ON A SEMISMOOTH LEAST SQUARES FORMULATION OF COMPLEMENTARITY PROBLEMS WITH GAP REDUCTION

Christian Kanzow and Stefania Petra

Preprint 251 July 2003

University of Würzburg
Institute of Applied Mathematics and Statistics
Am Hubland
97074 Würzburg
Germany

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

July 23, 2003

Abstract: We present a nonsmooth least squares reformulation of the complementarity problem and investigate its convergence properties. The global and local fast convergence results (under mild assumptions) are similar to some existing equation-based methods. In fact, our least squares formulation is obtained by modifying one of these equation-based methods (using the Fischer-Burmeister function) in such a way that we overcome a major drawback of this equation-based method. The resulting nonsmooth Levenberg-Marquardt-type method turns out to be significantly more robust than the corresponding equation-based method. This is illustrated by our numerical results using the MCPLIB test problem collection.

Keywords: Complementarity problems, nonlinear least squares reformulation, semismooth functions, global convergence, quadratic convergence.

1 Introduction

Given a continuously differentiable mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, the complementarity problem consists of finding a solution $x^* \in \mathbb{R}^n$ satisfying the following system of equations and inequalities:

$$x_i \geq 0, \;\; F_i(x) \geq 0, \;\; x_i F_i(x) = 0 \quad \forall i = 1, \ldots, n. \tag{1}$$

This problem plays an important role in the formulation of several economic equilibrium problems. In addition, there are a number of applications within the engineering field. For more details, the interested reader is referred to the survey paper [11] by Ferris and Pang or to the recent books [7, 8] by Facchinei and Pang.

Many algorithms for the solution of the complementarity problem are based on a suitable reformulation as a nonlinear and nonsmooth system of equations. The most popular one exploits the *Fischer-Burmeister function*

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

introduced in [12]. The Fischer-Burmeister function belongs to the class of *NCP-functions* ϕ which are defined by the property

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

Hence the complementarity problem is equivalent to the system of equations $\Phi_{FB}(x) = 0$ with $\Phi_{FB} : \mathbb{R}^n \to \mathbb{R}^n$ being given by

$$\Phi_{FB}(x) := \begin{pmatrix} \phi_{FB}(x_1, F_1(x)) \\ \vdots \\ \phi_{FB}(x_n, F_n(x)) \end{pmatrix}.$$

A number of important methods for the solution of the complementarity problem are based on this reformulation, see once again the book [8] and the references therein.

In order to motivate our approach, it will be convenient to divide the complementarity problem (1) into the following two subproblems:

- obtaining feasibility in the sense that $x_i \geq 0$ and $F_i(x) \geq 0$ for all $i = 1, \ldots, n$, and
- reducing the complementarity gap $x^T F(x) = \sum_{i=1}^n x_i F_i(x)$ down to zero.

Now the Fischer-Burmeister approach is quite effective in solving the feasibility problem. To this end, just note that $|\phi_{FB}(a, b)|$ becomes relatively large if either a or b is significantly negative. On the other hand, the Fischer-Burmeister function is much less effective in reducing the complementarity gap $x_i F_i(x)$ since ϕ_{FB} is quite flat on the positive orthant. For example, if a is a large number and b is, say, equal to one, then the product ab is a large number, whereas $|\phi_{FB}(a, b)|$ is close to zero.

In this paper, we try to overcome this problem by using a least squares formulation of the complementarity problem. To this end, let us introduce the function

$$\phi_+(a, b) := a_+ b_+ \;,$$

where $z_+ := \max\{0, z\}$ for $z \in \mathbb{R}$. We then define the mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^{2n}$ by

$$\Phi(x) := \begin{pmatrix} \vdots \\ \lambda\phi_{FB}(x_i, F_i(x)), & i = 1, \ldots, n \\ \vdots \\ (1-\lambda)\phi_+(x_i, F_i(x)), & i = 1, \ldots, n \\ \vdots \end{pmatrix}, \tag{2}$$

where $\lambda \in (0, 1)$ is a fixed but arbitrary parameter used as a weight between the first term and the second one. Hence we obtain Φ by adding some components to the definition of Φ_{FB}. These additional components are used in order to avoid the disadvantage of the Fischer-Burmeister function mentioned before.

A similar idea has been used before in [2] where the penalized Fischer-Burmeister function was introduced which, however, is based on a direct modification of ϕ_{FB}. Moreover, the corresponding Newton-type method in [2] (essentially taken from [5]) might have to take a number of gradient steps for difficult problems, whereas this is completely avoided in our approach.

To describe this approach, first note that $\Phi(x) = 0$ is an overdetermined system of equations having the property that

$$x^* \text{ solves } \Phi(x) = 0 \Longleftrightarrow x^* \text{ solves (1)}.$$

Hence we have a nonlinear least squares formulation of the complementarity problem with the additional property that the residual at the solution is zero. We therefore suggest a nonsmooth Gauss-Newton- or Levenberg-Marquardt-type method for the solution of this least squares problem. This method has the following advantages over existing methods based on Φ_{FB}:

- Faster reduction of the complementarity gap $x^T F(x)$.
- A Newton-type step can be accepted at each iteration; in particular, no gradient steps are necessary in order to get global convergence.
- The numerical results indicate that the method is more robust than those methods which are based on Φ_{FB}.

The organization of our paper is as follows: In Section 2, we state some important properties of the mapping Φ and the corresponding merit function

$$\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2. \tag{3}$$

The algorithm and its global and local convergence properties are given in Section 3. We then note in Section 4 that our approach can be extended to mixed complementarity problems, and we present our numerical results for the whole MCPLIB test problem collection. We conclude with some final remarks in Section 5.

A few words about our notation. Given a function $G : \mathbb{R}^n \to \mathbb{R}^m$ we denote by G_i its ith component function. If G is continuously differentiable we denote its Jacobian at a point $x \in \mathbb{R}^n$ by $G'(x)$. We say that the mapping G is an LC^1 function if G' is locally Lipschitzian. If the mapping G is locally Lipschitzian and if D_G is its set of differentiable points of G, we use

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

to denote Clarke's generalized Jacobian of G at x, see [3]. If G is a real valued mapping, the generalized Jacobian reduces to the generalized gradient of G at x. Furthermore we write

$$\partial_C G(x)^T = \partial G_1(x) \times \cdots \times \partial G_m(x)$$

for the C-subdifferential of G at x, where the right-hand side denotes a set of matrices whose ith column can be any element from the generalized gradient of $\partial G_i(x)$. Finally, if $M \in \mathbb{R}^{n\times n}$ is a matrix with elements m_{ij} and $I, J \subseteq \{1, \ldots, n\}$ are two given subsets, we write M_{IJ} for the $|I| \times |J|$ submatrix of M consisting of the elements m_{ij} with $i \in I$ and $j \in J$.

