JACOBIAN SMOOTHING METHODS FOR GENERAL NONLINEAR COMPLEMENTARITY PROBLEMS

Christian Kanzow [1,2] and Heiko Pieper [3]

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

[3] Stanford University
Department of Engineering-Economic
Systems and Operations Research
Terman Engineering Center
Stanford, CA 94305-4023
e-mail: pieper@stanford.edu

October 13, 1997

Abstract: We present a new algorithm for the solution of general (not necessarily monotone) complementarity problems. The algorithm is based on a reformulation of the complementarity problem as a nonsmooth system of equations by using the Fischer-Burmeister function. We use an idea by Chen, Qi and Sun and apply a Jacobian smoothing method (which is a mixture between nonsmooth Newton and smoothing methods) in order to solve this system. In contrast to Chen, Qi and Sun, however, our method can be applied to general complementarity problems. Extensive numerical results indicate that the new algorithm works very well. In particular, it can solve all complementarity problems from the MCPLIB and GAMSLIB libraries.

Key Words: Nonlinear complementarity problem, nonsmooth Newton method, smoothing method, global convergence, quadratic convergence.

---

[2]Current address (October 1, 1997 — September 30, 1998): Computer Sciences Department, University of Wisconsin – Madison, 1210 West Dayton Street, 53706 Madison, WI; e-mail: kanzow@cs.wisc.edu. The research of this author is being supported by the DFG (Deutsche Forschungsgemeinschaft).

1 Introduction

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. The *nonlinear complementarity problem* is to find a solution of the following system of equations and inequalities:

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

We denote this problem by NCP(F). It has a large number of important applications, and we refer the interested reader to the survey papers by Harker and Pang [22] and Ferris and Pang [17].

The basic idea of most algorithms for the solution of NCP(F) is to reformulate this problem as a nonlinear system of equations, as an optimization problem or as a parametric problem. Here we concentrate ourselves on the equation-based approach where problem NCP(F) is written equivalently as

$$\Phi(x) = 0 \tag{1}$$

for a suitable equation-operator $\Phi : \mathbb{R}^n \to \mathbb{R}^n$. For certain reasons, the operator Φ is usually nonsmooth, so that we cannot apply the classical Newton method in order to solve problem (1). Nevertheless, recent research shows that one can still design globally and locally fast convergent methods for the solution of (1). In the following, we give a short summary of the basic ideas of some of these methods which are related to this paper.

Nonsmooth Newton Methods: Instead of solving problem (1) by the classical Newton method, one can apply a nonsmooth Newton method based, e.g., on Clarke's [12] generalized Jacobian $\partial\Phi(x)$ of Φ at the point $x \in \mathbb{R}^n$. For example, the nonsmooth Newton methods by Kummer [30] and Qi and Sun [37] solve at each iteration the generalized Newton equation

$$V_k d = -\Phi(x^k), \tag{2}$$

where $V_k \in \partial\Phi(x^k)$. This method is locally superlinearly/quadratically convergent under certain assumptions, but (in contrast to the classical Newton method for smooth systems of equations) cannot be globalized in a simple way for general operators Φ. However, by using special functions Φ, several authors have recently presented globally and locally fast convergent nonsmooth Newton-type methods, see, e.g., [25, 16, 13, 28, 5].

One of the main advantages of most of these methods is the fact that they are usually well-defined for an arbitrary complementarity problem NCP(F).

Smoothing Methods: Another way to deal with the nonsmoothness of Φ is to approximate this function by a smooth operator $\Phi_\mu : \mathbb{R}^n \to \mathbb{R}^n$, where $\mu > 0$ denotes the smoothing parameter. The basic idea of the class of smoothing methods is then to solve a sequence of problems

$$\Phi_\mu(x) = 0 \tag{3}$$

and to force μ to go to 0. The advantage of this approach is that one can apply the standard Newton method for solving problem (3) so that one has to solve at each iteration the smoothing Newton equation

$$\Phi_\mu'(x^k) d = -\Phi_\mu(x^k). \tag{4}$$

Smoothing methods of this kind were considered, e.g., by Chen and Harker [6], Chen and Mangasarian [9], Kanzow [26], Gabriel and Moré [20], Burke and Xu [3, 43], Xu [41, 42], Hotta and Yoshise [23], Chen and Ye [11], Chen and Chen [4], Jiang [24] and Tseng [40]. In particular, the paper [3] by Burke and Xu initiated much of the recent research in this area.

The disadvantage of smoothing methods is that they usually require F to be at least a P_0-function in order to guarantee that the linear systems (4) are solvable. It seems difficult to make smoothing methods work on general complementarity problems where the Jacobian in (4) might be singular. This problem is also reflected by the fact that smoothing methods try to follow the so-called smoothing path which may not exist for non-P_0- or non-monotone problems.

Nevertheless, a sophisticated implementation like in the SMOOTH code by Chen and Mangasarian [9] seems to work quite well also for non-monotone problems, see [2].

Jacobian Smoothing Methods: The third class of algorithms for the solution of (1) is due to Chen, Qi and Sun [10]. They call it a smoothing Newton method, but we prefer the name Jacobian smoothing method in order to distinguish it better from the class of smoothing methods. These methods try to solve at each iteration the mixed Newton equation

$$\Phi'_\mu(x^k)d = -\Phi(x^k). \tag{5}$$

This linear system is a mixture between the nonsmooth Newton equation (2) and the smoothing Newton equation (4); it uses the unperturbed right-hand side from (2), but the smooth matrix from (4).

The algorithm and convergence theory developed by Chen et al. [10] still relies on the fact that the linear systems (5) are solvable at each iteration, and, similarly to the class of smoothing methods, this assumption is intimately related to F being a P_0-function. Hence also this Jacobian smoothing method is not well-defined for general complementarity problems.

Note that the Jacobian smoothing idea is also used in a couple of recent smoothing papers as a kind of hybrid step, see, e.g., [11, 4]. The main reason for doing this is that the Jacobian smoothing method helps (or simplifies) to prove local fast convergence.

Despite the fact that Jacobian smoothing methods are often viewed as a variation of smoothing methods, we take a different point of view: We view a Jacobian smoothing method as a suitable perturbation of a nonsmooth Newton method. In fact, the Jacobian smoothing method seems to be much closer to nonsmooth Newton methods than to smoothing methods since they do not try to follow any smoothing path. Instead, they also try to solve the unperturbed problem (1) directly by replacing the matrix $V_k \in \partial\Phi(x^k)$ in (2) by a suitable approximation $\Phi'_\mu(x^k)$.

Having this in mind, it seems reasonable to ask if one can modify the Jacobian smoothing method by Chen et al. [10] in such a way that it becomes well-defined for general complementarity problems. This is actually the main motivation for this paper, and the answer is positive.

In order to do this, however, we cannot consider the general class of smoothing methods used by Chen et al. [10]. Instead, we concentrate on one particular reformulation of the complementarity problem NCP(F) and fully exploit the (additional) properties of this special reformulation. It is based on the Fischer-Burmeister function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ defined by

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

see [18]. Then it is well-known and easy to see that problem NCP(F) is equivalent to problem (1) with Φ being defined by

$$\Phi(x) := \begin{pmatrix} \varphi(x_1, F_1(x)) \\ \vdots \\ \varphi(x_n, F_n(x)) \end{pmatrix}.$$

The globalization strategy for our algorithm is heavily based on the natural merit function $\Psi : \mathbb{R}^n \to \mathbb{R}$ given by

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

The corresponding smooth operator $\Phi_\mu : \mathbb{R}^n \to \mathbb{R}^n$ is defined similarly by

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

where $\varphi_\mu : \mathbb{R}^2 \to \mathbb{R}$ denotes Kanzow's [26] smooth approximation

$$\varphi_\mu(a,b) := \sqrt{a^2+b^2+2\mu} - a - b, \quad \mu > 0,$$

of the Fischer-Burmeister function.

