and locally a "good" descent direction (see Sections 3 and 6) seems a crucial point which is missing in [23].

8. Concluding remarks

In this paper we presented a new method for the solution of nonlinear complementarity problems NCP(F). The central idea was to reformulate NCP(F) as a semismooth system of equations $\Phi(x) = 0$, to which a generalized Newton method is applied. In contrast to other equation-based methods (see, e.g., [20,34,35]), the norm function $\Psi(x) = \|\Phi(x)\|^2$, which is the natural merit function for systems of equations, is differentiable, and the search direction d^k from the generalized Newton-equation $H^k d = -\Phi(x^k)$, $H^k \in \partial_B\Phi(x^k)$, is a descent direction for this merit function under some very mild assumptions. From the computational point of view the most attractive feature of the new algorithm is that the solution of a linear system of equations at each iteration is sufficient to guarantee global and quadratic convergence, even on degenerate problems. This contrasts sharply with recent algorithms proposed for the solution of NCP(F), where the solution of a (mixed) linear complementarity problem or of a quadratic program at each iteration is required to achieve the same results.

As far as we are aware of, the conditions guaranteeing that a stationary point of the merit function Ψ is a solution of NCP(F) are currently the weakest ones among those proved for methods of the kind considered in this paper. In fact, it is our feeling that these conditions are close to the boundary of what is possible to show. Also the assumptions

guaranteeing the solvability of the subproblems $H^k d = -\Phi(x^k)$, $H^k \in \partial_B \Phi(x^k)$, appear to be very weak; in particular, when far away from a solution, the index sets α and β as introduced in Section 5 are usually empty, and Lemma 20 (or Theorem 21) therefore states that each element $H^k \in \partial\Phi(x^k)$ is nonsingular if F is a P_0-function.

References

[1] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods* (Academic Press, New York, 1982).

[2] B. Chen and P.T. Harker, A noninterior continuation method for quadratic and linear programming, *SIAM J. on Optimization* 3 (1993) 503–515.

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

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

[5] S.P. Dirkse and M.C. Ferris, The PATH solver: A Non-monotone Stabilization Scheme for Mixed Complementarity Problems, *Optimization Methods and Software* 5 (1995) 123–156.

[6] F. Facchinei, Minimization of SC^1 functions and the Maratos effect, *Operations Research Letters* 17 (1995), 131–137.

[7] F. Facchinei and J. Soares, A new merit function for nonlinear complementarity problems and a related algorithm, Technical Report 15.94, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", Rome, Italy (1994) (to appear in: *SIAM J. on Optimization*).

[8] F. Facchinei and J. Soares, Testing a new class of algorithms for nonlinear complementarity problems, in: F. Giannessi and A. Maugeri, eds., *Variational Inequalities and Network Equilibrium Problems* (Plenum Press, New York, 1995) 69–83.

[9] M.C. Ferris and S. Lucidi, Nonmonotone stabilization methods for nonlinear equations, *J. of Optimization Theory and Applications* 81 (1994) 53–71.

[10] M.C. Ferris and D. Ralph, Projected gradient methods for nonlinear complementarity problems via normal maps, in: D.Z. Du, L. Qi and R.S. Womersley, eds., *Recent Advances in Nonsmooth Optimization* (World Scientific Publishers, Singapore) 57–87.

[11] A. Fischer, A special Newton-type optimization method, *Optimization* 24 (1992) 269–284.

[12] A. Fischer, A Newton-type method for positive semidefinite linear complementarity problems, *J. of Optimization Theory and Applications* 86 (1995) 585–608.

[13] A. Fischer, An NCP-function and its use for the solution of complementarity problems, in: D.Z. Du, L. Qi and R.S. Womersley, eds., *Recent Advances in Nonsmooth Optimization* (World Scientific Publishers, Singapore) 88–105.

[14] A. Fischer and C. Kanzow, On finite termination of an iterative method for linear complementarity problems, Preprint MATH-NM-10-1994, Institute for Numerical Mathematics, Technical University of Dresden, Dresden, Germany (1994) (to appear in: *Mathematical Programming*).

[15] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Mathematical Programming* (Series A) 53 (1992) 99–110.

[16] C. Geiger and C. Kanzow, On the resolution of monotone complementarity problems, *Computational Optimization and Applications* 5 (1996) 155–173.

[17] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone linesearch technique for Newton's method, *SIAM J. on Numerical Analysis* 23 (1986) 707–716.

[18] S.P. Han, J.-S. Pang and N. Rangaraj, Globally convergent Newton methods for nonsmooth equations, *Mathematics of Operations Research* 17 (1992) 586–607.

[19] P.T. Harker, Accelerating the convergence of the diagonalization and projection algorithm for finite-dimensional variational inequalities, *Mathematical Programming* (Series A) 41 (1988) 25–59.

[20] P.T. Harker and B. Xiao, Newton's method for the nonlinear complementarity problem: A B-differentiable equation approach, *Mathematical Programming* (Series A) 48 (1990) 339–357.

[21] W. Hock and K. Schittkowski, Test examples for nonlinear programming codes, Lectures Notes in Economics and Mathematical Systems 187 (Springer-Verlag, Berlin, 1981).