2 Properties of Φ and Ψ

In this section, we study several important properties of the mappings Φ and Ψ from (2) and (3), respectively. To this end, we begin with the equation operator Φ and note that it is (strongly) semismooth. For the definition of (strong) semismoothness, the reader is referred to the original papers [26, 25] or to the recent book [8].

Theorem 2.1 *The mapping Φ from (2) is semismooth. If F is an LC^1 function, then Φ is strongly semismooth.*

Theorem 2.1 is a direct consequence of the facts that ϕ_{FB} and ϕ_+ are strongly semismooth functions and that the composition of (strongly) semismooth functions is again (strongly) semismooth, see [13, 8].

We next investigate the structure of the C-subdifferential of Φ at a given point $x \in \mathbb{R}^n$. To this end, we first state a standard result regarding the generalized gradients of ϕ_{FB} and ϕ_+, cf. [2].

Lemma 2.2 *The generalized gradient of the function $\phi_{FB} : \mathbb{R}^2 \to \mathbb{R}$ at a point $(a, b) \in \mathbb{R}^2$ is equal to the set of all $(g_a, g_b)^T \in \mathbb{R}^2$ with*

$$(g_a, g_b) = \begin{cases} \left(\frac{a}{\|(a,b)\|} - 1, \frac{b}{\|(a,b)\|} - 1\right), & \text{if } (a, b) \neq (0, 0), \\ (\xi - 1, \zeta - 1), & \text{if } (a, b) = (0, 0), \end{cases}$$

and where (ξ, ζ) is any vector satisfying $\|(\xi, \zeta)\| \leq 1$. The generalized gradient of the function $\phi_+ : \mathbb{R}^2 \to \mathbb{R}$ at a point $(a, b)^T \in \mathbb{R}^2$ is equal to $\partial\phi_+(a, b) = \{(b_+\partial a_+, a_+\partial b_+)\}$,

where

$$\partial z_+ = \begin{cases} 1 & \text{if } z > 0, \\ [0,1] & \text{if } z = 0, \\ 0 & \text{if } z < 0. \end{cases}$$

As a consequence of Lemma 2.2, we obtain the following result.

Theorem 2.3 *Let $x \in \mathbb{R}^n$ be given. Then any matrix $H \in \partial_C \Phi(x)$ has the representation*

$$H = \begin{pmatrix} \lambda H_1 \\ (1-\lambda) H_2 \end{pmatrix},$$

where

$$H_1 \subseteq D_a(x) + D_b(x) F'(x) \quad \text{and} \quad H_2 \subseteq \tilde{D}_a(x) + \tilde{D}_b(x) F'(x)$$

with

$$D_a(x) = \mathrm{diag}\{a_i(x)\}, \;\; D_b(x) = \mathrm{diag}\{b_i(x)\}, \;\; \tilde{D}_a(x) = \mathrm{diag}\{\tilde{a}_i(x)\}, \;\; \tilde{D}_b(x) = \mathrm{diag}\{\tilde{b}_i(x)\}$$

being diagonal matices with entries $(a_i(x), b_i(x)) \in \partial \phi_{FB}(x_i, F_i(x))$ and $(\tilde{a}_i(x), \tilde{b}_i(x)) \in \partial \phi_+(x_i, F_i(x))$, where $\partial \phi_{FB}(x_i, F_i(x))$ and $\partial \phi_+(x_i, F_i(x))$ denote the sets from Lemma 2.2, with (a, b) being replaced by $(x_i, F_i(x))$.

Proof. By our definition of the C-subdifferential, we have

$$\partial_C \Phi(x)^T = \partial \Phi_1(x) \times \cdots \times \partial \Phi_{2n}(x),$$

where $\partial \Phi_i(x)$ denotes the generalized gradient of the i-th component function of Φ. Using Lemma 2.2, it follows that

$$\partial \Phi_i(x) \subseteq \lambda \big(a_i(x) e_i^T + b_i(x) \nabla F_i(x)^T \big) \quad \forall i \in \{1, \ldots, n\} \tag{4}$$

and

$$\partial \Phi_i(x) \subseteq (1-\lambda) \big(\tilde{a}_i(x) e_i^T + \tilde{b}_i(x) \nabla F_i(x)^T \big) \quad \forall i \in \{n+1, \ldots, 2n\}, \tag{5}$$

with $(a_i(x), b_i(x))$ and $(\tilde{a}_i(x), \tilde{b}_i(x))$ being the elements specified in the statement of our theorem. □

In order to prove fast local convergence, we need to show that every element $H \in \partial_C \Phi(x^*)$ has full rank n under a suitable assumption. This assumption will be the R-regularity condition. To this end, we define the index sets

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

and recall that a solution x^* of the complementarity problem is called *R-regular* if the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular and the Schur complement

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

is a P-matrix (see [27, 4]). Then we have the following result.

Theorem 2.4 *Let $x^* \in \mathbb{R}^n$ be an R-regular solution of the complementarity problem. Then all elements from the C-subdifferential $\partial\Phi_C(x^*)$ have full rank.*

Proof. Let $H \in \partial_C\Phi(x^*)$. In view of Theorem 2.3, we then have

$$H = \begin{pmatrix} \lambda H_1 \\ (1-\lambda)H_2 \end{pmatrix},$$

where H_1 is an element from $\partial_C\Phi_{FB}(x^*)$. It now follows from [9, 5] that each element $H_1 \in \partial_C\phi_{FB}(x^*)$ is nonsingular under the assumed R-regularity condition. Therefore we have $\operatorname{rank}(H) = n$, i.e., H has full rank. □

We next state a consequence of Theorem 2.4 that will play an important role in our convergence analysis.

Lemma 2.5 *Let $x^* \in \mathbb{R}^n$ be an R-regular solution of the complementarity problem. Then there exist constants $\varepsilon > 0$ and $c > 0$ such that*

$$\|(H^TH)^{-1}\| \le c$$

for all $H \in \partial_C\Phi(x)$ and all $x \in \mathbb{R}^n$ with $\|x - x^*\| \le \varepsilon$.

