A NEW CLASS OF SEMISMOOTH NEWTON-TYPE METHODS FOR NONLINEAR COMPLEMENTARITY PROBLEMS

Christian Kanzow[1] and Helmut Kleinmichel[2]

[1] University of Hamburg
Institute of Applied Mathematics
Bundesstrasse 55
D-20146 Hamburg
Germany
e-mail: kanzow@math.uni-hamburg.de

[2] Technical University of Dresden
Institute of Numerical Mathematics
D-01062 Dresden
Germany
e-mail: kleinmi@math.tu-dresden.de

January 14, 1997 (revised September 16, 1997)

Abstract: We introduce a new, one-parametric class of NCP-functions. This class subsumes the Fischer function and reduces to the minimum function in a limiting case of the parameter. This new class of NCP-functions is used in order to reformulate the nonlinear complementarity problem as a nonsmooth system of equations. We present a detailed investigation of the properties of the equation operator, of the corresponding merit function as well as of a suitable semismooth Newton-type method. Finally, numerical results are presented for this method being applied to a number of test problems.

Key Words: Nonlinear complementarity problems, Newton's method, generalized Jacobians, semismoothness, global convergence, quadratic convergence.

1 Introduction

Consider the *nonlinear complementarity problem*, NCP(F) for short, which is to find a solution of the system

$$x_i \geq 0, F_i(x) \geq 0, x_i F_i(x) = 0 \quad \forall i \in I := \{1, \dots, n\},$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. This problem has a number of important applications in operations research, economic equilibrium problems and in the engineering sciences. We refer the reader to the recent survey paper [11] by Ferris and Pang for a description of many of these applications.

The number of solution methods for the complementarity problem NCP(F) is enormous, see, for example, the paper [10] by Ferris and Kanzow for an overview. Many of these methods are based on a so-called NCP-function: An *NCP-function* is a mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}$ having the property

$$\varphi(a, b) = 0 \Longleftrightarrow a \geq 0, b \geq 0, ab = 0.$$

The two probably most prominent examples of an NCP-function are the minimum function

$$\varphi(a, b) = \min\{a, b\}$$

(see, e.g., Pang [23, 24] and Fischer and Kanzow [15]) as well as the Fischer function

$$\varphi(a, b) = \sqrt{a^2 + b^2} - a - b$$

(see, e.g., Fischer [13], Facchinei and Soares [8] and De Luca, Facchinei and Kanzow [4]).

In this paper, we investigate the properties of the following new class of functions:

$$\varphi_\lambda(a, b) := \sqrt{(a - b)^2 + \lambda ab} - a - b, \tag{1}$$

where λ is a fixed parameter such that $\lambda \in (0, 4)$. It is easy to see that the expression inside the square root in (1) is always nonnegative, i.e.

$$(a - b)^2 + \lambda ab \geq 0 \quad \forall\, a, b \in \mathbb{R}. \tag{2}$$

Hence φ_λ is at least well-defined. Moreover, it is elementary to see that φ_λ is an NCP-function. In the special case $\lambda = 2$, the NCP-function φ_λ obviously reduces to the Fischer function, whereas in the limiting case $\lambda \to 0$, the function φ_λ becomes a multiple of the minimum function. Hence the class (1) covers the currently most important NCP-functions so that a closer look at this new class of NCP-functions seems to be worthwile.

Now, if we define the equation operator $\Phi_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Phi_\lambda(x) := \begin{pmatrix} \varphi_\lambda(x_1, F_1(x)) \\ \vdots \\ \varphi_\lambda(x_n, F_n(x)) \end{pmatrix},$$

then it follows immediately from the definition of an NCP-function that

$$x^* \text{ solves NCP}(F) \iff x^* \text{ solves } \Phi_\lambda(x) = 0.$$

Alternatively, if $\Psi_\lambda : \mathbb{R}^n \to \mathbb{R}$ denotes the corresponding merit function

$$\Psi_\lambda(x) := \frac{1}{2}\Phi_\lambda(x)^T\Phi_\lambda(x) = \frac{1}{2}\|\Phi_\lambda(x)\|^2,$$

then we may rewrite the complementarity problem NCP(F) as the unconstrained minimization problem

$$\min \Psi_\lambda(x), \quad x \in \mathbb{R}^n.$$

In Sections 2 and 3, we will give a detailed discussion of the properties of Φ_λ and Ψ_λ, respectively. These properties will be applied in Section 4 in order to establish global and local fast convergence of a nonsmooth Newton-type method for the solution of the nonlinear complementarity problem NCP(F). Numerical results will be reported in Section 5. We then conclude with some final remarks in Section 6.

Some words about our notation. The n-dimensional Euclidian space is abbreviated by $\mathbb{R}^n$. If $x \in \mathbb{R}^n$ is any given vector, we denote its ith component by x_i, i.e., by using the letter i as a subscript. On the other hand, a superscript k usually denotes the kth iterate of a sequence $\{x^k\} \subseteq \mathbb{R}^n$. A function $G : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a C^1 mapping if it is continuously differentiable, and LC^1 mapping if it is differentiable with a locally Lipschitz continuous derivative. The Jacobian of G at a point $x \in \mathbb{R}^n$ will be denoted by $G'(x)$.

The index set $\{1, \dots, n\}$ will often be abbreviated by the capital letter I. If $M \in \mathbb{R}^{n\times n}, M = (m_{ij})$, is any given matrix and $J, K \subseteq I$, then M_{JK} denotes the submatrix in $\mathbb{R}^{|J|\times|K|}$ with elements $m_{ij}, i \in J, j \in K$. Similarly, if $d \in \mathbb{R}^n$ is any given vector with components $d_i, i \in I$, then d_J is the subvector in $\mathbb{R}^{|J|}$ with elements $d_i, i \in J$.

Finally, we will make use of the Landau symbols $o(\cdot)$ and $O(\cdot)$: If $\{\alpha^k\}$ and $\{\beta^k\}$ are two sequences in $\mathbb{R}$ with $\alpha^k > 0, \beta^k > 0$ for all $k \in \mathbb{N}$ and $\lim_{k\to\infty} \beta^k = 0$, then $\alpha^k = o(\beta^k)$ means that $\lim_{k\to\infty} \alpha^k/\beta^k = 0$, whereas $\alpha^k = O(\beta^k)$ just says that $\limsup_{k\to\infty} \alpha^k/\beta^k < +\infty$, i.e., that there is a constant $c > 0$ with $\alpha^k \le c\beta^k$ for all $k \in \mathbb{N}$.

In the following sections, we will further need some definitions and results from nonsmooth analysis. In order to avoid a complete section on this background material, we will introduce the relevant concepts within the subsequent sections directly before they are used. We basically assume, however, that the reader is more or less familiar with these concepts.

2 Properties of the Equation-Operator Φ_λ

We begin with a simple result for the function φ_λ whose elementary proof is left to the reader.

Lemma 2.1 *φ_λ is locally Lipschitz continuous and directionally differentiable everywhere.*

We next want to prove that φ_λ is actually strongly semismooth. To this end, we first recall from [21, 27, 26, 25] that a locally Lipschitz continuous and directionally differentiable function $G : \mathbb{R}^n \to \mathbb{R}^m$ is called *semismooth* at a point $x \in \mathbb{R}^n$ if

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

holds for every $d \to 0$ and every $H \in \partial G(x+d)$, and *strongly semismooth* at $x \in \mathbb{R}^n$ if

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

for $d \to 0$ and $H \in \partial G(x+d)$; here, $G'(x; d)$ denotes the usual directional derivative of G at x in the direction d, and

$$\partial G(x) = \text{conv}\{H \in \mathbb{R}^{m\times n} | \, \exists \{x^k\} \subseteq D_G : \lim_{k\to\infty} x^k \to x \text{ and } \lim_{k\to\infty} G'(x^k) = H\}$$

is Clarke's [2] generalized Jacobian of G at x, with D_G being the set of all points where G is differentiable. Note that $\partial G(x)$ is known to be a nonempty, convex and compact set.

Moreover, we call G (strongly) semismooth if it is (strongly) semismooth at each point $x \in \mathbb{R}^n$. We also mention that every C^1 function is semismooth and that every LC^1 function is strongly semismooth.

Lemma 2.2 *φ_λ is strongly semismooth.*

Proof. We first note that

$$(a - b)^2 + \lambda ab = 0 \Longleftrightarrow a = b = 0. \tag{3}$$

Hence the mapping φ_λ is an LC^1 mapping at any point $(a, b) \neq (0, 0)$. In particular, φ_λ is strongly semismooth at any point $(a, b) \neq (0, 0)$. So it remains to consider the origin.