The organization of this paper is as follows: The mathematical background and some preliminary results are summarized in Section 2. The Jacobian smoothing idea is discussed in more detail in Section 3. The algorithm together with some of its elementary properties is presented in Section 4. The global and local convergence analysis is part of Sections 5 and 6, respectively. Extensive and very encouraging numerical results are reported in Section 7, and Section 8 concludes this paper with some final remarks.

Some words about our notation. Let $G : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. Then $G'(x) \in \mathbb{R}^{m\times n}$ denotes the Jacobian of G at a point $x \in \mathbb{R}^n$, whereas the symbol $\nabla G(x)$ is used for the transposed Jacobian. In particular, if $m = 1$, the gradient $\nabla G(x)$ is viewed as a column vector. If $G : \mathbb{R}^n \to \mathbb{R}^m$ is only locally Lipschitzian, we can define Clarke's [12] *generalized Jacobian* as follows:

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

here, D_G denotes the set of differentiable points of G and conv$\mathcal{A}$ is the convex hull of a set $\mathcal{A}$. If $m = 1$, we call $\partial G(x)$ the *generalized gradient* of G at x for obvious reasons.

Usually, $\partial G(x)$ is difficult to compute, especially for $m > 1$. Instead, Proposition 2.6.2 (e) in Clarke [12] provides the overestimation

$$\partial G(x)^T \subseteq \partial G_1(x) \times \ldots \times \partial G_m(x),$$

where the right-hand side denotes the set of matrices in $\mathbb{R}^{n\times m}$ whose ith column is given by the generalized gradient of the ith component function G_i. Since this right-hand side is often easier to compute and motivated by the recent paper [34] by Qi, we write

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

and call $\partial_C G(x)$ the *C-subdifferential* of G at x. For the purpose of this paper, the C-subdifferential is considerably more important than the more familiar generalized Jacobian.

If $x \in \mathbb{R}^n$, we denote by $\|x\|$ the Euclidian norm of x. Similarly, $\|A\|$ denotes the spectral norm of a matrix $A \in \mathbb{R}^{n\times n}$ which is the induced matrix norm of the Euclidian vector norm. Occasionally, we will also write $\|\cdot\|_2$ in order to avoid any possible confusions. Sometimes we also need the Frobenius norm $\|A\|_F$ of a matrix $A \in \mathbb{R}^{n\times n}$.

If $A \in \mathbb{R}^{n\times n}$ is any given matrix and $\mathcal{A} \subseteq \mathbb{R}^{n\times n}$ is a nonempty set of matrices, we denote by $\mathrm{dist}(A, \mathcal{A}) := \inf_{B\in\mathcal{A}} \|A - B\|$ the distance between A and $\mathcal{A}$. This is sometimes also written as $\mathrm{dist}_2(A, \mathcal{A})$ in order to emphasize that the distance is measured using the spectral norm. Similarly, we write $\mathrm{dist}_F(A, \mathcal{A})$ if the distance is calculated by using the Frobenius norm. The (Euclidian) distance between a vector and a set of vectors of the same dimension is defined in an analogous way.

Finally, we make use of the Landau symbols $o(\cdot)$ and $O(\cdot)$: Let $\{\alpha_k\}$ and $\{\beta_k\}$ be two sequences of positive numbers such that $\beta_k \to 0$. Then we write $\alpha_k = o(\beta_k)$ if $\alpha_k/\beta_k \to 0$, and $\alpha_k = O(\beta_k)$ if $\limsup_{k\to\infty} \alpha_k/\beta_k < \infty$, i.e., if there exists a constant $c > 0$ such that $\alpha_k \leq c\beta_k$ for all $k \in \mathbb{N} := \{0, 1, 2, \ldots\}$.

2 Preliminaries

In this section, we summarize some of the known properties of the functions Φ, Φ_μ and Ψ which will be important for our subsequent analysis. In addition, we prove some preliminary results which will also be used later.

The first result follows directly from the definition of the C-subdifferential and Proposition 3.1 in [16].

Proposition 2.1. *For an arbitrary $x \in \mathbb{R}^n$, we have*

$$\partial_C \Phi(x)^T = D_a(x) + \nabla F(x) D_b(x) \tag{6}$$

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

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

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

$$a_i(x) = \xi_i - 1, \qquad b_i(x) = \rho_i - 1$$

for every $(\xi_i, \rho_i) \in \mathbb{R}^2$ such that $\|(\xi_i, \rho_i)\| \leq 1$ if $(x_i, F_i(x)) = (0, 0)$.

The next result follows from [16, 19] together with known results for (strongly) semismooth functions [37] and the recent theory of C-differentiable functions by Qi [34].

Proposition 2.2. *Assume that $\{x^k\} \subseteq \mathbb{R}^n$ is any convergent sequence with limit point $x^* \in \mathbb{R}^n$. Then the following statements hold:*

(a) The function Φ is semismooth so that

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

for any $H_k \in \partial_C \Phi(x^k)$.

(b) If F is continuously differentiable with a locally Lipschitzian Jacobian, then Φ is strongly semismooth so that

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

for any $H_k \in \partial_C \Phi(x^k)$.

The following result can be verified similarly to Lemma 3.7 in [27].

Proposition 2.3. *The function φ_μ satisfies the inequality*

$$|\varphi_{\mu_1}(a,b) - \varphi_{\mu_2}(a,b)| \leq \sqrt{2}|\sqrt{\mu_1} - \sqrt{\mu_2}|$$

for all $(a,b) \in \mathbb{R}^2$ and all $\mu_1, \mu_2 \geq 0$. In particular, we have

$$|\varphi_\mu(a,b) - \varphi(a,b)| \leq \sqrt{2}\sqrt{\mu}$$

for all $(a,b) \in \mathbb{R}^2$ and all $\mu > 0$.

As an immediate consequence of Proposition 2.3, we obtain

Corollary 2.4. *The function Φ_μ satisfies the inequality*

$$\|\Phi_{\mu_1}(x) - \Phi_{\mu_2}(x)\| \leq \kappa \, |\sqrt{\mu_1} - \sqrt{\mu_2}\,| \tag{7}$$

for all $x \in \mathbb{R}^n$ and $\mu_1, \mu_2 \geq 0$, where $\kappa := \sqrt{2n}$. In particular, we have

$$\|\Phi_\mu(x) - \Phi(x)\| \leq \kappa \, \sqrt{\mu}$$

for all $x \in \mathbb{R}^n$ and all $\mu \geq 0$.

We next state a result which is a minor extension of Proposition 3.4 of [16]. We omit its proof here since it can be carried out in a similar way as the one in [16].

Proposition 2.5. *The merit function Ψ is continuously differentiable with $\nabla\Psi(x) = V^T\Phi(x)$ for an arbitrary $V \in \partial_C \Phi(x)$.*

The following technical result will be used in the proof of our main global convergence result, Theorem 5.8 below.