Proof. The proof is similar to one given in [25] in a slightly different situation. If the claim is not true, there exists a sequence $\{x^k\}$ converging to x^* and a corresponding sequence of matrices $\{H_k\}$ with $H_k \in \partial_C\Phi(x^k)$ for all $k \in \mathbb{N}$ such that either $H_k^TH_k$ is singular or $\|(H_k^TH_k)^{-1}\| \to \infty$ on a subsequence. Noting that $H_k^TH_k$ is symmetric positive semidefinite, we have $\|(H_k^TH_k)^{-1}\| = \frac{1}{\lambda_{\min}(H_k^TH_k)}$ in the nonsingular case. Hence the condition $\|(H_k^TH_k)^{-1}\| \to \infty$ is equivalent to $\lambda_{\min}(H_k^TH_k) \to 0$. Since $\{x^k\} \to x^*$ and the mapping $x \mapsto \partial_C\Phi(x)$ is upper semicontinuous, it follows that the sequence $\{H_k\}$ is bounded and therefore has a convergent subsequence. Let H_* be a limiting element of such a subsequence. It then follows that either $H_*^TH_*$ is singular or $\lambda_{\min}(H_*^TH_*) = 0$ (note that the mapping $A \mapsto \lambda_{\min}(A^TA)$ is continuous), i.e., H_* is not of full rank. On the other hand, exploiting the fact that the mapping $x \mapsto \partial_C\Phi(x)$ is closed, we have $H_* \in \partial_C\Phi(x^*)$, so that H_* is of full rank by Theorem 2.4. This contradiction completes the proof. □

We next investigate the properties of the merit function Ψ from (3). To this end, it will be useful to rewrite this function as

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

with $\psi : \mathbb{R}^2 \to \mathbb{R}$ being defined by

$$\psi(a,b) := \frac{1}{2}\lambda^2\phi_{FB}^2(a,b) + \frac{1}{2}(1-\lambda)^2 a_+^2 b_+^2. \tag{6}$$

The following properties of ψ are crucial in order to state several interesting results for the corresponding merit function Ψ. Basically, the next result says that ψ shares all the nice properties of the merit function corresponding to the Fischer-Burmeister function ϕ_{FB}.

Lemma 2.6 *The mapping $\psi : \mathbb{R}^2 \to \mathbb{R}$ from (6) has the following properties:*

(a) ψ is continuously differentiable on $\mathbb{R}^2$.

(b) $\psi(a,b) \geq 0$ for all $a, b \in \mathbb{R}^2$.

(c) $\psi(a,b) = 0 \Longleftrightarrow a \geq 0, b \geq 0$ and $ab = 0$.

(d) $\frac{\partial\psi}{\partial a}(a,b)\frac{\partial\psi}{\partial b}(a,b) \geq 0$ for all $a, b \in \mathbb{R}^2$.

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

Proof. Statements (a) and (b) follow directly from the definition of ψ together with the fact that ϕ_{FB}^2 is known to be continuously differentiable on $\mathbb{R}^2$, see [14, 9]. Property (c) follows from the fact that ϕ_{FB} is an NCP-function. Hence it remains to show statements (d) and (e). Since both statements obviously hold for $(a,b) = (0,0)$, we can assume without loss of generality that $(a,b) \neq (0,0)$ for the rest of this proof.

In order to verify part (d), first note that we have

$$\begin{aligned}
&\frac{\partial\psi}{\partial a}(a,b)\frac{\partial\psi}{\partial b}(a,b) \\
&\quad = \lambda^4\phi_{FB}^2(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big)\Big(\frac{b}{\sqrt{a^2+b^2}} - 1\Big) + (1-\lambda)^4 a_+^3 b_+^3 + \lambda^2(1-\lambda)^2 t(a,b)
\end{aligned}$$

with $t : \mathbb{R}^2 \to \mathbb{R}$ being defined by

$$t(a,b) := \phi_{FB}(a,b)a_+b_+\Big[\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big)a_+ + \Big(\frac{b}{\sqrt{a^2+b^2}} - 1\Big)b_+\Big].$$

It is easy to see that it suffices to prove that $t(a,b) \geq 0$. Now, we obviously have $t(a,b) \geq 0$ if $a \geq 0$ and $b \geq 0$. On the other hand, in all other cases, we have $t(a,b) = 0$, so that statement (d) follows.

To prove part (e), we first recall that we have $(a,b) \neq (0,0)$. Furthermore, taking into account the fact that an unconstrained minimum of a continuously differentiable function is always a stationary point of this function and using the symmetry of the function ψ with respect to its arguments a and b, we only have to verify the implication

$$\frac{\partial\psi}{\partial a}(a,b) = 0 \Longrightarrow \psi(a,b) = 0.$$

To this end, we first note that

$$\frac{\partial\psi}{\partial a}(a,b) = \lambda^2\phi_{FB}(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big) + (1-\lambda)^2 a_+ b_+^2.$$

Using $\frac{\partial\psi}{\partial a}(a,b) = 0$, let us consider two cases: If $a \le 0$ or $b \le 0$, we have $a_+ b_+^2 = 0$ and therefore

$$0 = \frac{\partial\psi}{\partial a}(a,b) = \lambda^2\phi_{FB}(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big).$$

This implies

$$\phi_{FB}(a,b) = 0 \quad \text{or} \quad \frac{a}{\sqrt{a^2+b^2}} - 1 = 0$$

which, in turn, is equivalent to

$$\phi_{FB}(a,b) = 0 \quad \text{or} \quad \big(a > 0 \text{ and } b = 0\big).$$

Hence we immediately have $\psi(a,b) = 0$.

Now consider the second case where $a > 0$ and $b > 0$. Then we get $\phi_{FB}(a,b) \le 0$ and therefore

$$\phi_{FB}(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big) \ge 0.$$

Consequently, we obtain from

$$0 = \frac{\partial\psi}{\partial a}(a,b) = \lambda^2\phi_{FB}(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big) + (1-\lambda)^2 a_+ b_+^2$$

that both sums must be equal to zero. In particular, we therefore have

$$0 = \lambda^2\phi_{FB}(a,b)\Big(\frac{a}{\sqrt{a^2+b^2}} - 1\Big).$$

Hence we can argue as in the first case and see that $\psi(a,b) = 0$. □

Using Lemma 2.6, we obtain the following properties of the merit function Ψ in essentially the same way as they can be obtained for some other merit functions, see [5, 9, 14] as well as [19] regarding the compact level sets for monotone problems. We therefore do not state the corresponding proofs here. As for the definition of a P_0-matrix occuring in statement (b), the reader is referred to [4]. For the notion of a (uniform) P-function, see [22]. Furthermore, recall that the complementarity problem (1) is said to be *strictly feasible* if there is a vector $\hat{x} \in \mathbb{R}^n$ such that $\hat{x} > 0$ and $F(\hat{x}) > 0$.