To this end, let $d := (d_a, d_b)^T \in \mathbb{R}^2$ be any nonzero direction vector. Then φ_λ is smooth at the point $0 + d = d$, so that $H \in \partial\varphi_\lambda(d)$ is uniquely given by

$$H = \left(\frac{2(d_a - d_b) + \lambda d_b}{2\sqrt{(d_a - d_b)^2 + \lambda d_a d_b}} - 1, \frac{-2(d_a - d_b) + \lambda d_a}{2\sqrt{(d_a - d_b)^2 + \lambda d_a d_b}} - 1 \right).$$

Therefore, we obtain after some algebraic manipulations:

$$\begin{aligned} Hd &= \frac{(d_a - d_b)^2 + \lambda d_a d_b}{\sqrt{(d_a - d_b)^2 + \lambda d_a d_b}} - d_a - d_b \\ &= \sqrt{(d_a - d_b)^2 + \lambda d_a d_b} - d_a - d_b \\ &= \varphi_\lambda(d_a, d_b). \end{aligned}$$

On the other hand, the directional derivative of φ_λ at the origin in the direction d is given by

$$\begin{aligned}\varphi_\lambda'(0;d) &= \lim_{t\to 0+} \frac{\varphi_\lambda(0+td)-\varphi_\lambda(0)}{t} \\ &= \lim_{t\to 0+} \sqrt{(d_a-d_b)^2+\lambda d_a d_b} - d_a - d_b \\ &= \varphi_\lambda(d_a,d_b).\end{aligned}$$

Hence we have

$$Hd - \varphi_\lambda'(0;d) = 0$$

for all nonzero vectors $d \in \mathbb{R}^2$. In particular, it follows that φ_λ is strongly semismooth also in the origin. □

The following result is a simple consequence of Lemma 2.2.

Theorem 2.3 *The following statements hold:*

(a) Φ_λ *is semismooth.*

(b) If F *is an* LC^1 *mapping, then* Φ_λ *is strongly semismooth.*

Proof. We first recall that Φ_λ is (strongly) semismooth if and only if each component function is (strongly) semismooth, see [27]. We further recall that every C^1 mapping is semismooth and every LC^1 mapping is strongly semismooth. Hence, by Lemma 2.2, each component function of Φ_λ is a composition of semismooth functions if F is continuously differentiable and a composition of strongly semismooth functions if F is an LC^1 mapping. However, it is known that the composition of (strongly) semismooth functions is again (strongly) semismooth, see [21, 14]. Hence Φ_λ itself is (strongly) semismooth. □

The following is a very technical result which, however, turns out to be highly important in order to establish Theorem 2.8 below which, in turn, is the basis for a fast local rate of convergence of our algorithm to be described in Section 4.

Lemma 2.4 *Let* $f_\lambda : \mathbb{R}^2 \to \mathbb{R}$ *be defined by*

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

Then there is a constant $c = c_\lambda \in (0,2)$ *such that*

$$\|\nabla f_\lambda(a,b)\|^2 \le c_\lambda$$

for all nonzero vectors $(a,b)^T \in \mathbb{R}^2$.

Proof. Let $(a,b)^T \in \mathbb{R}^2$ be any nonzero vector. Then it follows from (3) that f_λ is continuously differentiable at (a,b), and after some computation we therefore obtain

$$\begin{aligned} \|\nabla f_\lambda(a,b)\|^2 &= \left(\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}}\right)^2 + \left(\frac{-2(a-b)+\lambda a}{2\sqrt{(a-b)^2+\lambda ab}}\right)^2 \\ &= \frac{8(a-b)^2+8\lambda ab+\lambda^2(a^2+b^2)-4\lambda(a^2+b^2)}{4(a-b)^2+4\lambda ab} \\ &= 2 - \frac{\lambda(4-\lambda)}{4}\frac{a^2+b^2}{a^2+b^2+(\lambda-2)ab}. \end{aligned} \tag{4}$$

Now it is easy to see that

$$\frac{a^2+b^2}{a^2+b^2+(\lambda-2)ab} \geq \frac{1}{2}. \tag{5}$$

To this end, we first note that

$$a^2+b^2+2ab-\lambda ab = (a+b)^2 - \lambda ab \geq 0$$

for all $a,b \in \mathbb{R}$. Therefore, we get

$$2a^2+2b^2 \geq a^2+b^2+\lambda ab - 2ab = (a-b)^2 + \lambda ab \geq 0$$

for all $a,b \in \mathbb{R}$, where the last inequality follows from (2). This immediately implies (5).

We now obtain from (4) and (5)

$$\|\nabla f_\lambda(a,b)\|^2 \leq 2 - \frac{\lambda(4-\lambda)}{8} =: c_\lambda.$$

Obviously, c_λ is a constant such that $c_\lambda \in (0,2)$. □

Based on the previous lemma, we are now in the position to present an overestimate of Clarke's [2] generalized Jacobian of Φ_λ at an arbitrary point $x \in \mathbb{R}^n$.

Proposition 2.5 *For an arbitrary $x \in \mathbb{R}^n$, we have*

$$\partial\Phi_\lambda(x) \subseteq D_a(x) + D_b(x)F'(x),$$

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

$$a_i(x) = \frac{2(x_i - F_i(x)) + \lambda F_i(x)}{2\sqrt{(x_i-F_i(x))^2 + \lambda x_i F_i(x)}} - 1, \quad b_i(x) = \frac{-2(x_i - F_i(x)) + \lambda x_i}{2\sqrt{(x_i-F_i(x))^2 + \lambda x_i F_i(x)}} - 1$$

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

$$a_i(x) = \xi_i - 1, \;\; b_i(x) = \chi_i - 1 \textit{ for any } (\xi_i,\chi_i) \in \mathbb{R}^2 \textit{ such that } \|(\xi_i,\chi_i)\| \leq \sqrt{c_\lambda}$$

if $(x_i, F_i(x)) = (0,0)$, where $c_\lambda \in (0,2)$ denotes the constant from Lemma 2.4.

Proof. By Proposition 2.6.2 (e) in Clarke [2], we have

$$\partial\Phi_\lambda(x)^T \subseteq \partial\Phi_{\lambda,1}(x) \times \ldots \times \partial\Phi_{\lambda,n}(x),$$

where $\Phi_{\lambda,i}$ denotes the ith component function of Φ_λ. If i is such that $(x_i, F_i(x)) \neq (0,0)$, then it is easy to check that $\Phi_{\lambda,i}$ is continuously differentiable, cf. (3). Hence it follows immediately that

$$\begin{aligned}\partial\Phi_{\lambda,i}(x) &= \{\nabla\Phi_{\lambda,i}(x)\} \\ &= \Bigg\{\left(\frac{2(x_i - F_i(x)) + \lambda F_i(x)}{2\sqrt{(x_i - F_i(x))^2 + \lambda x_i F_i(x)}} - 1\right) e_i + \\ &\quad \left(\frac{-2(x_i - F_i(x)) + \lambda x_i}{2\sqrt{(x_i - F_i(x))^2 + \lambda x_i F_i(x)}} - 1\right) \nabla F_i(x)\Bigg\}.\end{aligned}$$

If, on the other hand, i is such that $(x_i, F_i(x)) = (0,0)$, then we obtain from Lemma 2.4 that each element of $\partial\Phi_{\lambda,i}(x)$ can be represented in the form

$$\{(\xi_i - 1)e_i + (\chi_i - 1)\nabla F_i(x)\},$$

where $(\xi_i, \chi_i) \in \mathbb{R}^2$ is an arbitrary vector such that $\|(\xi_i, \chi_i)\| \leq \sqrt{c_\lambda}$. From these equalities, the statement of the lemma follows easily. □

We stress that, since $c_\lambda < 2$ in view of Lemma 2.4, it cannot happen that the diagonal elements $a_i(x)$ and $b_i(x)$ as defined in Proposition 2.5 are both equal to zero for the same index $i \in I$. However, it is not clear from Proposition 2.5 that these elements are nonpositive (namely for those indices for which $(x_i, F_i(x)) = (0,0)$). We therefore need the following result.