Lemma 2.6. *Let* $\{x^k\} \subseteq \mathbb{R}^n$ *and* $\{\mu_k\} \subseteq \mathbb{R}$ *be two sequences with* $\{x^k\} \to x^*$ *for some* $x^* \in \mathbb{R}^n$ *and* $\{\mu_k\} \downarrow 0$. *Then*

$$\lim_{k\to\infty} \nabla \Psi_{\mu_k}(x^k) = \nabla \Psi(x^*)$$

and

$$\lim_{k\to\infty} \Phi'_{\mu_k}(x^k)^T \Phi(x^k) = \nabla \Psi(x^*).$$

Proof. Since Ψ_μ is differentiable for all $\mu > 0$, we have

$$\nabla \Psi_{\mu_k}(x^k) = \Phi'_{\mu_k}(x^k)^T \Phi_{\mu_k}(x^k) = \sum_{i\in I} \varphi_{\mu_k}(x_i^k, F_i(x^k)) \nabla \Phi_{\mu_k,i}(x^k),$$

where $\Phi_{\mu_k,i}$ denotes the ith component function of Φ_{μ_k}. On the other hand, for arbitrary $V \in \partial_C \Phi(x^*)$, we obtain from Proposition 2.5:

$$\nabla \Psi(x^*) = V^T \Phi(x^*) = \sum_{i\in I} \varphi(x_i^*, F_i(x^*)) V_i^T,$$

where V_i^T denotes the ith column of the matrix V^T. Now let

$$\beta(x^*) := \{i \,|\, x_i^* = F_i(x^*) = 0\}.$$

We consider two cases:

Case 1: $i \notin \beta(x^*)$.
Then the Fischer-Burmeister function is continuously differentiable at $(x_i^*, F_i(x^*))$, and the ith column of V^T is single valued and equal to $\nabla \Phi_i(x^*)$ (cf. Proposition 2.1). In particular, all limits exist, and from the continuity of φ and ∇F, we obtain:

$$\lim_{k\to\infty} \varphi_{\mu_k}(x_i^k, F_i(x^k)) \nabla \Phi_{\mu_k,i}(x^k) = \varphi(x_i^*, F_i(x^*)) \nabla \Phi_i(x^*) = \varphi(x_i^*, F_i(x^*)) V_i^T.$$

Case 2: $i \in \beta(x^*)$.
Since

$$\frac{\partial \varphi_\mu}{\partial a}(a,b) \in (-2,0) \quad \text{and} \quad \frac{\partial \varphi_\mu}{\partial a}(a,b) \in (-2,0)$$

for all $(a,b) \in \mathbb{R}^2$ and $\mu > 0$, the sequence $\{\nabla \Phi_{\mu_k,i}(x^k)\}$ is bounded for $k \to \infty$. Since

$$\lim_{k\to\infty} \varphi_{\mu_k}(x_i^k, F_i(x^k)) = \varphi(x_i^*, F_i(x^*)) = 0,$$

we therefore have

$$\lim_{k\to\infty} \varphi_{\mu_k}(x_i^k, F_i(x^k)) \nabla \Phi_{\mu_k,i}(x^k) = 0.$$

Since we also have $\varphi(x_i^*, F_i(x^*)) V_i^T = 0$ for all $i \in \beta(x^*)$, the first statement follows from Cases 1 and 2.

The second statement is easier to establish than the first one since we multiply by $\Phi(x^k)$ and not by $\Phi_{\mu_k}(x^k)$. The proof would be similar to the one just given. $\square$

We conclude this section by stating another technical result which will also be utilized in our global convergence analysis.

Lemma 2.7. *Let* $\{x^k\}, \{d^k\} \subseteq \mathbb{R}^n$ *and* $\{t_k\} \subseteq \mathbb{R}$ *be sequences with* $x^{k+1} := x^k + t_k d^k$ *such that* $\{x^k\} \to x^*$, $\{d^k\} \to d^*$ *and* $\{t_k\} \downarrow 0$ *for certain vectors* $x^*, d^* \in \mathbb{R}^n$. *Furthermore let* $\{\mu_k\} \subseteq \mathbb{R}$ *be a sequence with* $\{\mu_k\} \downarrow 0$. *Then*

$$\lim_{k\to\infty} \frac{\Psi_{\mu_k}(x^k + t_k d^k) - \Psi_{\mu_k}(x^k)}{t_k} = \nabla\Psi(x^*)^T d^*.$$

Proof. From Proposition 2.5 and the Mean Value Theorem, we obtain that, for each $k \in \mathbb{N}$, there exists a vector $\xi^k \in \mathbb{R}^n$ on the line segment between x^k and x^{k+1} (that is $\xi^k = x^k + \theta_k d^k$ for some $\theta_k \in [0, t_k]$) such that

$$\Psi_{\mu_k}(x^k + t_k d^k) - \Psi_{\mu_k}(x^k) = t_k \nabla\Psi_{\mu_k}(\xi^k)^T d^k.$$

Dividing by t_k gives

$$\frac{\Psi_{\mu_k}(x^k + t_k d^k) - \Psi_{\mu_k}(x^k)}{t_k} = \nabla\Psi_{\mu_k}(\xi^k)^T d^k.$$

Since ξ^k lies between x^k and x^{k+1}, it follows that $\{\xi^k\} \to x^*$. Therefore, we can apply the first statement of Lemma 2.6, so that passing to the limit, we get

$$\lim_{k\to\infty} \frac{\Psi_{\mu_k}(x^k + t_k d^k) - \Psi_{\mu_k}(x^k)}{t_k} = \lim_{k\to\infty} \nabla\Psi_{\mu_k}(\xi^k)^T d^k = \nabla\Psi(x^*)^T d^*.$$

This completes the proof. □

3 Jacobian Smoothing

The basic idea of our algorithm to be presented in Section 4 is to replace the generalized Newton equation

$$V_k d = -\Phi(x^k), \quad V_k \in \partial_C \Phi(x^k),$$

by the linear system

$$\Phi'_{\mu_k}(x^k) d = -\Phi(x^k),$$

i.e., we replace the element V_k from the C-subdifferential $\partial_C\Phi(x^k)$ by the (existing) Jacobian $\Phi'_{\mu_k}(x^k)$ of the smoothed operator Φ_{μ_k}. In order to guarantee local fast convergence of this iteration, we have to control the difference between $\Phi'_{\mu_k}(x^k)$ and the set $\partial_C\Phi(x^k)$. A first result in this direction is established in

Lemma 3.1. *Let* $x \in \mathbb{R}^n$ *be arbitrary but fixed. Then we have*

$$\lim_{\mu\downarrow 0} \operatorname{dist}(\Phi'_\mu(x), \partial_C\Phi(x)) = 0. \tag{8}$$

Proof. From the definition of Φ_μ, we have for all $\mu > 0$,

$$\Phi_\mu'(x) = \operatorname{diag}\left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - 1 \right) + \operatorname{diag}\left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - 1 \right) F'(x).$$

We consider the distance of the columns of the transposed Jacobians.