Theorem 2.7 *The merit function Ψ from (3) has the following properties:*

(a) Ψ is continuously differentiable with $\nabla\Psi(x) = H^T\Phi(x)$, where $H \in \partial_C\Phi(x)$ can be chosen arbitrarily.

(b) *If x^* is a stationary point of Ψ and $F'(x^*)$ is a P_0-matrix, then x^* is a solution of the complementarity problem.*

(c) *If F is either a uniform P-function or if F is monotone and the complementarity problem (1) is strictly feasible, then the level sets*

$$\mathcal{L}(c) := \big\{x \in \mathbb{R}^n \,\big|\, \Psi(x) \le c\big\}$$

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

We close this section by noting that there are some other merit functions which share the properties from Theorem 2.7, see [19, 2]. However, we are not aware of any merit function having stronger properties, while there are a couple of merit functions (including the Fischer-Burmeister merit function) which satisfy only some weaker conditions, see [21, 20, 14].

3 Algorithm and Convergence

We first state our algorithm for the solution of the complementarity problem (1). It is a Levenberg-Marquardt-type method applied to the nonlinear least squares problem

$$\min \Psi(x) = \frac{1}{2}\|\Phi(x)\|^2,$$

where, of course, Φ and Ψ denote the mappings from (2) and (3), respectively.

Algorithm 3.1 (Semismooth Levenberg-Marquardt Method)

(S.0) Let $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$ and $\varepsilon \ge 0$. Choose any $x^0 \in \mathbb{R}^n$. Set $k := 0$.

(S.1) If $\|\nabla\Psi(x^k)\| \le \varepsilon$: STOP.

(S.2) Choose $H_k \in \partial_C\Phi(x^k)$, $\lambda_k > 0$ and find a solution $d^k \in \mathbb{R}^n$ of

$$\big(H_k^T H_k + \lambda_k I\big)d = -\nabla\Psi(x^k). \tag{7}$$

(S.3) Compute $t_k = \max\{\beta^l \,|\, l = 0, 1, 2, \dots\}$ such that

$$\Psi(x^k + t_k d^k) \le \Psi(x^k) + \sigma t_k \nabla\Psi(x^k)^T d^k. \tag{8}$$

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

We now investigate the convergence properties of our algorithm. To this end, we assume that the termination parameter ε is equal to zero and that Algorithm 3.1 generates an infinite sequence. We further note that Algorithm 3.1 is well defined since $\lambda_k > 0$ and since one can easily see that the search direction d^k is always a descent direction for the merit function Ψ.

We first state a global convergence result. For the sake of simplicity, we assume that λ_k is given by

$$\lambda_k := \|\nabla\Psi(x^k)\|, \tag{9}$$

although several other choices of λ_k yield the same result including the more realistic choice

$$\lambda_k := \min\{c_1, c_2\|\nabla\Psi(x^k)\|\}$$

for certain constants $c_1, c_2 > 0$. Note that these choices are consistent with the requirements for local superlinear/quadratic convergence in Theorem 3.3 below.

Theorem 3.2 *Algorithm 3.1 is well defined for an arbitrary complementarity problem. Furthermore, every accumulation point of a sequence $\{x^k\}$ generated by Algorithm 3.1 is a stationary point of Ψ.*

Proof. As noticed above the algorithm is well defined. Let x^* be any accumulation point of the sequence $\{x^k\}$ and $\{x^k\}_K$ a subsequence converging to x^*. Suppose that $\nabla\Psi(x^*) \neq 0$. Due to the monotone decrease of the sequence $\{\Psi(x^k)\}$ and the fact that $\{\Psi(x^k)\}_K$ converges to $\Psi(x^*)$, it follows that the entire sequence $\{\Psi(x^k)\}$ converges to $\Psi(x^*)$. In particular, we therefore have

$$\Psi(x^{k+1}) - \Psi(x^k) \to 0.$$

On the other hand, we obtain

$$\Psi(x^{k+1}) - \Psi(x^k) \leq \sigma t_k \nabla\Psi(x^k)^T d^k \leq 0$$

by Step (S.3) in Algorithm 3.1 and the descent property of the search direction d^k. Hence we have

$$\{t_k \nabla\Psi(x^k)^T d^k\}_K \to 0. \tag{10}$$

Using the definition of the Levenberg-Marquardt direction gives

$$t_k \nabla\Psi(x^k)^T d^k = -t_k \nabla\Psi(x^k)^T (H_k^T H_k + \lambda_k I)^{-1} \nabla\Psi(x^k). \tag{11}$$

Since $\{x^k\}_K \to x^*$, we get from the upper semicontinuity of the C-subdifferential that the sequence $\{H_k\}_K$ is bounded. Without loss of generality, we therefore have $\{H_k\}_K \to H_*$ for some matrix $H_* \in \partial_C\Phi(x^*)$. Since $\nabla\Psi$ is continuous, we also obtain $\{\nabla\Psi(x^k)\}_K \to \nabla\Psi(x^*)$ and therefore $\{\lambda_k\}_K \to \lambda_*$ with $\lambda_* := \|\nabla\Psi(x^*)\| > 0$, cf. (9). Using these arguments, it follows that the matrices

$$H_k^T H_k + \lambda_k I$$

converge to the symmetric positive definite matrix $H_*^T H_* + \lambda_* I$ on the subset $K \subseteq \mathbb{N}$. From (10), (11) we therefore obtain

$$\{t_k\}_K \to 0.$$

Now, for each $k \in \mathbb{N}$, let $l_k \in \mathbb{N}$ be the uniquely defined exponent such that $t_k = \beta^{l_k}$. It follows that the Armijo-rule in Step (S.3) is not satisfied for β^{l_k-1} for sufficiently large $k \in K$. Hence, we have

$$\frac{\Psi(x^k + \beta^{l_k-1}d^k) - \Psi(x^k)}{\beta^{l_k-1}} > \sigma \nabla\Psi(x^k)^T d^k \tag{12}$$

for all these $k \in K$. From the Levenberg-Marquardt equation we obtain $\{d^k\}_K \to d^*$, with d^* being the solution of the linear system

$$\big(H_*^T H_* + \lambda_* I\big)d = -\nabla\Psi(x^*).$$

Taking into account that $\{d^k\}_K \to d^*, \{x^k\}_K \to x^*$ and $\{t_k\}_K \to 0$, we obtain from (12) that

$$\nabla\Psi(x^*)^T d^* \geq \sigma \nabla\Psi(x^*)^T d^*.$$

Hence $\nabla\Psi(x^*)^T d^* \geq 0$, since $\sigma \in (0, \frac{1}{2})$. On the other hand, we have

$$\nabla\Psi(x^*)^T d^* = -\nabla\Psi(x^*)^T (H_*^T H_* + \lambda_* I)^{-1} \nabla\Psi(x^*) < 0.$$

This contradiction shows that x^* is a stationary point of Ψ. □

Recall that Theorem 2.7 (b) gives a relatively mild condition for a stationary point of Ψ to be a solution of the complementarity problem (1). Furthermore, the existence of a stationary point follows, e.g., under the assumptions of Theorem 2.7 (c).