Proposition 2.6 *Any $H \in \partial\Phi_\lambda(x)$ can be written in the form*

$$H = D_a + D_b F'(x),$$

where $D_a, D_b \in \mathbb{R}^{n\times n}$ are negative semidefinite diagonal matrices such that their sum $D_a + D_b$ is negative definite.

Proof. We first show that the statement holds for any matrix from the B-subdifferential

$$\partial_B\Phi_\lambda(x) := \{H \in \mathbb{R}^{n\times n} \,|\, \exists\{x^k\} \subseteq D_{\Phi_\lambda} : \{x^k\} \to x \text{ and } \{\Phi_\lambda'(x)\} \to H\},$$

where D_{Φ_λ} denotes the set of differentiable points of Φ_λ. So let $H \in \partial_B\Phi_\lambda(x)$. By definition, there exists a sequence $\{x^k\}$ converging to x with $(x_i^k, F_i(x^k)) \neq (0,0)$ for all i and all k such that $\{\Phi_\lambda'(x^k)\} \to H$. For each k, there exist negative semidefinite diagonal matrices $D_a^k, D_b^k \in \mathbb{R}^{n\times n}$ with $D_a^k + D_b^k$ negative definite and

$$\Phi_\lambda'(x^k) = D_a^k + D_b^k F'(x^k),$$

see Proposition 2.5. Since it it easy to see that the elements of D_a^k and D_b^k are bounded for all k, there exist subsequences $\{D_a^k\}_K$ and $\{D_b^k\}_K$ converging to diagonal matrices D_a and D_b, respectively. Obviously, D_a and D_b are also negative semidefinite. Moreover, their sum $D_a + D_b$ is negative definite since $c_\lambda < 2$ by Lemma 2.4. Hence, taking the limit $k \to \infty$ on the subset K gives

$$H = \lim_{k \in K} \Phi'_\lambda(x^k) = \lim_{k \in K}(D_a^k + D_b^k F'(x^k)) = D_a + D_b F'(x).$$

This is the desired representation for an arbitrary element $H \in \partial_B \Phi_\lambda(x)$.

Now let H be an element from the generalized Jacobian $\partial \Phi_\lambda(x)$. By definition, $\partial \Phi_\lambda(x)$ is the convex hull of the B-subdifferential $\partial_B \Phi_\lambda(x)$. Hence there is an integer $m > 0$ and numbers $\lambda_i \geq 0$ with $\sum_{i=1}^m \lambda_i = 1$ as well as matrices $H_i \in \partial_B \Phi_\lambda(x)$ with

$$H = \sum_{i=1}^m \lambda_i H_i.$$

In view of the first part of the proof, each H_i can be written in the form

$$H_i = D_a^i + D_b^i F'(x)$$

with negative semidefinite diagonal matrices D_a^i, D_b^i such that their sum is negative definite. Therefore

$$H = \sum_{i=1}^m \lambda_i H_i = \left(\sum_{i=1}^m \lambda_i D_a^i \right) + \left(\sum_{i=1}^m \lambda_i D_b^i \right) F'(x) =: D_a + D_b F'(x)$$

with $D_a := \sum_{i=1}^m \lambda_i D_a^i$ and $D_b := \sum_{i=1}^m \lambda_i D_b^i$. From the corresponding properties of D_a^i and D_b^i, it is very easy to see that also D_a, D_b are negative semidefinite diagonal matrices and that their sum is negative definite. This completes the proof. □

We next present a new characterization of the class of P-matrices which will turn out to be useful in the proof of the subsequent result. We note, however, that the sufficiency part of this result can also be found in Yamashita and Fukushima [31], and that Gabriel and Moré [16] recently proved a similar characterization.

Before stating our new characterization, we first recall that a matrix $M \in \mathbb{R}^{n \times n}$ is called a P-matrix if, for every nonzero vector $x \in \mathbb{R}^n$, there is an index $i_0 = i_0(x) \in I$ such that $x_{i_0}[Mx]_{i_0} > 0$. For a number of equivalent formulations, we refer the interested reader to the excellent book [3] by Cottle, Pang and Stone.

Proposition 2.7 *A matrix of the form*

$$D_a + D_b M$$

is nonsingular for all positive (negative) semidefinite diagonal matrices $D_a, D_b \in \mathbb{R}^{n \times n}$ such that $D_a + D_b$ is positive (negative) definite if and only if $M \in \mathbb{R}^{n \times n}$ is a P-matrix.

Proof. The proof of the sufficiency part is a generalization of the proof given for Theorem 3.3 in [15]: Assume that M is a P-matrix. Let $D_a, D_b \in \mathbb{R}^{n \times n}$ be any two positive semidefinite diagonal matrices such that their sum $D_a + D_b$ is positive definite. Let $p \in \mathbb{R}^n$ be an arbitrary vector such that

$$(D_a + D_b M)p = 0. \tag{6}$$

Let $D_a = diag(a_1, \dots, a_n)$ and $D_b = diag(b_1, \dots, b_n)$. Then, equation (6) can be rewritten as

$$a_i p_i + b_i [Mp]_i = 0 \quad (i \in I).$$

Multiplying the ith equation by p_i $(i \in I)$ yields

$$a_i (p_i)^2 + b_i p_i [Mp]_i = 0 \quad (i \in I). \tag{7}$$

Assume that $p \neq 0$. Since M is a P-matrix, we therefore have the existence of an index $i_0 \in I$ such that $p_{i_0} \neq 0$ and $p_{i_0}[Mp]_{i_0} > 0$. Since $a_{i_0}, b_{i_0} \geq 0$ and $a_{i_0} + b_{i_0} > 0$ by our assumptions on the diagonal matrices D_a and D_b, we therefore get from (7) the contradiction

$$0 = a_{i_0}(p_{i_0})^2 + b_{i_0} p_{i_0} [Mp]_{i_0} > 0.$$

Hence $p = 0$ and $D_a + D_b M$ is indeed nonsingular.

To prove the converse result, assume that M is not a P-matrix. Then there is a vector $p \neq 0$ such that for all $i \in I$ we have $p_i = 0$ or $p_i[Mp]_i \leq 0$. If $p_i = 0$, then let $a_i := 1$ and $b_i := 0$. Otherwise, define $a_i := |[Mp]_i|$ and $b_i := |p_i|$. Let $D_a := diag(a_1, \dots, a_n)$ and $D_b := diag(b_1, \dots, b_n)$. Obviously, the diagonal matrices D_a and D_b are positive semidefinite. Furthermore, their sum $D_a + D_b$ is positive definite; this follows immediately from the definitions of a_i and b_i $(i \in I)$. Moreover, it is not difficult to see that $(D_a + D_b M)p = 0$ holds, which proves the necessary part of the lemma. □

In the next result, we want to show that all elements in the generalized Jacobian $\partial \Phi_\lambda(x^*)$ are nonsingular if x^* is an R-regular solution of NCP(F). We want to prove this result by exploiting the characterization of P-matrices from Proposition 2.7.

First of all, however, we recall the definition of R-regularity. To this end, given a fixed solution x^* of NCP(F), we use the index sets

$$\begin{aligned} \alpha &:= \{i \in I \,|\, x_i^* > 0 = F_i(x^*)\}, \\ \beta &:= \{i \in I \,|\, x_i^* = 0 = F_i(x^*)\}, \\ \gamma &:= \{i \in I \,|\, x_i^* = 0 < F_i(x^*)\}. \end{aligned}$$

Then x^* is called an *R-regular* solution of NCP(F) if the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular and if the Schur-complement

$$F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha} F'(x^*)_{\alpha\alpha}^{-1} F'(x^*)_{\alpha\beta} \in \mathbb{R}^{|\beta| \times |\beta|}$$

is a P-matrix, cf. Robinson [28].