To this end, let us define

$$\beta(x) := \{ i \mid x_i = F_i(x) = 0 \}.$$

If we denote the ith component function of Φ_μ by $\Phi_{\mu,i}$, we obtain

$$\lim_{\mu \downarrow 0} \nabla \Phi_{\mu,i}(x) = \begin{cases} \left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2}} - 1 \right) e_i + \left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2}} - 1 \right) \nabla F_i(x) & \text{for } i \notin \beta(x), \\ -e_i - \nabla F_i(x) & \text{for } i \in \beta(x). \end{cases}$$

Hence the assertion follows from Proposition 2.1 (with $(\xi_i, \rho_i) = (0,0)$ for $i \in \beta(x)$). □

It is an immediate consequence of Lemma 3.1 that we can find, for every fixed $\delta > 0$, a parameter $\bar{\mu} = \bar{\mu}(x, \delta) > 0$ such that

$$\operatorname{dist}(\Phi_\mu'(x), \partial_C \Phi(x)) \leq \delta$$

for all $0 < \mu \leq \bar{\mu}$. However, it does not follow from Lemma 3.1 how we can choose this threshold value $\bar{\mu}$. On the other hand, it is important for the design of our algorithm to have an explicit expression of a possible value of $\bar{\mu}$. This is made more precise in Proposition 3.4 below whose proof is based on the following two observations.

Lemma 3.2. *Let $x \in \mathbb{R}^n$ and $\mu > 0$ be arbitrary but fixed. Then*

$$\left[\operatorname{dist}_F \left(\nabla \Phi_\mu(x), \partial_C \Phi(x)^T \right) \right]^2 = \sum_{i=1}^n \left[\operatorname{dist}_2 \left(\nabla \Phi_{\mu,i}(x), \partial \Phi_i(x) \right) \right]^2.$$

Proof. Let V_i be the ith column of a matrix V. Then, using the definition of the C-subdifferential, it is easy to see that

$$\inf_{V \in \partial_C \Phi(x)^T} \sum_{i=1}^n \| \nabla \Phi_{\mu,i}(x) - V_i \|_2^2 = \sum_{i=1}^n \inf_{H_i \in \partial \Phi_i(x)} \| \nabla \Phi_{\mu,i}(x) - H_i \|_2^2.$$

Using this and the definition of the Frobenius norm, we obtain

$$\begin{aligned} \left[\operatorname{dist}_F \left(\nabla \Phi_\mu(x), \partial_C \Phi(x)^T \right) \right]^2 &= \inf_{V \in \partial_C \Phi(x)^T} \| \nabla \Phi_\mu(x) - V \|_F^2 \\ &= \inf_{V \in \partial_C \Phi(x)^T} \sum_{i=1}^n \| \nabla \Phi_{\mu,i}(x) - V_i \|_2^2 \\ &= \sum_{i=1}^n \inf_{H_i \in \partial \Phi_i(x)} \| \nabla \Phi_{\mu,i}(x) - H_i \|_2^2 \\ &= \sum_{i=1}^n \left[\operatorname{dist}_2 \left(\nabla \Phi_{\mu,i}(x), \partial \Phi_i(x) \right) \right]^2. \end{aligned}$$

This completes the proof. $\square$

Lemma 3.3. *Let $\mu > 0$ be arbitrary but fixed. Then the function $f : (0,\infty) \to \mathbb{R}$, defined by*

$$f(\tau) := \frac{1}{\sqrt{\tau}} - \frac{1}{\sqrt{\tau + 2\mu}},$$

is strictly decreasing in $\tau > 0$.

Proof. The function f is continuously differentiable with

$$f'(\tau) = -\frac{1}{2}\frac{1}{(\sqrt{\tau})^3} + \frac{1}{2}\frac{1}{\sqrt{\tau+2\mu}^3} = -\frac{1}{2}\left(\frac{1}{(\sqrt{\tau})^3} - \frac{1}{\sqrt{\tau+2\mu}^3}\right).$$

Hence we have $f'(\tau) < 0$ for all $\tau > 0$. This implies our assertion. $\square$

We now come to the main result of this section.

Proposition 3.4. *Let $x \in \mathbb{R}^n$ be arbitrary but fixed. Assume that x is not a solution of* NCP(F). *Let us define the constants*

$$\gamma(x) := \max_{i \notin \beta(x)} \{\|x_i e_i + F_i(x)\nabla F_i(x)\|\} \geq 0$$

and

$$\alpha(x) := \min_{i \notin \beta(x)} \{x_i^2 + F_i(x)^2\} > 0,$$

where $\beta(x) := \{i |\, x_i = F_i(x) = 0\}$. Let $\delta > 0$ be given, and define

$$\bar{\mu}(x,\delta) := \begin{cases} 1 & \text{if } \left(\frac{n\gamma(x)^2}{\delta^2} - \alpha(x)\right) \leq 0, \\ \frac{\alpha(x)^2}{2}\left(\frac{\delta^2}{n\gamma(x)^2 - \delta^2\alpha(x)}\right) & \text{otherwise.} \end{cases}$$

Then

$$dist_F(\Phi_\mu'(x), \partial_C\Phi(x)) \leq \delta$$

for all μ such that $0 < \mu \leq \bar{\mu}(x,\delta)$.

Proof. We first note that $\{1,\dots,n\} \setminus \beta(x) \neq \emptyset$ since x is not a solution of NCP(F) by assumption. Hence $\alpha(x) > 0$. Furthermore, since $\|A\|_F = \|A^T\|_F$ for an arbitrary matrix $A \in \mathbb{R}^{n\times n}$, we obtain

$$\begin{aligned} \operatorname{dist}_F\left(\Phi_\mu'(x), \partial_C\Phi(x)\right) &= \operatorname{dist}_F\left(\nabla\Phi_\mu(x), \partial_C\Phi(x)^T\right) \\ &= \sqrt{\textstyle\sum_{i=1}^n \left[\operatorname{dist}_2\left(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)\right)\right]^2}. \end{aligned} \tag{9}$$

from Lemma 3.2. Hence it is sufficient to consider the distance between the ith columns of $\nabla\Phi_\mu(x)$ and $\partial_C\Phi(x)^T$. To this end, we recall that these columns are given by

$$\nabla\Phi_{\mu,i}(x) = \frac{\partial\varphi_\mu}{\partial a}(x_i, F_i(x))e_i + \frac{\partial\varphi_\mu}{\partial b}(x_i, F_i(x))\nabla F_i(x)$$

and

$$\partial \Phi_i(x) = \begin{cases} \frac{\partial \varphi}{\partial a}(x_i, F_i(x))e_i + \frac{\partial \varphi}{\partial b}(x_i, F_i(x))\nabla F_i(x) & \text{if } i \notin \beta(x), \\ (\xi_i - 1)e_i + (\rho_i - 1)\nabla F_i(x) & \text{if } i \in \beta(x), \end{cases}$$

respectively, where $(\xi_i, \rho_i) \in \mathbb{R}^2$ denotes any vector such that $\|(\xi_i, \rho_i)\| \leq 1$, see Proposition 2.1. We distinguish two cases:

Case 1: $i \in \beta(x)$:
Then $(x_i, F_i(x)) = (0, 0)$ and therefore

$$\nabla \Phi_{\mu,i}(x) = -e_i - \nabla F_i(x).$$

Hence, taking $(\xi_i, \rho_i) = (0, 0)$, we see that

$$\nabla \Phi_{\mu,i}(x) \in \partial \Phi_i(x)$$

so that

$$\text{dist}_2\left(\nabla \Phi_{\mu,i}(x), \partial \Phi_i(x)\right) = 0 \tag{10}$$

for all $i \in \beta(x)$.