We next investigate the rate of convergence of Algorithm 3.1. Obviously, this rate of convergence depends on the choice of the Levenberg-Marquardt parameter λ_k.

Theorem 3.3 *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1. Assume that x^* is an accumulation point of $\{x^k\}$ such that x^* is an R-regular solution of the complementarity problem* (1)*. Then the following statements hold:*

(a) The entire sequence $\{x^k\}$ converges to x^* if $\{\lambda_k\}$ is bounded.

(b) The full stepsize $t_k = 1$ is always accepted for k sufficiently large so that $x^{k+1} = x^k + d^k$ provided that $\lambda_k \to 0$.

(c) The rate of convergence is Q-superlinear if $\lambda_k \to 0$.

(d) The rate of convergence is Q-quadratic if $\lambda_k = O(\|\nabla\Psi(x^k)\|)$ and, in addition, F is an LC^1-function.

Proof. (a) To establish that the entire sequence $\{x^k\}$ converges to x^*, we first note that an R-regular solution is an isolated solution of the complementarity problem, see [27]. Since Algorithm 3.1 generates a decreasing sequence $\{\Psi(x^k)\}$ and x^* is a solution of the complementarity problem, it follows that the entire sequnce $\{\Psi(x^k)\}$ converges to zero.

Hence every accumulation point of the sequence $\{x^k\}$ is a solution of (1). Consequently, x^* is an isolated accumulation point of the sequence $\{x^k\}$.

Now let $\{x^k\}_K$ denote any subsequence converging to x^*, and note that x^* is a stationary point of Ψ. For all $k \in \mathbb{N}$, we have

$$\|x^{k+1} - x^k\| = t_k \|d^k\| \leq \|d^k\| \leq \|(H_k^T H_k + \lambda_k I)^{-1}\| \, \|\nabla \Psi(x)\|.$$

From $\{\nabla \Psi(x^k)\}_K \to 0$, Lemma 2.5 and the assumed boundedness of $\{\lambda_k\}$, we immediately obtain $\{\|x^{k+1} - x^k\|\}_K \to 0$. Hence statement (a) follows from [23, Lemma 4.10].

(b), (c) First we prove that

$$\|x^k + d^k - x^*\| = o(\|x^k - x^*\|) \tag{13}$$

for all $k \in \mathbb{N}$ sufficiently large. In view of part (a), we know that the entire sequence $\{x^k\}$ converges to the R-regular solution x^*. Hence it follows from Lemma 2.5 that there is a constant $c > 0$ such that

$$\|(H_k^T H_k + \lambda_k I)^{-1}\| \leq c \quad \forall k \in \mathbb{N}.$$

Furthermore, the sequence $\{H_k\}$ is bounded. We can therefore assume without loss of generality that we also have

$$\|H_k^T\| \leq c \quad \forall k \in \mathbb{N}.$$

Using Theorem 2.7 (a) and $\Phi(x^*) = 0$, we then obtain for all $k \in \mathbb{N}$ sufficiently large that

$$\begin{aligned}
\|x^k + d^k - x^*\| &= \|x^k - (H_k^T H_k + \lambda_k I)^{-1} \nabla \Psi(x^k) - x^*\| \\
&\leq \|(H_k^T H_k + \lambda_k I)^{-1}\| \, \|\nabla \Psi(x^k) - (H_k^T H_k + \lambda_k I)(x^k - x^*)\| \\
&\leq c \, \|H_k^T \Phi(x^k) - H_k^T H_k (x^k - x^*) - \lambda_k (x^k - x^*)\| \\
&= c \, \|H_k^T \big(\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\big) - \lambda_k (x^k - x^*)\| \\
&\leq c \, \big(\|H_k^T\| \, \|(\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| + \lambda_k \|x^k - x^*\|\big) \\
&\leq c\big(c\|\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| + \lambda_k \|x^k - x^*\|\big).
\end{aligned}$$

Since Φ is semismooth by Theorem 2.1, it follows that

$$\|\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| = o(\|x^k - x^*\|)$$

see [26, 25, 8]. Using the fact that $\lambda_k \to 0$ by assumption, we therefore obtain (13).

In order to prove that the full step is eventually accepted, we next show that

$$\lim_{k \to \infty} \frac{\Psi(x^k + d^k)}{\Psi(x^k)} = 0 \tag{14}$$

and

$$1 + \sigma \frac{\nabla \Psi(x^k)^T d^k}{\Psi(x^k)} \geq 1 - 2\sigma > 0 \tag{15}$$

holds for all sufficiently large $k \in \mathbb{N}$. Since $\Psi(x^k) \neq 0$ for all $k \in \mathbb{N}$, we get from Theorem 2.7 (a) that

$$\begin{aligned}\frac{\nabla\Psi(x^k)^T d^k}{\Psi(x^k)} &= -\frac{(H_k^T\Phi(x^k))^T(H_k^T H_k + \lambda_k I)^{-1} H_k^T \Phi(x_k)}{\frac{1}{2}\|\Phi(x^k)\|} \\ &\geq -\frac{\Phi(x^k)^T H_k (H_k^T H_k)^{-1} H_k^T \Phi(x_k)}{\frac{1}{2}\|\Phi(x^k)\|} \qquad (16) \\ &\geq -\frac{\Phi(x^k)^T \Phi(x_k)}{\frac{1}{2}\|\Phi(x^k)\|} \\ &= -2,\end{aligned}$$

where the second inequality in (16) follows from

$$d^T A d \leq \lambda_{\max}(A)\|d\|^2 \quad \forall d \in \mathbb{R}^n$$

and all symmetric matrices $A \in \mathbb{R}^{n\times n}$ by noting that the maximal eigenvalue of the symmetric matrix $A := H_k(H_k^T H_k)^{-1} H_k^T$ is equal to one. The inequality (15) now follows from (16).