Theorem 2.8 *Assume that $x^* \in \mathbb{R}^n$ is an R-regular solution of NCP(F). Then all elements in the generalized Jacobian $\partial\Phi_\lambda(x^*)$ are nonsingular.*

Proof. Let $H \in \partial\Phi_\lambda(x^*)$ be arbitrary but fixed. Due to Lemma 2.5, there are diagonal matrices $D_a := D_a(x^*) \in \mathbb{R}^{n\times n}$ and $D_b := D_b(x^*) \in \mathbb{R}^{n\times n}$ such that

$$H = D_a + D_b F'(x^*). \tag{8}$$

Without loss of generality, we assume that we can write

$$F'(x^*) = \begin{pmatrix} F'(x^*)_{\alpha\alpha} & F'(x^*)_{\alpha\beta} & F'(x^*)_{\alpha\gamma} \\ F'(x^*)_{\beta\alpha} & F'(x^*)_{\beta\beta} & F'(x^*)_{\beta\gamma} \\ F'(x^*)_{\gamma\alpha} & F'(x^*)_{\gamma\beta} & F'(x^*)_{\gamma\gamma} \end{pmatrix} \tag{9}$$

and, similarly,

$$D_a = \begin{pmatrix} D_{a,\alpha} & 0 & 0 \\ 0 & D_{a,\beta} & 0 \\ 0 & 0 & D_{a,\gamma} \end{pmatrix}, \tag{10}$$

$$D_b = \begin{pmatrix} D_{b,\alpha} & 0 & 0 \\ 0 & D_{b,\beta} & 0 \\ 0 & 0 & D_{b,\gamma} \end{pmatrix}, \tag{11}$$

where $D_{a,\alpha} := (D_a)_{\alpha\alpha}$ etc. Now let $p \in \mathbb{R}^n$ be an arbitrary vector with $Hp = 0$. If we partition the vector p appropriately as $p = (p_\alpha, p_\beta, p_\gamma)$, and if we take into account (8)–(11), the homogeneous linear system $Hp = 0$ can be rewritten as follows:

$$D_{a,\alpha}p_\alpha + D_{b,\alpha}\left(F'(x^*)_{\alpha\alpha}p_\alpha + F'(x^*)_{\alpha\beta}p_\beta + F'(x^*)_{\alpha\gamma}p_\gamma\right) = 0_\alpha, \tag{12}$$

$$D_{a,\beta}p_\beta + D_{b,\beta}\left(F'(x^*)_{\beta\alpha}p_\alpha + F'(x^*)_{\beta\beta}p_\beta + F'(x^*)_{\beta\gamma}p_\gamma\right) = 0_\beta, \tag{13}$$

$$D_{a,\gamma}p_\gamma + D_{b,\gamma}\left(F'(x^*)_{\gamma\alpha}p_\alpha + F'(x^*)_{\gamma\beta}p_\beta + F'(x^*)_{\gamma\gamma}p_\gamma\right) = 0_\gamma. \tag{14}$$

From Proposition 2.5, we obtain

$$D_{a,\alpha} = 0_\alpha, D_{a,\gamma} = -\kappa I_\gamma, D_{b,\alpha} = -\kappa I_\alpha, \text{ and } D_{b,\gamma} = 0_\gamma$$

for a certain constant $\kappa > 0$. Hence (14) immediately gives

$$p_\gamma = 0_\gamma, \tag{15}$$

so that (12) becomes

$$F'(x^*)_{\alpha\alpha}p_\alpha + F'(x^*)_{\alpha\beta}p_\beta = 0_\alpha.$$

Since x^* is an R-regular solution of NCP(F), the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular. Hence we obtain from the previous equation

$$p_\alpha = -F'(x^*)^{-1}_{\alpha\alpha}F'(x^*)_{\alpha\beta}p_\beta. \tag{16}$$

Replacing this in (13), using (15), and rearranging terms yields

$$\left[D_{a,\beta} + D_{b,\beta}\left(F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha}F'(x^*)_{\alpha\alpha}^{-1}F'(x^*)_{\alpha\beta}\right)\right]p_\beta = 0_\beta. \tag{17}$$

Since the Schur-complement

$$F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha}F'(x^*)_{\alpha\alpha}^{-1}F'(x^*)_{\alpha\beta}$$

is a P-matrix in view of the assumed R-regularity of x^*, and since both diagonal matrices $D_{a,\beta}, D_{b,\beta}$ can be assumed to be negative semidefinite such that their sum is negative definite by Proposition 2.6, it follows immediately from (17) and Proposition 2.7 that

$$p_\beta = 0_\beta.$$

But then we also have $p_\alpha = 0_\alpha$ because of (16), so that $p = 0$. This shows that H is nonsingular. □

3 Properties of the Merit Function Ψ_λ

In this section, we investigate the properties of the merit function Ψ_λ. We begin with the following result which will be crucial for the subsequent analysis as well as for the design of our algorithm in the next section.

Theorem 3.1 *The function Ψ_λ is continuously differentiable with $\nabla\Psi_\lambda(x) = H^T\Phi_\lambda(x)$ for any $H \in \partial\Phi_\lambda(x)$.*

Proof. By using Proposition 2.5 and the NCP-function property of φ_λ, the proof can be carried out in essentially the same way as the one for Proposition 3.4 in [8]. □

For the following results, it will be convenient to rewrite the merit function $\Psi_\lambda : \mathbb{R}^n \to \mathbb{R}$ as

$$\Psi_\lambda(x) = \frac{1}{2}\Phi_\lambda(x)^T\Phi_\lambda(x) = \sum_{i\in I}\psi_\lambda(x_i, F_i(x)),$$

where $\psi_\lambda : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$\psi_\lambda(a,b) := \frac{1}{2}(\varphi_\lambda(a,b))^2 = \frac{1}{2}\left(\sqrt{(a-b)^2 + \lambda ab} - a - b\right)^2. \tag{18}$$

In order to establish some important results for the merit function Ψ_λ, we will take a closer look at the properties of ψ_λ. We begin with the following simple lemma.

Lemma 3.2 *ψ_λ is continuously differentiable with $\nabla\psi_\lambda(0,0) = 0$.*

Proof. The continuous differentiability of ψ_λ can either be verified in the same way as for Theorem 3.1 or by direct calculation. The fact that $\nabla\psi_\lambda(0,0) = 0$ then follows from the first order optimality conditions in unconstrained minimization since the origin is a global minimizer of the function ψ_λ. □

The next result states that the two partial derivatives of the function ψ_λ cannot have opposite signs.

Lemma 3.3 *It holds that*

$$\frac{\partial\psi_\lambda}{\partial a}(a,b)\frac{\partial\psi_\lambda}{\partial b}(a,b) \geq 0$$

for all $a, b \in \mathbb{R}$.

Proof. In view of Lemma 3.2, the result holds if $(a,b) = (0,0)$. Hence assume that $(a,b) \neq (0,0)$. Then we can use the chain rule in order to write

$$\frac{\partial\psi_\lambda}{\partial a}(a,b) = \varphi_\lambda(a,b)\left(\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}} - 1\right).$$

We first show that

$$\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}} - 1 \leq 0. \tag{19}$$

Assume the contrary, i.e., assume that

$$2(a-b)+\lambda b > 2\sqrt{(a-b)^2+\lambda ab}.$$

Since the right-hand side of this inequality is nonnegative, we can square both sides in order to obtain

$$4(a-b)^2 + \lambda^2 b^2 + 4\lambda(a-b)b > 4(a-b)^2 + 4\lambda ab.$$

This reduces to

$$\lambda(\lambda-4)b^2 > 0$$

which, however, contradicts the general assumption that $\lambda \in (0,4)$. Hence (19) holds.

On the other hand, we have

$$\frac{\partial\psi_\lambda}{\partial b}(a,b) = \varphi_\lambda(a,b)\left(\frac{-2(a-b)+\lambda a}{2\sqrt{(a-b)^2+\lambda ab}} - 1\right),$$

and in a similar way as in the first part of this proof, we can show that

$$\frac{-2(a-b)+\lambda a}{2\sqrt{(a-b)^2+\lambda ab}} - 1 \leq 0.$$

Thus

$$\begin{aligned}\frac{\partial\psi_\lambda}{\partial a}(a,b)\frac{\partial\psi_\lambda}{\partial b}(a,b) &= (\varphi_\lambda(a,b))^2\left(\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}}-1\right)\left(\frac{-2(a-b)+\lambda a}{2\sqrt{(a-b)^2+\lambda ab}}-1\right)\\ &\geq 0,\end{aligned}$$

as desired. □

The following lemma gives a very strong characterization of a global minimum of the function ψ_λ in terms of the partial derivatives of this function.

Lemma 3.4 *It holds that*

$$\psi_\lambda(a,b)=0 \Longleftrightarrow \nabla\psi_\lambda(a,b)=0 \Longleftrightarrow \frac{\partial\psi_\lambda}{\partial a}(a,b)=0 \Longleftrightarrow \frac{\partial\psi_\lambda}{\partial b}(a,b)=0.$$

Proof. If $\psi_\lambda(a,b)=0$, then (a,b) is a global minimizer of the function ψ_λ. Hence $\nabla\psi_\lambda(a,b)=(0,0)$ by the necessary optimality condition in unconstrained minimization and Lemma 3.2. On the other hand, if (a,b) is a stationary point, then, of course, the two partial derivatives vanish at the point (a,b). So it remains to prove the implication

$$\frac{\partial\psi_\lambda}{\partial a}(a,b)=0 \Longrightarrow \psi_\lambda(a,b)=0 \tag{20}$$

as well as the corresponding implication if the partial derivative with respect to the second argument vanishes. Since the proof of the latter implication is similar to the one of (20), we just consider (20) in the following.