Case 2: $i \notin \beta(x)$:
In this case, we have

$$\partial \Phi_i(x) = \{\nabla \Phi_i(x)\}.$$

By a simple calculation, we therefore get

$$\begin{aligned}
& \text{dist}_2\left(\nabla \Phi_{\mu,i}(x), \partial \Phi_i(x)\right) \\
= \; & \|\nabla \Phi_{\mu,i}(x) - \nabla \Phi_i(x)\| \\
= \; & \left\| \left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - 1 \right) e_i + \left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - 1 \right) \nabla F_i(x) \right. \\
& \left. - \left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2}} - 1 \right) e_i - \left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2}} - 1 \right) \nabla F_i(x) \right\| \\
= \; & \left\| x_i e_i \left(\frac{1}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - \frac{1}{\sqrt{x_i^2 + F_i(x)^2}} \right) \right. \\
& \left. + F_i(x) \nabla F_i(x) \left(\frac{1}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - \frac{1}{\sqrt{x_i^2 + F_i(x)^2}} \right) \right\| \\
= \; & \left\| \left(\frac{1}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} - \frac{1}{\sqrt{x_i^2 + F_i(x)^2}} \right) (x_i e_i + F_i(x) \nabla F_i(x)) \right\| \\
= \; & \left(\frac{1}{\sqrt{x_i^2 + F_i(x)^2}} - \frac{1}{\sqrt{x_i^2 + F_i(x)^2 + 2\mu}} \right) \|x_i e_i + F_i(x) \nabla F_i(x)\|.
\end{aligned}$$

In view of the definitions of the constants $\alpha(x)$ and $\gamma(x)$, we therefore obtain by using Lemma 3.3:

$$\begin{aligned}\operatorname{dist}_2(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)) &\leq \left(\frac{1}{\sqrt{\alpha(x)}} - \frac{1}{\sqrt{\alpha(x)+2\mu}}\right)\gamma(x) \\ &= \left(\frac{\sqrt{\alpha(x)+2\mu} - \sqrt{\alpha(x)}}{\sqrt{\alpha(x)}\sqrt{\alpha(x)+2\mu}}\right)\gamma(x) \\ &\leq \left(\frac{\sqrt{2\mu}}{\sqrt{\alpha(x)}\sqrt{\alpha(x)+2\mu}}\right)\gamma(x),\end{aligned}$$

where the latter inequality follows from the elementary fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b \geq 0$. We now want to show that

$$\left(\frac{\sqrt{2\mu}}{\sqrt{\alpha(x)}\sqrt{\alpha(x)+2\mu}}\right)\gamma(x) \leq \frac{\delta}{\sqrt{n}} \tag{11}$$

for all $0 < \mu \leq \bar{\mu}(x, \delta)$ which then implies

$$\operatorname{dist}_2\left(\nabla\Phi_{\mu,i}(x), \partial\Phi_i(x)\right) \leq \frac{\delta}{\sqrt{n}}. \tag{12}$$

If $\gamma(x) = 0$, then inequality (11) holds trivially (for arbitrary $\mu > 0$). Hence we assume that $\gamma(x) > 0$. Then an easy calculation shows that (11) is equivalent to

$$\alpha(x)^2 \geq 2\mu\left(\frac{n\gamma(x)^2}{\delta^2} - \alpha(x)\right). \tag{13}$$

Hence, if $\frac{n\gamma(x)^2}{\delta^2} - \alpha(x) \leq 0$, inequality (11) is satisfied for any $\mu > 0$, in particular for all $\mu \in (0, 1]$. Otherwise we obtain the following upper bound from (13):

$$\mu \leq \frac{\alpha(x)^2}{2}\left(\frac{\delta^2}{n\gamma(x)^2 - \delta^2\alpha(x)}\right) =: \bar{\mu}(x, \delta).$$

Putting together (9), (10) and (12), we therefore obtain

$$\operatorname{dist}_F(\Phi'_\mu(x), \partial_C\Phi(x)) \leq \sqrt{\sum_{i=1}^n \frac{\delta^2}{n}} = \delta$$

for all $0 < \mu \leq \bar{\mu}(x, \delta)$. □

The constant $\bar{\mu}(x, \delta)$ defined in Proposition 3.4 will play a central role in the design of our algorithm to be described in the following section.

We also note that, since $\|A\| \leq \|A\|_F$ for an arbitrary matrix $A \in \mathbb{R}^{n\times n}$, it follows from Proposition 3.4 that

$$\operatorname{dist}(\Phi'_\mu(x), \partial_C\Phi(x)) \leq \delta$$

for all μ with $0 < \mu \leq \bar{\mu}(x, \delta)$.

4 Algorithm

In this section, we give a detailed description of our Jacobian smoothing method and state some of its elementary properties. In particular, we show that the algorithm is well-defined for an arbitrary complementarity problem.

Basically, we try to take the Jacobian smoothing method from Chen et al. [10]. In addition, we incorporate a gradient step in a similar (but slightly different) way as this is done by some nonsmooth Newton methods [13, 28, 5]. Unfortunately, the introduction of these gradient steps makes the updating rules for our smoothing parameter μ_k as well as the convergence theory considerably more technical and complicated. However, it is this gradient step which makes the algorithm applicable to a general nonlinear complementarity problem.

In fact, this is also the reason why we concentrate us on the Fischer-Burmeister function: Its merit function Ψ is smooth due to Proposition 2.5, whereas the same does not seem to hold for the general class of smoothing functions considered in [10].

We now state our algorithm formally.

Algorithm 4.1. *(Jacobian Smoothing Method)*

(S.0) Choose $x^0 \in \mathbb{R}^n$, $\lambda, \alpha, \eta, \rho \in (0,1)$, $\gamma > 0$, $\sigma \in (0, \frac{1}{2}(1-\alpha))$, $p > 2$ *and* $\epsilon \geq 0$. *Set* $\beta_0 := \|\Phi(x^0)\|$, $\kappa := \sqrt{2n}$, $\mu_0 := (\frac{\alpha}{2\kappa}\beta_0)^2$ *and* $k := 0$.

(S.1) If $\|\nabla\Psi(x^k)\| \leq \epsilon$: *STOP.*

(S.2) Find a solution $d^k \in \mathbb{R}^n$ *of the linear system*

$$\Phi'_{\mu_k}(x^k)d = -\Phi(x^k). \quad \textbf{(Newton step)} \tag{14}$$

If the system (14) *is not solvable or if the condition*

$$\Phi(x^k)^T\Phi'_{\mu_k}(x^k)d^k \leq -\rho\|d^k\|^p \tag{15}$$

is not satisfied, set

$$d^k := -\nabla\Psi(x^k). \quad \textbf{(Gradient step)} \tag{16}$$

(S.3) Find the smallest m_k *in* $\{0,1,2,\dots\}$ *such that*

$$\Psi_{\mu_k}(x^k + \lambda^{m_k}d^k) \leq \Psi_{\mu_k}(x^k) - 2\sigma\lambda^{m_k}\Psi(x^k) \tag{17}$$

if d^k *is given by* (14), *and such that*

$$\Psi(x^k + \lambda^{m_k}d^k) \leq \Psi(x^k) - \sigma\lambda^{m_k}\|d^k\|^2. \tag{18}$$

if d^k *is given by* (16). *Set* $t_k := \lambda^{m_k}$ *and* $x^{k+1} := x^k + t_kd^k$.