[22] H. Jiang, Unconstrained minimization approaches to nonlinear complementarities, Applied Mathematics Report AMR 94/33, School of Mathematics, University of New South Wales, Sydney, Australia (1994).
[23] H. Jiang and L. Qi, A nonsmooth equations approach to nonlinear complementarities, Applied Mathematics Report AMR 94/31, School of Mathematics, University of New South Wales, Sydney, Australia (1994) (to appear in: *SIAM J. on Control and Optimization*).
[24] N.H. Josephy, Newton's method for generalized equations, MRC Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin-Madison, WI (1979).
[25] C. Kanzow, Some equation-based methods for the nonlinear complementarity problem, *Optimization Methods and Software* 3 (1994) 327–340.
[26] C. Kanzow, Nonlinear complementarity as unconstrained optimization, *J. of Optimization Theory and Applications* 88 (1996) 139–155.
[27] C. Kanzow, Global convergence properties of some iterative methods for linear complementarity problems, Preprint 72, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany (1993) revised (1994) (to appear in: *SIAM J. on Optimization*).
[28] M. Kojima, Computational methods for solving the nonlinear complementarity problem, Keio Engineering Reports 27 (1974) 1–41.
[29] O.L. Mangasarian, Equivalence of the complementarity problem to a system of nonlinear equations, *SIAM J. on Applied Mathematics* 31 (1976) 89–92.
[30] O.L. Mangasarian and M.V. Solodov, Nonlinear complementarity as unconstrained and constrained minimization, *Mathematical Programming* (Series B) 62 (1993) 277–297.
[31] R. Mifflin, Semismooth and semiconvex functions in constrained optimization, *SIAM J. on Control and Optimization* 15 (1977) 957–972.
[32] J.J. Moré, Global methods for nonlinear complementarity problems, Preprint MCS-P429-0494, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL (1994).
[33] J.J. Moré and W.C. Rheinboldt, On P- and S-functions and related classes of n-dimensional nonlinear mappings, *Linear Algebra and its Applications* 6 (1973) 45–68.
[34] J.-S. Pang, Newton's method for B-differentiable equations, *Mathematics of Operations Research* 15 (1990) 311–341.
[35] J.-S. Pang, A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems, *Mathematical Programming* (Series A) 51 (1991) 101–132.
[36] J.-S. Pang and S.A. Gabriel, NE/SQP: A robust algorithm for the nonlinear complementarity problem, *Mathematical Programming* (Series A) 60 (1993) 295–337.
[37] J.-S. Pang and L. Qi, Nonsmooth equations: motivation and algorithms, *SIAM J. on Optimization* 3 (1993) 443–465.
[38] L. Qi, A convergence analysis of some algorithms for solving nonsmooth equations, *Mathematics of Operations Research* 18 (1993) 227–244.
[39] L. Qi and H. Jiang, Karush-Kuhn-Tucker equations and convergence analysis of Newton methods and Quasi-Newton methods for solving these equations, Technical Report AMR 94/5, School of Mathematics, University of New South Wales, Australia (1994) (to appear in: *SIAM J. on Control and Optimization*).
[40] L. Qi and J. Sun, A nonsmooth version of Newton's method, *Mathematical Programming* (Series A) 58 (1993) 353–368.
[41] D. Ralph, Global convergence of damped Newton's method for nonsmooth equations via the path search, *Mathematics of Operations Research* 19 (1994) 352–389.
[42] S.M. Robinson, Strongly regular generalized equations, *Mathematics of Operations Research* 5 (1980) 43–62.
[43] S.M. Robinson, Implicit B-differentiability in generalized equations, Technical Summary Report 2854, Mathematics Research Center, University of Wisconsin-Madison, WI (1985).
[44] S.M. Robinson, An implicit function theorem for B-differentiable functions, IIASA Working Paper 88-67, Laxemburg, Austria (1988).
[45] S.M. Robinson, Newton's method for a class of nonsmooth functions, Manuscript, Department of Industrial Engineering, University of Wisconsin-Madison, WI (1988).
[46] S.M. Robinson, Normal maps induced by linear transformations, *Mathematics of Operations Research* 17 (1992) 691–714.
[47] S.M. Robinson, Generalized equations, in: A. Bachem, M. Groetschel and B. Korte, eds., *Mathematical programming: the state of the art* (Springer-Verlag, Berlin, 1983) 346–367.

[48] H. Scarf, *The Computation of Economic Equilibria* (Yale University Press, New Haven, CT, 1973).
[49] P.K. Subramanian, Gauss-Newton methods for the complementarity problem, *J. of Optimization Theory and Applications* 77 (1993) 467–482.
[50] R.L. Tobin, Variable dimension spatial price equilibrium algorithm, *Mathematical Programming* (Series A) 40 (1988) 33–51.
[51] P.L. Toint, A non-monotone trust-region algorithm for nonlinear optimization subject to convex constraints, Report 92/94, Department of Mathematics, FUNDP, Namur, Belgium (1994).
[52] P. Tseng, Growth behavior of a class of merit functions for the nonlinear complementarity problem, Technical Report, Department of Mathematics, University of Washington, Seattle, WA (1994).
[53] L.T. Watson, Solving the nonlinear complementarity problem by a homotopy method, *SIAM J. on Control and Optimization* 17 (1979) 36–46.
[54] B. Xiao and P.T. Harker, A nonsmooth Newton method for variational inequalities, I: Theory, *Mathematical Programming* (Series A) 65 (1994) 151–194.
[55] B. Xiao and P.T. Harker, A nonsmooth Newton method for variational inequalities, II: Numerical results, *Mathematical Programming* (Series A) 65 (1994) 195–216.