In view of Lemma 3.2, we can assume that $(a,b)\neq(0,0)$. Then we have

$$\frac{\partial\psi_\lambda}{\partial a}(a,b)=\varphi_\lambda(a,b)\left(\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}}-1\right).$$

If $\varphi_\lambda(a,b)=0$, we are done. Hence assume that

$$\frac{2(a-b)+\lambda b}{2\sqrt{(a-b)^2+\lambda ab}}-1=0,$$

i.e.,

$$2(a-b)+\lambda b=2\sqrt{(a-b)^2+\lambda ab}. \tag{21}$$

Squaring both sides yields after some rearrangements:

$$\lambda(\lambda-4)b^2=0.$$

This implies $b=0$ since $\lambda\in(0,4)$. Now, substituting $b=0$ in (21) gives

$$2a=2\sqrt{a^2}=2|a|$$

and therefore $a \geq 0$. However, since φ_λ is an NCP-function, we thus obtain $\varphi_\lambda(a,b) = 0$ and therefore also $\psi_\lambda(a,b) = 0$. This completes the proof. $\square$

We are now in the position to state and prove one of the main results of this section. We only have to recall that a matrix $M \in \mathbb{R}^{n\times n}$ is called a P_0-matrix if, for any nonzero $x \in \mathbb{R}^n$, there is an index $i_0 = i_0(x) \in I$ with $x_{i_0} \neq 0$ and $x_{i_0}[Mx]_{i_0} \geq 0$, see [3].

Theorem 3.5 *Assume that $x^* \in \mathbb{R}^n$ is a stationary point of Ψ_λ such that the Jacobian $F'(x^*)$ is a P_0-matrix. Then x^* is a solution of the nonlinear complementarity problem NCP(F).*

Proof. Based on the previous results, the proof is essentially the same as the one given, e.g., for the (squared) Fischer-function by Facchinei and Soares [8]. For the sake of completeness and in order to show how the previous results apply here, we include a full proof.

We first recall that we can write

$$\Psi_\lambda(x) = \sum_{i\in I} \psi_\lambda(x_i, F_i(x))$$

with ψ_λ being defined in (18).

Now, since $x^* \in \mathbb{R}^n$ is a stationary point, we have

$$0 = \nabla\Psi_\lambda(x^*) = \frac{\partial\psi_\lambda}{\partial a}(x^*, F(x^*)) + F'(x^*)^T \frac{\partial\psi_\lambda}{\partial b}(x^*, F(x^*)), \tag{22}$$

where

$$\frac{\partial\psi_\lambda}{\partial a}(x^*, F(x^*)) := \left(\ldots, \frac{\partial\psi_\lambda}{\partial a}(x_i^*, F_i(x^*)), \ldots\right)^T \in \mathbb{R}^n$$

and, similarly,

$$\frac{\partial\psi_\lambda}{\partial b}(x^*, F(x^*)) := \left(\ldots, \frac{\partial\psi_\lambda}{\partial b}(x_i^*, F_i(x^*)), \ldots\right)^T \in \mathbb{R}^n.$$

Componentwise, equation (22) becomes

$$\frac{\partial\psi_\lambda}{\partial a}(x_i^*, F_i(x^*)) + \left[F'(x^*)^T \frac{\partial\psi_\lambda}{\partial b}(x^*, F(x^*))\right]_i = 0 \quad \forall i \in I.$$

Premultiplying the ith equation by $\frac{\partial\psi_\lambda}{\partial b}(x_i^*, F_i(x^*))$ yields

$$\frac{\partial\psi_\lambda}{\partial a}(x_i^*, F_i(x^*))\frac{\partial\psi_\lambda}{\partial b}(x_i^*, F_i(x^*)) + \frac{\partial\psi_\lambda}{\partial b}(x_i^*, F_i(x^*)) \left[F'(x^*)^T \frac{\partial\psi_\lambda}{\partial b}(x^*, F(x^*))\right]_i = 0 \tag{23}$$

for all $i \in I$. Now assume that $\frac{\partial \psi_\lambda}{\partial b}(x^*, F(x^*)) \neq 0$. By assumption, $F'(x^*)$ and therefore also $F'(x^*)^T$ is a P_0-matrix. Hence there is an index $i_0 \in I$ such that $\frac{\partial \psi_\lambda}{\partial b}(x^*_{i_0}, F_{i_0}(x^*)) \neq 0$ and

$$\frac{\partial \psi_\lambda}{\partial b}(x^*_{i_0}, F_{i_0}(x^*)) \left[F'(x^*)^T \frac{\partial \psi_\lambda}{\partial b}(x^*, F(x^*)) \right]_{i_0} \geq 0. \tag{24}$$

On the other hand, we also have

$$\frac{\partial \psi_\lambda}{\partial a}(x^*_{i_0}, F_{i_0}(x^*)) \frac{\partial \psi_\lambda}{\partial b}(x^*_{i_0}, F_{i_0}(x^*)) \geq 0 \tag{25}$$

by Lemma 3.3. In view of (23)–(25), we obtain

$$\frac{\partial \psi_\lambda}{\partial a}(x^*_{i_0}, F_{i_0}(x^*)) \frac{\partial \psi_\lambda}{\partial b}(x^*_{i_0}, F_{i_0}(x^*)) = 0.$$

Lemma 3.4 therefore implies

$$\psi_\lambda(x^*_{i_0}, F_{i_0}(x^*)) = 0.$$

This, in turn, implies $\frac{\partial \psi_\lambda}{\partial b}(x^*_{i_0}, F_{i_0}(x^*)) = 0$ by Lemma 3.4, a contradiction to the choice of the index $i_0 \in I$. Hence $\frac{\partial \psi_\lambda}{\partial b}(x^*, F(x^*)) = 0$ which, again by Lemma 3.4, implies $\Psi_\lambda(x^*) = 0$ so that x^* solves NCP(F). □

Lemma 3.6 *There exist constants $c_1 > 0$ and $c_2 > 0$ such that*

$$c_1 |\min\{a, b\}| \leq |\varphi_\lambda(a, b)| \leq c_2 |\min\{a, b\}| \tag{26}$$

for all $a, b \in \mathbb{R}$.

Proof. Let $\lambda \in (0, 4)$ be fixed. We will show that the inequalities (26) hold with the following constants:

$$c_1 := c_1(\lambda) := 1 - \lambda/4 < 1 \text{ and } c_2 := c_2(\lambda) := 2 + \sqrt{\lambda} > 2.$$

First note that we have

$$0 < c_1^2 < c_1^2 + 2c_1 < 4c_1 = 4 - \lambda \tag{27}$$

and

$$0 < \lambda + 2(c_2 - 2). \tag{28}$$

Now we consider four cases:
Case 1: $a \geq 0, b \geq 0$.
Then it is easy to see that $\varphi_\lambda(a, b) \leq 0$. Hence

$$|\varphi_\lambda(a, b)| = a + b - \sqrt{(a - b)^2 + \lambda ab}.$$

Since $ab \geq 0$, we therefore obtain

$$|\varphi_\lambda(a,b)| \leq a+b-\sqrt{(a-b)^2} = a+b-|a-b| = 2|\min\{a,b\}| \leq c_2|\min\{a,b\}|.$$

In order to derive the left-hand inequality in (26), we first note that

$$\begin{aligned} & 2c_1(a+b)\min\{a,b\} \leq 2c_1(ab+ab) = 4c_1 ab \\ = \; & (4-\lambda)ab \leq (4-\lambda)ab + c_1^2\min^2\{a,b\}, \end{aligned}$$

so that

$$(a+b)^2-(4-\lambda)ab \leq c_1^2\min^2\{a,b\} - 2c_1(a+b)\min\{a,b\} + (a+b)^2.$$