(S.4) If

$$\|\Phi(x^{k+1})\| \leq \max\left\{\eta\beta_k, \frac{1}{\alpha}\|\Phi(x^{k+1}) - \Phi_{\mu_k}(x^{k+1})\|\right\}, \tag{19}$$

then set

$$\beta_{k+1} := \|\Phi(x^{k+1})\|$$

and choose μ_{k+1} such that

$$0 < \mu_{k+1} \leq \min\left\{\left(\frac{\alpha}{2\kappa}\beta_{k+1}\right)^2, \frac{\mu_k}{4}, \bar{\mu}(x^{k+1}, \gamma\beta_{k+1})\right\}. \tag{20}$$

If (19) *is not satisfied and $d^k = -\nabla\Psi(x^k)$, then set*

$$\beta_{k+1} := \beta_k$$

and choose μ_{k+1} such that

$$0 < \mu_{k+1} \leq \min\left\{\left(\frac{\alpha}{2\kappa}\|\Phi(x^{k+1})\|\right)^2, \left(\frac{\|\Phi(x^k)\| - \|\Phi(x^{k+1})\|}{2\kappa}\right)^2, \frac{\mu_k}{4}\right\}. \tag{21}$$

If none of the above conditions is met, set $\beta_{k+1} := \beta_k$ and $\mu_{k+1} := \mu_k$.

(S.5) Set $k \leftarrow k+1$, and return to Step (S.1).

For convenience of presentation, we assume implicitly throughout the theoretical part of this paper that the termination parameter ϵ is equal to 0 and that the algorithm does not terminate after a finite number of iterations.

Before we start to investigate the properties of Algorithm 4.1, we give some comments on it: In Step (S.2), we try to solve the (mixed) Newton equation (14) which is the main computational effort of our method. If the solution of this linear system does not provide a direction of sufficient decrease (in the sense of (15)), we switch to the steepest descent direction of the merit function Ψ.

In Step (S.3), we perform a line search. The line search rule depends on the search direction chosen in Step (S.2): If d^k is the Newton direction, the line search in (17) is used as a globalization strategy. Note that this line search condition is exactly the same as in Chen et al. [10]. On the other hand, if d^k is a gradient step, we use the standard Armijo rule in (18).

The complicated part of the algorithm is in Step (S.4), where we update the parameter μ_k. The first part of the updating rules (where condition (19) is satisfied) is also used by Chen et al. [10]. The second part is due to the gradient step. In the following list, we give some more detailed comments on the role on these two updating rules:

(a) In both updating rules, namely in (20) and (21), we reduce μ_k at least by a factor of $1/4$. This is reasonable since we want to force μ_k to go to 0.

(b) The last part of the updating rule (20) controls the distance between our smooth Jacobian and the C-subdifferential, see Lemma 4.2 (b) below.

(c) The remaining parts of the updating rules (20) and (21) are important in order to guarantee that Algorithm 4.1 is well-defined and globally convergent. We will exploit these rules several times in our convergence proofs.

We now turn to the analysis of Algorithm 4.1. To this end, we introduce the index set

$$K = \{0\} \cup \left\{ k \in \mathbb{N} \;\middle|\; \|\Phi(x^k)\| \le \max\left\{ \eta\beta_{k-1}, \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\| \right\} \right\}. \qquad (22)$$

We stress that, compared to the updating rule (19), there is a shift of the indices in the definition of the index set K!

We can prove the following result.

Lemma 4.2. *The following two statements hold:*

(a) We have

$$\|\Phi(x^k) - \Phi_{\mu_k}(x^k)\| \le \alpha\|\Phi(x^k)\| \qquad (23)$$

for all $k \ge 0$.

(b) We have

$$\mathrm{dist}_F(\Phi'_{\mu_k}(x^k), \partial_C\Phi(x^k)) \le \gamma\|\Phi(x^k)\| \qquad (24)$$

for all $k \in K$ *with* $k \ge 1$.

Proof. (a) We distinguish three cases:
Case 1: $k \in K$:
Then we obtain from (20) and Corollary 2.4:

$$\|\Phi(x^k) - \Phi_{\mu_k}(x^k)\| \le \kappa\sqrt{\mu_k} \le \frac{\alpha}{2}\beta_k \le \alpha\beta_k = \alpha\|\Phi(x^k)\|.$$

Case 2: $k \notin K$ and the $(k-1)$st step is a Newton step (i.e., μ_k is not updated by (21)):
In this case, we have $\mu_k = \mu_{k-1}$, so that we obtain from (19):

$$\|\Phi(x^k) - \Phi_{\mu_k}(x^k)\| = \|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\| < \alpha\|\Phi(x^k)\|.$$

Case 3: $k \notin K$ and the $(k-1)$st step is a gradient step (i.e., μ_k is updated by (21)):
Then we obtain from Corollary 2.4 and (21):

$$\|\Phi(x^k) - \Phi_{\mu_k}(x^k)\| \le \kappa\sqrt{\mu_k} \le \frac{\alpha}{2}\|\Phi(x^k)\| \le \alpha\|\Phi(x^k)\|.$$

Statement (a) now follows from these three cases.

(b) Statement (b) follows immediately from the definition of the threshold value $\bar{\mu}(x,\delta)$ in Proposition 3.4 and the updating rule (20). $\square$

As a consequence of Lemma 4.2, we obtain

Theorem 4.3. *Algorithm 4.1 is well-defined.*

Proof. We only have to show that the exponent m_k in the line search rules (17)/(18) is finite for any $k \in \mathbb{N}$. In case of a gradient step, this is well-known since we use the standard Armijo-rule. In case of a Newton step, we can use Part (a) of Lemma 4.2 and prove the finiteness of m_k is essentially the same way as this was done in [10, Lemma 3.1]. □

5 Global Convergence

The aim of this section is to show that any accumulation point of a sequence generated by Algorithm 4.1 is at least a stationary point of Ψ. Unfortunately, the analysis is somewhat technical due to the different updating rules for Newton and gradient steps in Algorithm 4.1. We therefore need a couple of preliminary results. Some of them, however, are of interest by their own.

We begin our global convergence analysis with the following observation.

Lemma 5.1. *Let $\{x^k\} \subseteq \mathbb{R}^n$ be a sequence generated by Algorithm 4.1. Assume that $\{x^k\}$ has an accumulation point x^* which is a solution of* $\mathrm{NCP}(F)$. *Then the index set K is infinite and $\{\mu_k\} \to 0$.*

Proof. Assume that K is finite. Then it follows from (19) and the updating rules for β_k in Step (S.4) of Algorithm 4.1 that there is a $k_0 \in \mathbb{N}$ such that

$$\beta_k = \beta_{k_0}$$

and

$$\|\Phi(x^{k+1})\| > \max\left\{\eta\beta_k, \frac{1}{\alpha}\|\Phi(x^{k+1}) - \Phi_{\mu_k}(x^{k+1})\|\right\} \geq \eta\beta_k = \eta\beta_{k_0}$$

for all $k \in \mathbb{N}$ with $k \geq k_0$. However, this contradicts the fact that x^* is a solution of $\mathrm{NCP}(F)$ so that we have $\Phi(x^*) = 0$.