To verify (14), we only have to show that

$$\lim_{k\to\infty} \frac{\|\Phi(x^k + d^k)\|}{\|\Phi(x^k)\|} = 0 \qquad (17)$$

holds. To this end, we first note that there exists a constant $\alpha > 0$ such that

$$\|\Phi(x^k)\| \geq \alpha\|x^k - x^*\| \qquad (18)$$

for all $k \in \mathbb{N}$ sufficiently large. This follows from the simple observation that $\|\Phi(x)\| \geq \lambda\|\Phi_{FB}(x)\|$ together with the fact that all elements $V \in \partial\Phi_{FB}(x^*)$ are nonsingular under the R-regularity condition as well as [24, Proposition 3]. Using (18) and (13), we obtain

$$\begin{aligned}\frac{\|\Phi(x^k + d^k)\|}{\|\Phi(x^k)\|} &\leq \frac{\|\Phi(x^k + d^k)\|}{\alpha\|x^k - x^*\|} \\ &= \frac{\|\Phi(x^k + d^k) - \Phi(x^*)\|}{\alpha\|x^k - x^*\|} \\ &\leq \frac{L\|x^k + d^k - x^*\|}{\alpha\|x^k - x^*\|} \\ &\to 0,\end{aligned}$$

where $L > 0$ denotes the local Lipschitz constant of Φ. Hence (17) holds.

Using (14) and (15), we see that the condition

$$\Psi(x^k + d^k) \leq \Psi(x^k) + \sigma\nabla\Psi(x^k)^T d^k$$

or, equivalently,

$$\frac{\Psi(x^k + d^k)}{\Psi(x^k)} \leq 1 + \sigma \frac{\nabla\Psi(x^k)^T d^k}{\Psi(x^k)}$$

is satisfied for all $k \in \mathbb{N}$ sufficiently large. Hence the stepsize $t_k = 1$ is eventually accepted in the line search criterion, and we have $x^{k+1} = x^k + d^k$ for all $k \in \mathbb{N}$ large enough. Hence Q-superlinear convergence of $\{x^k\}$ to x^* follows from (13).

(d) The proof is essentially the same as for the local superlinear convergence. To this end, we only note that F being an LC^1 function implies that Φ is stronlgy semismooth by Theorem 2.1, and that the relation

$$\|\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| = O(\|x^k - x^*\|^2).$$

holds for strongly semismooth functions, see [26, 25, 8]. □

Note that the previous proof is similar to one given in [16]. We stress, however, that [16] considers a Levenberg-Marquardt-type method for a square system of equations, whereas we are dealing with a nonsquare (overdetermined) system.

4 Extension to Mixed Complementarity Problems and Computational Results

4.1 Extension to Mixed Complementarity Problems

In this subsection, we would like to point out that the approach presented for the standard complementarity problem (1) can actually be extended to the more general mixed complementarity problem. We only sketch the idea here and do not state any formal results.

In order to introduce the mixed complementarity problem, it is quite convenient to consider the variational inequality problem first. For a given function $F : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty, closed and convex set $X \subseteq \mathbb{R}^n$, this variational inequality problem consists in finding a point $x^* \in X$ such that

$$F(x^*)^T(x - x^*) \geq 0 \quad \forall x \in X.$$

It is well-known and easy to see that this variational inequality problem is equivalent to the complementarity problem (1) when X is equal to the nonnegative orthant, i.e., if $X = [0, \infty)$. On the other hand, if $X = [l, u]$ is a general box with lower bounds $l = (l_1, \ldots, l_n)^T$ and upper bounds $u = (u_1, \ldots, u_n)^T$ satisfying $-\infty \leq l_i < u_i \leq +\infty$ for all $i \in \{1, \ldots, n\}$, we obtain the mixed complementarity problem.

In order to present a reformulation of this mixed complementarity problem, let us introduce the following partition of the index set $I := \{1, \ldots, n\}$:

$$\begin{aligned} I_l &:= \{i \in I \mid -\infty < l_i < u_i = \infty\}, \\ I_u &:= \{i \in I \mid -\infty = l_i < u_i < \infty\}, \\ I_{lu} &:= \{i \in I \mid -\infty < l_i < u_i < \infty\}, \\ I_f &:= \{i \in I \mid -\infty = l_i < u_i = \infty\}. \end{aligned}$$

We now define the operator $\bar{\Phi} : \mathbb{R}^n \to \mathbb{R}^{2n}$ componentwise as follows ($i = 1, \ldots, n$):

$$\bar{\Phi}_i(x) := \begin{cases} \lambda\phi_{FB}(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ -\lambda\phi_{FB}(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ \lambda\phi_{FB}(x_i - l_i, \phi_{FB}(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -\lambda F_i(x) & \text{if } i \in I_f, \end{cases}$$

$$\bar{\Phi}_{n+i}(x) := \begin{cases} (1-\lambda)\phi_+(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ (1-\lambda)\phi_+(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ (1-\lambda)(\phi_+(x_i - l_i, F_i(x)) + \phi_+(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -(1-\lambda)F_i(x) & \text{if } i \in I_f. \end{cases}$$

Note that the first n components of $\bar{\Phi}$ correspond to the Fischer-Burmeister-type reformulation of the mixed complementarity problem first introduced by Billups [1] and further investigated in [10]. The last n components are again used in order to reduce the complementarity gap at the current point x.

It follows in a relatively simple way from these references that the mixed complementarity problem is equivalent to the overdetermined system of equations $\bar{\Phi}(x) = 0$. Furthermore, one can show that the corresponding merit function

$$\bar{\Psi}(x) := \frac{1}{2}\|\bar{\Phi}(x)\|^2$$

is continuously differentiable. Algorithm 3.1 can therefore be extended in a straightforward way to the nonlinear least squares problem

$$\min \bar{\Psi}(x) = \frac{1}{2}\|\bar{\Phi}(x)\|^2, \quad x \in \mathbb{R}^n,$$

and it is rather straigthforward to see that this method has similar global and local convergence properties.

4.2 Numerical Results

We implemented Algorithm 3.1 (or, more precisely, the extension from the previous subsection) in MATLAB and tested the algorithm on the MCPLIB test problem collection, see

[6] (note that we use a newer version of this test problem collection). The implementation corresponds exactly to the statement of Algorithm 3.1 except that we use a nonmonotone line search as introduced by Grippo, Lampariello and Lucidi [15]. To be more precise, we use the standard (monotone) Armijo rule during the first five iterations and then switch to the nonmonotone line search where the maximum of the function values $\Psi(x^k)$ is taken over the last ten iterations, see [15] for further details.