Since $c_1\min\{a,b\} \leq \min\{a,b\} \leq a+b$, this implies

$$\sqrt{(a-b)^2+\lambda ab} \leq |c_1\min\{a,b\} - (a+b)| = a+b-c_1\min\{a,b\}$$

and therefore

$$c_1|\min\{a,b\}| = c_1\min\{a,b\} \leq a+b-\sqrt{(a-b)^2+\lambda ab} = |\varphi_\lambda(a,b)|.$$

Case 2: $a<0, b<0$.
Then we have $\varphi_\lambda(a,b) > 0$ so that

$$|\varphi_\lambda(a,b)| = \sqrt{(a-b)^2+\lambda ab} - a - b.$$

This implies

$$|\varphi_\lambda(a,b)| > \sqrt{(a-b)^2} - a - b = |a-b| - a - b = 2|\min\{a,b\}| > c_1|\min\{a,b\}|$$

since $ab > 0$. In order to prove the right-hand inequality of (26), we assume without loss of generality that $|a| \geq |b|$. Since $\lambda = (c_2-2)^2$ in view of the definition of c_2, we obtain from (28) the inequality

$$0 < (\lambda+2c_2-4)|a|\,|b| = (c_2^2-2c_2)|a|\,|b| \leq (c_2^2-2c_2)|a|^2$$

and therefore

$$(\lambda-4)|a|\,|b| \leq c_2^2|a|^2 - 2c_2|a|(|a|+|b|) = c_2^2|a|^2 + 2c_2|a|(a+b).$$

This implies

$$(a+b)^2 + (\lambda-4)ab \leq (c_2|a|+a+b)^2$$

which, since $c_2|a|+a+b = c_2|a|-|a|-|b| \geq (c_2-2)|a| \geq 0$, gives

$$\sqrt{(a-b)^2+\lambda ab} - a - b \leq c_2|a| = c_2\max\{|a|,|b|\} = c_2|\min\{a,b\}|.$$

Case 3: $a \geq 0, b < 0$.
In this case we have $ab \leq 0$ and $\varphi_\lambda(a,b) \geq 0$, so that

$$|\varphi_\lambda(a,b)| = \sqrt{(a-b)^2 + \lambda ab} - a - b.$$

We therefore have

$$|\varphi_\lambda(a,b)| \leq \sqrt{(a-b)^2} - a - b = 2|\min\{a,b\}| < c_2|\min\{a,b\}|.$$

In order to verify the other inequality in (26), we consider two subcases.
Subcase 3a: $a + b > 0$.
Then $a > 0, b = \min\{a,b\} \leq 0$ and therefore because of (27):

$$\begin{aligned} 0 &< c_1^2(a+b) - 2c_1 b = c_1^2 a + 2c_1 a + c_1^2 b - 2c_1(a+b) \\ &< (4-\lambda)a - c_1^2|b| - 2c_1(a+b) \end{aligned}$$

so that

$$c_1^2|b|^2 + 2c_1|b|(a+b) \leq (4-\lambda)a|b| = -(4-\lambda)ab.$$

Hence

$$c_1^2|b|^2 + 2c_1|b|(a+b) + (a+b)^2 \leq (a+b)^2 - (4-\lambda)ab = (a-b)^2 + \lambda ab.$$

Taking the square root on both sides gives

$$0 \leq c_1|b| + a + b \leq \sqrt{(a-b)^2 + \lambda ab}.$$

Hence we obtain

$$c_1|\min\{a,b\}| = c_1|b| \leq \sqrt{(a-b)^2 + \lambda ab} - a - b = |\varphi_\lambda(a,b)|.$$

Subcase 3b: $a + b \leq 0$.
Because of (27), we have:

$$0 \leq a^2 c_1^2 \leq -abc_1^2 \leq -ab(4-\lambda)$$

so that

$$ac_1 \leq \sqrt{-ab(4-\lambda)} \leq \sqrt{(a+b)^2 - ab(4-\lambda)} = \sqrt{(a-b)^2 + \lambda ab}.$$

On the other hand, we also have $(1-c_1)(a+b) \leq 0$ because of $c_1 < 1$. Hence $a + b - c_1 b \leq ac_1$ and therefore

$$a + b - c_1 b \leq ac_1 \leq \sqrt{(a-b)^2 + \lambda ab}.$$

This implies

$$c_1|\min\{a,b\}| = c_1|b| = -c_1 b \leq \sqrt{(a-b)^2 + \lambda ab} - a - b = |\varphi_\lambda(a,b)|.$$

Case 4: $a < 0, b \geq 0$.
This case can be verified in a similar way as Case 3, so we omit the details here. □

Note that, in particular, Lemma 3.6 implies a result by Tseng [30] who compares the growth behaviour of the Fischer function with the growth behaviour of the minimum function.

As a consequence of Lemma 3.6 and known results for the minimum function (see, e.g., Facchinei and Soares [8] as well as Kanzow and Fukushima [19]), we therefore obtain the following two theorems. In the statement of these results, we call the function $F : \mathbb{R}^n \to \mathbb{R}^n$ a uniform P-function if there is a modulus $\mu > 0$ such that

$$\max_{i \in I}(x_i - y_i)(F_i(x) - F_i(y)) \geq \mu \|x - y\|^2$$

holds for all $x, y \in \mathbb{R}^n$.

Theorem 3.7 *If F is a uniform P-function, then the level sets*

$$\mathcal{L}(c) := \{x \in \mathbb{R}^n |\, \Psi_\lambda(x) \leq c\}$$

are compact for any fixed $c \in \mathbb{R}$.

Theorem 3.8 *If F is a Lipschitz-continuous uniform P-function, then there exists a constant $c_3 > 0$ such that*

$$\|x - x^*\|^2 \leq c_3 \Psi_\lambda(x)$$

for all $x \in \mathbb{R}^n$, where x^* denotes the unique solution of NCP(F).

We note that Theorem 3.7 is important because it guarantees that any sequence generated by our algorithm to be presented in the following section will be bounded under the stated assumptions. On the other hand, Theorem 3.8 says that the function $\sqrt{\Psi_\lambda}$ provides a global error bound for the nonlinear complementarity problem NCP(F).

4 Algorithm and Convergence

The following algorithm tries to solve the nonlinear complementarity problem by solving the equivalent nonlinear system of equations

$$\Phi_\lambda(x) = 0.$$

It applies a nonsmooth Newton-type method as introduced and investigated by Kummer [20], Qi and Sun [27] as well as Qi [26] to this nonlinear system of equations. The method is globalized by using the smooth merit function Ψ_λ.

We stress that this algorithm looks very similar to the classical Newton method applied to a smooth system of equations. The only difference is that, here, we have to choose an element $H \in \partial\Phi_\lambda(x)$ instead of taking the (usually not existing) classical Jacobian of Φ_λ at x.

The following is a precise statement of our algorithm.

Algorithm 4.1 *(Semismooth Newton-type Method)*

(S.0) *(Initialization)*
Choose $\lambda \in (0,4), x^0 \in \mathbb{R}^n, \rho > 0, \beta \in (0,1), \sigma \in (0,1/2), p > 2, \varepsilon \geq 0,$ *and set* $k := 0$.

(S.1) *(Termination Criterion)*
If $\|\nabla\Psi_\lambda(x^k)\| \leq \varepsilon$, *stop.*

(S.2) *(Search Direction Calculation)*
Select an element $H_k \in \partial\Phi_\lambda(x^k)$. *Find a solution* $d^k \in \mathbb{R}^n$ *of the linear system*

$$H_k d = -\Phi_\lambda(x^k).$$

If this system is not solvable or if the descent condition

$$\nabla\Psi_\lambda(x^k)^T d^k \leq -\rho\|d^k\|^p$$

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

(S.3) *(Line Search)*
Compute $t_k := \max\{\beta^\ell \mid \ell = 0,1,2,\ldots\}$ *such that*

$$\Psi_\lambda(x^k + t_k d^k) \leq \Psi_\lambda(x^k) + \sigma t_k \nabla\Psi_\lambda(x^k)^T d^k.$$

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

The global and local convergence properties of Algorithm 4.1 are summarized in the following theorem. We implicitly assume in this result that the termination parameter ε in Algorithm 4.1 is equal to zero, and that the algorithm generates an infinite sequence.

Theorem 4.2 *(a) Every accumulation point of a sequence* $\{x^k\}$ *generated by Algorithm 4.1 is a stationary point of* Ψ_λ.

(b) Suppose that x^* *is an isolated accumulation point of a sequence* $\{x^k\}$ *generated by Algorithm 4.1. Then the entire sequence* $\{x^k\}$ *converges to the point* x^*.