Hence K is an infinite set. The updating rules for μ_k therefore immediately imply that the whole sequence $\{\mu_k\}$ converges to 0. □

We will also need the following simple result.

Lemma 5.2. *The following two statements hold:*

(a) If d^k is given by (14), we have

$$\|\Phi_{\mu_k}(x^{k+1})\| < \|\Phi_{\mu_k}(x^k)\|.$$

(b) If $d^k = -\nabla\Psi(x^k)$ and if μ_k is updated by (21), then

$$\|\Phi_{\mu_{k+1}}(x^{k+1})\| \leq \|\Phi_{\mu_{k+1}}(x^k)\|.$$

(Note the difference between the index μ_k and μ_{k+1} in statements (a) and (b).)

Proof. Part (a) follows immediately from the line search rule (17).

(b) Let $d^k = -\nabla\Psi(x^k)$ and assume (19) is not satisfied. From (18), we have $\|\Phi(x^k)\| - \|\Phi(x^{k+1})\| =: c_k > 0$. Therefore, together with Corollary 2.4, we get

$$\begin{aligned}
\|\Phi_{\mu_{k+1}}(x^{k+1})\| &\le \|\Phi_{\mu_{k+1}}(x^{k+1}) - \Phi(x^{k+1})\| + \|\Phi(x^{k+1})\| \\
&\le \kappa\sqrt{\mu_{k+1}} + \|\Phi(x^k)\| - c_k \\
&\le \|\Phi_{\mu_{k+1}}(x^k)\| + \|\Phi(x^k) - \Phi_{\mu_{k+1}}(x^k)\| + \kappa\sqrt{\mu_{k+1}} - c_k \\
&\le \|\Phi_{\mu_{k+1}}(x^k)\| + 2\kappa\sqrt{\mu_{k+1}} - c_k \\
&\le \|\Phi_{\mu_{k+1}}(x^k)\|,
\end{aligned}$$

where the last inequality follows from the special choice of μ_{k+1} made in (21). □

As a simple consequence of this result, we obtain the following

Corollary 5.3. *If $k \notin K$, then*

$$\|\Phi_{\mu_k}(x^k)\| \le \|\Phi_{\mu_k}(x^{k-1})\|.$$

Proof. First assume that $k \notin K$ and the updating rule (21) is active (i.e., d^{k-1} is a gradient step). Taking into account the shift of indices in the definition of the set K, we directly obtain from Lemma 5.2 (b)

$$\|\Phi_{\mu_k}(x^k)\| \le \|\Phi_{\mu_k}(x^{k-1})\|.$$

On the other hand, if (21) is not active (i.e., d^{k-1} is a Newton direction), then we have $\mu_k = \mu_{k-1}$ and therefore

$$\|\Phi_{\mu_k}(x^k)\| = \|\Phi_{\mu_{k-1}}(x^k)\| < \|\Phi_{\mu_{k-1}}(x^{k-1})\|$$

by Lemma 5.2 (a). This completes the proof. □

Using these preliminary results, we are now able to show that the iterates x^k stay in a certain level set. To this end, we first note that, in all standard descent methods, the iterates would stay in the level set belonging to the level $\Psi(x^0)$ of Ψ at the initial iterate x^0. This is no longer true for our algorithm basically because we minimize different merit functions in our line search rules, namely Ψ when using a gradient step, and Ψ_{μ_k} when using a Newton step. (Note that a decrease in one merit function does not necessarily imply a decrease in the other.) Fortunately, our following result shows that the possible increase in Ψ can't be too dramatic. In fact, this result shows that all iterates x^k stay in a level set whose level can be made arbitrarily close to the level $\Psi(x^0)$.

Proposition 5.4. *The sequence $\{x^k\}$ generated by Algorithm 4.1 remains in the level set*

$$\mathcal{L}_0 := \{x \in \mathbb{R}^n \mid \Psi(x) \le (1+\alpha)^2\Psi(x^0)\}. \tag{25}$$

Proof. We define the following two index sets:

$$K_1 := \left\{ k \in K \,\middle|\, \eta\beta_{k-1} \geq \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\| \right\} \tag{26}$$

and

$$K_2 := \left\{ k \in K \,\middle|\, \eta\beta_{k-1} < \frac{1}{\alpha}\|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\| \right\}. \tag{27}$$

Then $K = \{0\} \cup K_1 \cup K_2$, where K is defined in (22). Assume K consists of $k_0 = 0 < k_1 < k_2 < \ldots$ (notice that K is not necessarily infinite). Let $k \in \mathbb{N}$ be an arbitrary but fixed index and k_j the largest number in K such that $k_j \leq k$. Then we have

$$\mu_k \leq \mu_{k_j} \quad \text{and} \quad \beta_k = \beta_{k_j}$$

in view of the updating rules in Step (S.4) of Algorithm 4.1. We divide the proof into three parts.

(a) In this part, we show that the following inequality holds:

$$\|\Phi(x^k)\| \leq \beta_{k_j} + 2\kappa\sqrt{\mu_{k_j}}. \tag{28}$$

If $k_j = k$, this inequality is obviously true since $\beta_{k_j} = \|\Phi(x^{k_j})\|$ in this case. Hence assume that $k_j < k$ in the following. From Corollary 5.3, we obtain

$$\|\Phi_{\mu_l}(x^l)\| \leq \|\Phi_{\mu_l}(x^{l-1})\|$$

for all $k_j < l < k_{j+1}$. Since $k < k_{j+1}$, this implies

$$\|\Phi_{\mu_l}(x^l)\| \leq \|\Phi_{\mu_l}(x^{l-1})\|$$

for all $k_j < l \leq k$ or, equivalently,

$$\|\Phi_{\mu_{l+1}}(x^{l+1})\| \leq \|\Phi_{\mu_{l+1}}(x^l)\|$$

for all l such that $k_j \leq l \leq k-1$. Then, by Corollary 2.4, we get for all l such that $k_j \leq l \leq k-1$:

$$\begin{aligned} \|\Phi_{\mu_{l+1}}(x^{l+1})\| + \kappa\sqrt{\mu_{l+1}} &\leq \|\Phi_{\mu_{l+1}}(x^l)\| + \kappa\sqrt{\mu_{l+1}} \\ &\leq \|\Phi_{\mu_l}(x^l)\| + \|\Phi_{\mu_{l+1}}(x^l) - \Phi_{\mu_l}(x^l)\| + \kappa\sqrt{\mu_{l+1}} \\ &\leq \|\Phi_{\mu_l}(x^l)\| + \kappa(\sqrt{\mu_l} - \sqrt{\mu_{l+1}}) + \kappa\sqrt{\mu_{l+1}} \\ &= \|\Phi_{\mu_l}(x^l)\| + \kappa\sqrt{\mu_l}. \end{aligned} \tag{29}$$

This inequality together with Corollary 2.4 gives

$$\begin{aligned}
\|\Phi(x^k)\| &\le \|\Phi_{\mu_k}(x^k)\| + \|\Phi(x^k) - \Phi_{\mu_k}(x^k)\| \\
&\le \|\Phi_{\mu_k}(x^k)\| + \kappa\sqrt{\mu_k} \\
&\le \|\Phi_{\mu_{k-1}}(x^{k-1})\| + \kappa\sqrt{\mu_{k-1}} \\
&\vdots \\
&\le \|\Phi_{\mu_{k_j}}(x^{k_j})\| + \kappa\sqrt{\mu_{k_j}} \\
&\le \|\Phi(x^{k_j})\| + \|\Phi_{\mu_{k_j}}(x^{k_j}) - \Phi(x^{k_j})\| + \kappa\sqrt{\mu_{k_j}} \\
&\le \|\Phi(x^{k_j})\| + \kappa\sqrt{\mu_{k_j}} + \kappa\sqrt{\mu_{k_j}} \\
&= \beta_{k_j} + 2\kappa\sqrt{\mu_{k_j}},
\end{aligned} \tag{30}$$

where the dots indicate the repeated use of (29). This shows that (28) holds for arbitrary $k \in \mathbb{N}$.