We terminate the iteration if one of the following conditions are satisfied

$$\|\bar{\Phi}(x^k)\| \leq 10^{-11} \quad \text{or} \quad \|\nabla\bar{\Psi}(x^k)\| \leq 10^{-6} \quad \text{or} \quad k > 300,$$

and we choose $\lambda_k := 0$ for all $k \in \mathbb{N}$ so that our Levenberg-Marquardt method becomes a Gauss-Newton-type algorithm. The other parameters used in our implementation are $\lambda = 0.1, \beta = 0.55, \sigma = 10^{-4}$. The procedure for calculating an element $H_k \in \partial_C\Phi(x^k)$ is similar to one given in [5] for the Fischer-Burmeister equation operator.

Our numerical results are summarized in Table 1 for small dimensional problems and in Table 2 for large dimensional ones. In these tables the first column gives the name of the problem; Dim is the number of the variables in the problem; $\bar{\Psi}(x^0)$ gives the value of the merit function at the starting point; Nit denotes the number of iterations; $\bar{\Psi}(x^f)$ and $\|\nabla\bar{\Psi}(x^f)\|$ denote the values of $\bar{\Psi}(x)$ and $\|\nabla\bar{\Psi}(x)\|$ at the final iterate $x = x^f$. Note that Nit is equal to the number of linear subproblems solved.

Table 1: Numerical results for MCPLIB test problems

Problem	Dim	$\bar{\Psi}(x^0)$	Nit	$\bar{\Psi}(x^f)$	$\|\nabla\bar{\Psi}(x^f)\|$
badfree	5	4.600000e–01	2	3.802055e–13	1.120124e–06
bertsekas	15	3.936098e–03	38	2.754293e–16	1.468242e–07
billups	1	3.451182e–05	30	2.153258e–12	7.538314e–06
choi	13	7.709002e–03	5	2.649619e–16	1.278982e–09
colvdual	20	5.488000e+01	19	8.785822e–12	1.096167e–05
colvnlp	15	6.207596e+01	6	4.033072e–15	2.298057e–07
cycle	1	5.173835e+01	5	3.703547e–21	7.746281e–10
degen	2	1.000000e–01	5	6.295417e–17	1.122182e–09
duopoly	63	2.132546e+02	—	—	—
ehl-k40	41	1.042178e+04	32	2.335817e–14	6.724183e–06
ehl-k60	61	3.797546e+04	43	4.583751e–14	1.039908e–05
ehl-k80	81	9.363011e+04	50	1.115490e–12	3.133144e–03
ehl-kost	101	1.878951e+05	113	1.021911e–12	5.417297e–03
electric	158	2.609736e+08	33	8.195661e–13	2.421314e–06
explcp	16	3.200000e–01	19	5.723100e–16	3.383225e–09
forcebsm	184	3.944244e+03	239	3.095727e–12	2.489442e–07
forcedsa	186	3.948661e+03	25	2.971468e–16	2.437821e–09
freebert	15	1.509811e+04	10	6.212865e–14	2.194618e–06
gafni	5	1.300358e+03	10	6.470323e–13	3.651867e–05

Table 1: Numerical results for MCPLIB test problems (continued)

Problem	Dim	$\bar{\Psi}(x^0)$	Nit	$\bar{\Psi}(x^f)$	$\|\nabla\bar{\Psi}(x^f)\|$
games	16	6.006634e+01	6	7.154607e–12	8.895455e–05
hanskoop	14	1.185959e+02	—	—	—
hydroc06	29	1.766604e+05	5	6.763084e–15	6.398054e–04
hydroc20	99	4.104414e+05	9	6.306107e–16	1.027084e–04
jel	6	9.561221e+02	7	5.187223e–18	2.058253e–08
josephy	4	2.281054e–02	3	1.059211e–22	1.355775e–10
kojshin	4	2.281054e–02	3	1.079806e–22	1.368936e–10
mathinum	3	6.215376e+02	4	3.024771e–12	6.673325e–07
mathisum	4	5.216473e+00	8	2.199525e–16	1.559273e–08
methan08	31	6.463832e+06	4	2.488696e–14	4.527670e–03
nash	10	5.426293e+02	4	2.354630e–19	7.456865e–09
ne-hard	3	1.155892e+04	24	2.464179e–15	8.990290e–07
pgvon106	106	1.536653e+02	21	6.110093e–12	8.964999e–07
pies	42	5.267785e+08	27	7.026768e–13	8.659322e–03
powell	16	6.807131e–04	5	8.057886e–17	1.061372e–08
powell-mcp	8	9.316746e+01	2	2.728284e–13	6.048681e–06
qp	4	3.300000e+00	5	6.295416e–17	1.122089e–09
scarfanum	13	6.994871e–05	3	3.605079e–12	2.824943e–06
scarfasum	14	6.994871e–05	3	3.604873e–12	2.809442e–06
scarfbsum	40	1.123239e+02	19	3.920227e–13	1.655680e–04
shubik	45	1.638873e–01	—	—	—
simple-ex	17	9.561639e+00	—	—	—
simple-red	13	2.250785e+02	11	1.037775e–19	3.444415e–10
sppe	27	1.216934e+02	4	4.790941e–19	3.603064e–10
tinloi	146	4.001771e–01	9	1.416760e–12	1.286753e–03
tobin	42	3.236481e+00	2	1.474633e–14	1.354525e–05

Table 1 shows that the algorithm was able to solve almost all problems from the MCPLIB collection including a number of examples which are known to be very hard. We have failures only on problems `duopoly`, `hanskoop`, `shubik` and `simple-ex`. By changing the parameters we can solve the first two problems with high precision, but we still fail on `shubik`. Choosing $\beta = 0.9$ we also succeed to solve `simple-ex` in 49 steps, or even in 10, for another choice of λ. The `billups` problem was constructed by Billups [1] in order to make almost all methods fail on this problem. We consider it as an accident that we could solve it.

Table 2: Numerical results for MCPLIB test problems

Problem	Dim	$\bar{\Psi}(x^0)$	Nit	$\bar{\Psi}(x^f)$	$\|\nabla\bar{\Psi}(x^f)\|$
bert-oc	5000	5.129446e+01	10	6.254441e–12	3.734828e–07
bratu	5625	7.886962e+00	10	1.800397e–14	2.026905e–08
bishop	1645	2.157671e+11	—	—	—
lincont	419	3.956019e+03	21	5.335966e–18	9.678748e–07
obstacle	2500	3.371445e–04	7	3.232956e–12	1.645663e–06
opt-cont	288	8.693982e+02	8	3.973421e–14	3.040037e–08
opt-cont31	1024	2.435888e+02	9	1.175346e–12	1.578468e–07
opt-cont127	4096	7.867371e+01	12	3.232490e–12	2.584229e–07
opt-cont255	8192	5.184548e+01	41	4.567335e–12	3.173779e–07
opt-cont511	16384	3.848816e+01	17	4.924145e–13	1.000753e–07
trafelas	2904	5.124999e+03	143	6.567286e–17	1.585157e–09

In Table 2 we see that we are also able to solve all larger problems with the only exception of problem `bishop`. In particular, we can solve the relatively difficult examples `lincont` and `trafelas`.