(c) Assume that x^* *is an accumulation point of a sequence* $\{x^k\}$ *generated by Algorithm 4.1 such that* x^* *is an R-regular solution of NCP(F). Then:*

(i) The entire sequence $\{x^k\}$ *converges to* x^*.

(ii) The search direction d^k *is eventually given by the solution of the linear system* $H_k d = -\Phi_\lambda(x^k)$ *in Step (S.2).*

(iii) The full stepsize $t_k = 1$ *is accepted for all* k *sufficiently large.*

(iv) The rate of convergence is Q-superlinear.

(v) If, in addition, F *is an* LC^1 *mapping, then the rate of convergence is Q-quadratic.*

Proof. Based on our results stated in Sections 2 and 3, the proof is basically the same as the one for Theorem 3.1 in De Luca, Facchinei and Kanzow [4] for a related algorithm. The only difference is that our statement (b) is slightly different, so that we have to ensure that the R-regularity condition in part (c) implies that x^* is an isolated accumulation point of the sequence $\{x^k\}$.

Suppose it is not. Since $\{\Psi_\lambda(x^k)\}$ is monotonically decreasing, and since the accumulation point x^* solves NCP(F), it follows that $\{\Psi_\lambda(x^k)\} \to 0$. Hence each accumulation point of the sequence $\{x^k\}$ is a solution of NCP(F). Since, however, an R-regular solution is known to be locally unique, x^* is necessarily also an isolated accumulation point of the sequence $\{x^k\}$. But then part (i) of (c) follows from part (b).

By using Theorems 2.3 and 3.1, we can prove all other statements in exactly the same way as for the corresponding Theorem in [4] (see also [6, 18]). □

Note that Theorem 4.2 (a) and (b) only give a global convergence result to stationary points of the merit function Ψ_λ whereas we are much more interested in finding a global minimizer of Ψ_λ and hence a solution of NCP(F). Fortunately, Theorem 3.5 provides a rather weak assumption for such a stationary point to be a solution of NCP(F). We stress that the P_0-matrix property used in Theorem 3.5 is satisfied, in particular, for the large class of P_0-functions and therefore especially for all monotone functions F, see [22].

The existence of an accumulation point and thus of a stationary point of Ψ_λ is guaranteed by Theorem 3.7. In view of our numerical experience, however, the sequence $\{x^k\}$ generated by Algorithm 4.1 remains bounded and therefore admits an accumulation point for all test problems, even if the assumptions from Theorem 3.7 are not met.

Finally, we stress that the local convergence theory for Algorithm 4.1 as given in part (c) of Theorem 4.2 does not require any nondegeneracy assumption for the solution x^*.

5 Numerical Results

We implemented Algorithm 4.1 in MATLAB and tested it on a SUN SPARC 20 station using basically all complementarity problems and all available starting points from the test problem collection MCPLIB by Dirkse and Ferris [5], see also Ferris and Rutherford [12]. Two exceptions are the von Thünen problems `pgvon105` and `pgvon106` whose implementations within the MCPLIB library seems to be somewhat crucial [9] so that we deleted them from our tables. Nevertheless, we stress that our algorithm was able to solve problem `pgvon105` with an accuracy of $\Psi(x^k) \approx 10^{-13}$ after $k = 44$ iterations, whereas for problem `pgvon106` the algorithm stopped after a few iterations with $\Psi(x^k) \approx 10^{-5}$ since the steplength was getting too small.

Within our implementation, we incorporated some (partially heuristic) strategies in order to improve the numerical results to some extent. In particular, we used the following strategies:

(a) We replaced the monotone Armijo rule from Step (S.3) of Algorithm 4.1 by a nonmonotone line search as suggested by Grippo, Lampariello and Lucidi [17] and further investigated by Toint [29] for unconstrained minimization.

(b) We implemented a simple backtracking strategy as described, e.g., in [7], in order to avoid possible domain violations (note that the function F in many examples from the MCPLIB library is not defined everywhere).

Since the above two strategies are more or less standard techniques within our algorithmic framework, we do not restate them here precisely. Instead, we mention that we terminated our iteration if one of the following conditions was satisfied:

$$k > k_{\max}, \Psi(x^k) < \varepsilon \text{ or } t_k < t_{\min},$$

and that we used the following settings for our parameters:

$$\rho = 10^{-8}, \beta = 0.5, \sigma = 10^{-4}, p = 2.1,$$

and

$$k_{\max} = 200, t_{\min} = 10^{-12}, \varepsilon = 10^{-12}.$$

Using this algorithmic environment, we made some preliminary test runs using different values of our parameter λ. In view of these preliminary experiments, it seems that the choice $\lambda = 2$, i.e., the Fischer function, gives the best global convergence properties, whereas taking λ close to zero usually results in a faster local convergence.

This motivated us to use a dynamic choice of λ for our test runs. More precisely, we update λ at each iteration using the following rules:

(a) Set $\lambda = 2$ at the beginning of each iteration.

(b) If $\Psi(x^k) \leq \gamma_1$, then set $\lambda := \Psi(x^k)$, else set $\lambda := \min\{c_1\Psi(x^k), \lambda\}$.

(c) If $\Psi(x^k) \leq \gamma_2$, then set $\lambda := \min\{c_2, \lambda\}$.

Basically, this strategy uses the value $\lambda = 2$ as long as we are far away from a solution, and reduces λ fast if we are getting close to a solution of the nonlinear complementarity problem NCP(F).

Our implementation uses the following values for the parameters involved in this updating scheme for λ :

$$\gamma_1 = 10^{-2}, \gamma_2 = 10^{-4}, c_1 = 10, \text{ and } c_2 = 10^{-8}.$$

The numerical results which we obtained with this algorithm are summarized in Table 1. In this table, we report the following data:

problem:	name of the test problem in MCPLIB
n:	number of variables
SP:	number of starting point provided by MCPLIB
k:	number of iterations
F-ev.:	number of F-evaluations
F'-ev.:	number of Jacobian evaluations
$\Psi(x^f)$:	value of $\Psi(x)$ at the final iterate $x = x^f$
$\\|\nabla\Psi(x^f)\\|$:	value of $\\|\nabla\Psi(x)\\|$ at the final iterate $x = x^f$
G:	number of gradient steps
N:	number of Newton steps.

The results reported in Table 1 are, in our opinion, quite impressive. The algorithm was able to solve almost all examples, most of them without any difficulties. For example, the famous Hansen-Koopmans problem `hanskoop`, which is supposed to be a difficult problem, was solved in approximately ten iterations for all starting points.

There are just three examples which our algorithm was not able to solve. The first one is the `billups` example which, in fact, was constructed by Billups [1] in order to make almost all state-of-the-art methods to fail on this problem. By incorporating a simple hybrid technique as discussed in Chapter 8 of the manuscript [18], it is no problem to solve even this example.

The second failure is on problem `colvdual` (second starting point). This problem is known to be very hard. We were able to solve it by using different parameter settings, but then we usually got difficulties with some other test problems.

Finally, the third failure is on the Kojima-Shindo problem `kojshin` when using the fourth starting point. Here we have a failure due to our nonmonotone line search rule since this problem can be solved within a very few iterations when using the standard (monotone) Armijo rule. So this is one of the rare situations where the nonmonotone line search is worse than its monotone counterpart.

6 Concluding Remarks

In this paper, we introduced a new class of NCP-functions which, in particular, covers the well-known Fischer function. We further investigated the theoretical and

Table 1: Numerical results for Algorithm 4.1.

problem	n	SP	k	F-ev.	F'-ev.	$\Psi(x^f)$	$\|\nabla\Psi(x^f)\|$	G	N
bertsekas	15	1	30	230	31	1.5e-17	1.2e-6	1	29
bertsekas	15	2	28	205	29	7.9e-17	8.2e-7	0	28
bertsekas	15	3	12	24	13	7.3e-15	1.2e-5	0	12
billups	1	1	—	—	—	—	—	—	—
colvdual	20	1	33	81	34	1.8e-14	9.1e-5	0	33
colvdual	20	2	—	—	—	—	—	—	—
colvnlp	15	1	14	28	15	4.5e-17	2.3e-6	0	14
colvnlp	15	2	16	18	17	2.7e-25	1.2e-10	0	16
cycle	1	1	3	5	4	3.5e-17	1.7e-8	0	3
explcp	16	1	19	24	20	1.3e-26	3.2e-13	0	19
hanskoop	14	1	9	11	10	5.9e-13	1.7e-5	1	8
hanskoop	14	2	11	15	12	7.0e-14	9.0e-6	1	10
hanskoop	14	3	9	11	10	2.3e-15	1.1e-6	1	8
hanskoop	14	4	8	12	9	4.9e-14	2.7e-6	0	8
hanskoop	14	5	17	21	18	5.1e-17	2.4e-7	1	16
josephy	4	1	27	36	28	1.7e-17	1.1e-7	0	27
josephy	4	2	6	10	7	5.8e-13	1.9e-5	0	6
josephy	4	3	59	71	60	1.6e-18	3.3e-8	0	59
josephy	4	4	5	6	6	6.6e-25	2.1e-11	0	5
josephy	4	5	4	5	5	1.4e-17	9.7e-8	0	4
josephy	4	6	7	10	8	6.2e-17	2.0e-7	0	7
kojshin	4	1	14	15	15	2.3e-27	1.2e-12	0	14
kojshin	4	2	7	12	8	9.9e-26	1.1e-11	0	7
kojshin	4	3	11	12	12	4.4e-14	7.4e-6	0	11
kojshin	4	4	—	—	—	—	—	—	—
kojshin	4	5	5	6	6	4.6e-26	5.5e-12	0	5
kojshin	4	6	6	8	7	2.5e-27	1.7e-12	0	6
mathinum	3	1	4	5	5	2.1e-15	2.6e-7	0	4
mathinum	3	2	5	6	6	5.8e-15	5.9e-7	0	5
mathinum	3	3	8	10	9	0.0	0.0	0	8
mathinum	3	4	7	8	8	2.3e-19	3.7e-9	0	7
mathisum	4	1	5	7	6	1.1e-19	2.2e-9	0	5
mathisum	4	2	6	7	7	1.4e-13	2.5e-6	0	6
mathisum	4	3	9	12	10	4.0e-29	3.1e-14	0	9
mathisum	4	4	6	7	7	4.7e-24	1.4e-11	0	6
nash	10	1	8	9	9	2.7e-24	3.4e-10	0	8
nash	10	2	9	15	10	7.4e-21	1.8e-8	0	9

Table 1 (continued): Numerical results for Algorithm 4.1.

problem	n	SP	k	F-ev.	F'-ev.	$\Psi(x^f)$	$\|\nabla\Psi(x^f)\|$	G	N
powell	16	1	9	11	10	1.5e-18	4.2e-9	0	9
powell	16	2	11	14	12	1.5e-23	9.2e-11	0	11
powell	16	3	20	21	21	3.7e-13	8.6e-6	1	19
powell	16	4	10	11	11	2.2e-15	1.1e-6	1	9
scarfanum	13	1	8	11	9	7.6e-20	1.3e-8	0	8
scarfanum	13	2	10	17	11	4.4e-17	2.9e-7	0	10
scarfanum	13	3	8	11	9	7.7e-15	4.9e-6	0	8
scarfasum	14	1	6	8	7	7.0e-19	1.8e-7	0	6
scarfasum	14	2	9	16	10	1.9e-17	5.4e-7	0	9
scarfasum	14	3	10	16	11	3.6e-18	3.6e-7	0	10
scarfbnum	39	1	19	24	20	1.0e-26	4.3e-11	0	19
scarfbnum	39	2	25	36	26	5.8e-28	1.2e-11	1	24
scarfbsum	40	1	20	29	21	1.4e-15	1.3e-5	0	20
scarfbsum	40	2	28	36	29	1.4e-14	4.1e-5	4	24
sppe	27	1	7	8	8	3.1e-14	1.2e-6	0	7
sppe	27	2	8	40	9	6.7e-13	6.8e-6	2	6
tobin	42	1	9	11	10	1.3e-13	2.1e-6	0	9
tobin	42	2	14	15	15	2.2e-13	3.0e-6	0	14

numerical properties of a corresponding nonsmooth Newton-type method. Theoretically, it turned out that our whole class of algorithms has essentially the same (strong) properties as the Fischer function. On the other hand, the numerical results presented in the previous section are quite promising, but still preliminary, and we believe that a more sophisticated dynamic choice of our parameter could lead to substantial improvements. We leave this as a future research topic.

We finally note that the results of this paper do not hold for the boundary values $\lambda = 0$ and $\lambda = 4$. In fact, if $\lambda = 0$, then neither ψ_λ nor Ψ_λ are continuously differentiable, so that Algorithm 4.1 is not even well-defined in this case, whereas for $\lambda = 4$ it is easy to see that $\varphi_\lambda(a, b) = 0$ for all $a, b \geq 0$ so that φ_4 is not an NCP-function.

Acknowledgement. The authors would like to thank Martin Zupke for pointing out a gap in an earlier version of this paper. His comments led us to the introduction of Proposition 2.6.

References

[1] S.C. Billups: *Algorithms for Complementarity Problems and Generalized*

Equations. Ph.D. Thesis, Computer Sciences Department, University of Wisconsin, Madison, WI, August 1995.

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

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

[4] T. De Luca, F. Facchinei and C. Kanzow: *A semismooth equation approach to the solution of nonlinear complementarity problems.* Mathematical Programming 75, 1996, pp. 407–439.

[5] S.P. Dirkse and M.C. Ferris: *MCPLIB: A collection of nonlinear mixed complementarity problems.* Optimization Methods and Software 5, 1995, pp. 123–156.

[6] F. Facchinei, A. Fischer and C. Kanzow: *Inexact Newton methods for semismooth equations with applications to variational inequality problems.* In: G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Applications.* Plenum Press, New York, NY, 1996, pp. 125–139.

[7] F. Facchinei, A. Fischer and C. Kanzow: *A semismooth Newton method for variational inequalities: The case of box constraints.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems. State of the Art.* SIAM, Philadelphia, PA, 76–90.

[8] F. Facchinei and J. Soares: *A new merit function for nonlinear complementarity problems and a related algorithm.* SIAM Journal on Optimization 7, 1997, pp. 225–247.

[9] M.C. Ferris: Private communication, September 1996.

[10] M.C. Ferris and C. Kanzow: *Recent developments in the solution of nonlinear complementarity problems.* Preprint, in preparation.

[11] M.C. Ferris and J.-S. Pang: *Engineering and economic applications of complementarity problems.* SIAM Review, to appear.

[12] M.C. Ferris and T.F. Rutherford: *Accessing realistic mixed complementarity problems within MATLAB.* In: G. Di Pillo and F. Giannessi (eds.): *Nonlinear Optimization and Applications.* Plenum Press, New York, NY, pp. 141–153.

[13] A. Fischer: *A special Newton-type optimization method.* Optimization 24, 1992, pp. 269–284.

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

[15] A. Fischer and C. Kanzow: *On finite termination of an iterative method for linear complementarity problems.* Mathematical Programming 74, 1996, pp. 279–292.

[16] S.A. Gabriel and J.J. Moré: *Smoothing of mixed complementarity problems.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems. State of the Art.* SIAM, Philadelphia, PA, 105–116.

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

[18] C. Kanzow: *Semismooth Newton-type Methods for the Solution of Nonlinear Complementarity Problems.* Habilitation Thesis, Institute of Applied Mathematics, University of Hamburg, Hamburg, April 1997.

[19] C. Kanzow and M. Fukushima: *Equivalence of the generalized complementarity problem to differentiable unconstrained minimization.* Journal of Optimization Theory and Applications 90, 1996, pp. 581–603.

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

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

[22] 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, pp. 45–68.

[23] J.-S. Pang: *Newton's method for B-differentiable equations.* Mathematics of Operations Research 15, 1990, pp. 311–341.

[24] J.-S. Pang: *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems.* Mathematical Programming 51, 1991, pp. 101–131.

[25] J.-S. Pang and L. Qi: *Nonsmooth equations: Motivation and algorithms.* SIAM Journal on Optimization 3, 1993, pp. 443–465.

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

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

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

[29] Ph.L. Toint: *Non-monotone trust-region algorithms for nonlinear optimization subject to convex constraints.* Mathematical Programming 77, 1997, pp. 69–94.

[30] P. Tseng: *Growth behaviour of a class of merit functions for the nonlinear complementarity problem.* Journal of Optimization Theory and Applications 89, 1996, pp. 17–37.

[31] N. Yamashita and M. Fukushima: *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems.* Mathematical Programming 76, 1997, pp. 469–491.