We close this section with some interesting remarks:

- During all the iterations, no domain violations occured, i.e., the function F was always defined at the iterates x^k. This is very much in contrast to the pure Fischer-Burmeister approach where domain violations occur quite frequently and special (heuristic) rules have to be used so that the method can go on.

- If we apply our algorithm with $\lambda = 1$ in the definition of Φ or $\bar{\Phi}$, then our method reduces to the standard Fischer-Burmeister approach since the last n components in the definition of Φ or $\bar{\Phi}$ get cancelled. Doing so, we get failures on seven (compared to four) test problems among the small-dimensional examples, namely `duopoly`, `ehl-k40`, `electric`, `forcebsm`, `forcedsa`, `shubik`, `simple-ex`. Furthermore, we sometimes have to take a significantly higher number of iterations for some other examples, e.g., the solution of `ehl-k60`, `ehl-k80` and `ehl-kost` now takes more than 200 iterations.

- A further comparison with the pure Fischer-Burmeister approach can be obtained by having a look at the numerical results presented in [29] where four different Fischer-Burmeister-type algorithms are compared with each other (two of them use constrained reformulations of the complementarity problem and therefore have to solve more complicated subproblems). Even the best method in [29] produces more failures than our algorithm. This is interesting to note especially because many of the more difficult test problems from the MCPLIB collection (like `electric`, `forcebsm`, `forcedsa`) have been completely excluded from the numerical results in [29].

Altogether this indicates that our new approach is certainly more robust and sometimes also more efficient than the underlying Fischer-Burmeister method.

5 Conclusions

We have introduced a new method for the solution of (mixed and nonlinear) complementarity problems. This method uses a nonlinear least squares reformulation of the complementarity problem and applies a Levenberg-Marquardt-type method to this reformulated problem. The main idea of our method is to overcome one of the main disadvantages of the well-known Fischer-Burmeister method and to take special attention to the reduction of the complementarity gap. The numerical results indicate that the new method is significantly more robust than the corresponding Fischer-Burmeister equation-based algorithm.

While we have illustrated our technique by modifying the equation-based method using the Fischer-Burmeister function, it should be clear that our idea can also be used in order to modify other equation-based methods, see, for example, [28] for a summary of many of these equation reformulations. More precisely, assume we have a reformulation of the complementarity problem (1) as a square system of equations $\Phi_A(x) = 0$ with $\Phi_A : \mathbb{R}^n \to \mathbb{R}^n$. Suppose further that $\Phi_B : \mathbb{R}^n \to \mathbb{R}^m$ is any mapping with the property that $\Phi_B(x) = 0$ whenever x is a solution of (1). Then it is easy to see that x^* is a solution of the complementarity problem (1) if and only if x^* is a solution of the overdetermined system of equations $\Phi(x) = 0$, where $\Phi : \mathbb{R}^n \to \mathbb{R}^{n+m}$ is now defined by

$$\Phi(x) := \begin{pmatrix} \Phi_A(x) \\ \Phi_B(x) \end{pmatrix}.$$

Assuming that Φ and the corresponding merit function $\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2$ have similar properties as those stated in Section 2 for the functions from (2) and (3), we can apply the Levenberg-Marquardt method from Algorithm 3.1 to the least squares problem

$$\min \Psi(x)$$

in order to solve the complementarity problem (1). The convergence theory from Section 3 still holds for this approach. Of course, the crucial part is the definition of the mapping Φ_B which depends on the properties of the mapping Φ_A.

References

[1] S.C. Billups: *Algorithms for complementarity problems and generalized equations.* Ph.D. thesis, Computer Sciences Department, University of Wisconsin, Madison, 1995.

[2] B. Chen, X. Chen and C. Kanzow: *A penalized Fischer-Burmeister NCP-function.* Mathematical Programming 88, 2000, pp. 211–216.

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

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

[5] 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.

[6] S.P. Dirkse and M.C. Ferris: *MCPLIB: A collection of nonlinear mixed complementarity problems.* Optimization Methods and Software 5, 1995, pp. 319–345.

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

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

[9] 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.

[10] M.C. Ferris, C. Kanzow and T.S. Munson: *Feasible descendent algorithms for mixed complemetarity problems.* Mathematical Programming 86, 1999, pp. 475–497.

[11] M.C. Ferris and J.S. Pang: *Engineering and economic applications of complementarity problems.* SIAM Review 39, 1997, pp. 669–713.

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

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

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

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

[16] H. Jiang: *Global convergence analysis of the generalized Newton and Gauss-Newton methods of the Fischer-Burmeister equation for the complementarity problem.* Mathematics of Operations Research 24, 1999, pp. 529–543.

[17] C. Kanzow: *An inexact QP-based method for nonlinear complementarity problems.* Numerische Mathematik 80, 1998, pp. 557–577.

[18] C. Kanzow: *Strictly feasible equation-based methods for mixed complementarity problems.* Numerische Mathematik 89, 2001, pp. 135–160.

[19] C. Kanzow, N. Yamashita and M. Fukushima: *New NCP-functions and their properties.* Journal of Optimization Theory and Applications 94, 1997, pp. 115–135.

[20] Z.Q. Luo and P. Tseng: *A new class of merit functions for the nonlinear complementarity problem.* In M.C. Ferris and J.S. Pang (eds.): *Complementarity and Variational Problems: State of the Art.* SIAM, Philadelphia, PA, 1997, pp. 204–225.

[21] O.L. Mangasarian and M.V. Solodov: *Nonlinear complementarity as unconstrained and constrained optimization.* Mathematical Programming 62, 1993, pp. 277–298.

[22] 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. Moré and D.C. Sorensen: *Computing a trust region step.* SIAM Journal on Scientific and Statistical Computing 4, 1983, pp. 553–572.

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

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

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

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

[28] D. Sun and L. Qi: *On NCP-functions.* Computational Optimization and Applications 13, 1999, pp. 201–230.

[29] M. Ulbrich: *Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems.* SIAM Journal on Optimization 11, 2001, pp. 889–916.