(b) In this part, we show that

$$\sqrt{\mu_{k_j}} \le \frac{1}{2^{j+1}}\frac{\alpha}{\kappa}\|\Phi(x^0)\|$$

and

$$\beta_{k_j} \le r^j\|\Phi(x^0)\|,$$

where

$$r := \max\{\frac{1}{2}, \eta\}.$$

Indeed, for $j = 0$, we have $k_0 = 0$ and therefore

$$\sqrt{\mu_{k_0}} = \sqrt{\mu_0} = \frac{\alpha}{2\kappa}\|\Phi(x^0)\|$$

and

$$\beta_{k_0} = \beta_0 = r^0\|\Phi(x^0)\|$$

by the definitions of μ_0 and β_0. For $j \ge 1$, Step (S.4) of Algorithm 4.1 shows that

$$\beta_{k_j} \le \eta\beta_{k_j-1} = \eta\beta_{k_{j-1}} \le r\beta_{k_{j-1}} \quad \text{for } k_j \in K_1,$$

and, using Corollary 2.4,

$$\beta_{k_j} \le \frac{1}{\alpha}\|\Phi(x^{k_j}) - \Phi_{\mu_{k_j-1}}(x^{k_j})\| \le \frac{\kappa}{\alpha}\sqrt{\mu_{k_j-1}} \le \frac{\kappa}{\alpha}\sqrt{\mu_{k_{j-1}}} \le \frac{1}{2}\beta_{k_{j-1}} \le r\beta_{k_{j-1}} \quad \text{for } k_j \in K_2.$$

Similarly, we obtain

$$\mu_{k_j} \le \frac{1}{4}\mu_{k_j-1} \le \frac{1}{4}\mu_{k_{j-1}}.$$

From the definitions of μ_0 and β_0, we thus have

$$\sqrt{\mu_{k_j}} \le \frac{1}{2^j}\sqrt{\mu_0} = \frac{1}{2^{j+1}}\frac{\alpha}{\kappa}\|\Phi(x^0)\| \tag{31}$$

and

$$\beta_{k_j} \leq r^j \beta_0 = r^j \|\Phi(x^0)\|. \tag{32}$$

This completes the proof of Part (b).

(c) In this part, we now want to verify the statement of our Proposition. Using Parts (a) and (b), we obtain

$$\begin{aligned}
\|\Phi(x^k)\| &\leq \beta_{k_j} + 2\kappa\sqrt{\mu_{k_j}} \\
&\leq r^j \|\Phi(x^0)\| + \frac{\alpha}{2^j}\|\Phi(x^0)\| \\
&\leq r^j (1+\alpha)\|\Phi(x^0)\| \\
&\leq (1+\alpha)\|\Phi(x^0)\|.
\end{aligned} \tag{33}$$

Hence $x^k \in \mathcal{L}_0$. □

Remark 5.5. *We explicitly point out that the proof of Proposition 5.4 showed that the following inequality holds for all $k \in \mathbb{N}$:*

$$\|\Phi(x^k)\| \leq r^j (1+\alpha)\|\Phi(x^0)\|,$$

where, if $K = \{k_0, k_1, k_2, \dots\}$ with $k_0 = 0$, the index $j \in \mathbb{N}$ is defined to be the largest integer $k_j \in K$ such that $k_j \leq k$.

As an immediate consequence of Remark 5.5, we obtain

Proposition 5.6. *Let $\{x^k\}$ be a sequence generated by Algorithm 4.1 and assume that the index set K is infinite. Then each accumulation point of the sequence $\{x^k\}$ is a solution of* NCP(F).

Proof. Let x^* be an accumulation point of the sequence $\{x^k\}$, and let $\{x^k\}_L$ be a subsequence converging to x^*. Since K is infinite by assumption, we obtain from Remark 5.5:

$$\|\Phi(x^*)\| = \lim_{k \in L} \|\Phi(x^k)\| \leq \lim_{j \to \infty} r^j (1+\alpha)\|\Phi(x^0)\| = 0,$$

where the exponent $j \in \mathbb{N}$ is defined as in Remark 5.5. Hence x^* is a solution of NCP(F). □

In our next result, we consider the situation that x^* is a limit point of a subsequence which consists of gradient steps only.

Proposition 5.7. *Let $\{x^k\}$ be a sequence generated by Algorithm 4.1 and let $\{x^k\}_L$ be a subsequence converging to a point $x^* \in \mathbb{R}^n$. If $d^k = -\nabla\Psi(x^k)$ for all $k \in L$, then x^* is a stationary point of Ψ.*

Proof. If the index set K is infinite, the accumulation point x^* is a solution of NCP(F) by Proposition 5.6. Hence x^* is a global minimum and therefore a stationary point of Ψ.

So let K be finite. Then, without loss of generality, we can assume that $K \cap L = \emptyset$ so that the updating rule (21) is active for all $k \in L$. This, in particular, implies that $\{\mu_k\} \to 0$.

Let $\hat{k}$ be the largest number in K (which exists since K is finite). Then we obtain from the updating rules in Step (S.4) of Algorithm 4.1 for all $k > \hat{k}$:

$$\mu_k \leq \mu_{\hat{k}}, \quad \beta_k = \beta_{\hat{k}} = \|\Phi(x^{\hat{k}})\|, \tag{34}$$

$$\|\Phi(x^k)\| > \eta\beta_{k-1} = \eta\|\Phi(x^{\hat{k}})\| > 0 \tag{35}$$

and

$$\alpha\|\Phi(x^k)\| > \|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\|. \tag{36}$$

From (35), we get

$$\Psi(x^k) > \eta^2\Psi(x^{\hat{k}}) > 0 \tag{37}$$

for all $k > \hat{k}$.

The proof is by contradiction: Assume that $\nabla\Psi(x^*) \neq 0$. Our first aim is to show that $\liminf_{k\in L} t_k = 0$. Suppose that $\liminf_{k\in L} t_k = t_* > 0$. Since $d^k = -\nabla\Psi(x^k)$ for all $k \in L$, we obtain from the Armijo-rule (18):

$$\Psi(x^{k+1}) - \Psi(x^k) \leq -\sigma t_k \|\nabla\Psi(x^k)\|^2 \leq -\frac{c}{2} \tag{38}$$

for all $k \in L$ sufficiently large, where $c := \sigma t_* \|\nabla\Psi(x^*)\|^2 > 0$. Since $\{\mu_k\} \to 0$, Corollary 2.4 shows that

$$|\Psi_{\mu_k}(x^{k+1}) - \Psi(x^{k+1})| \leq \frac{c}{4} \quad \text{and} \quad |\Psi_{\mu_k}(x^k) - \Psi(x^k)| \leq \frac{c}{4}$$

for all $k \in \mathbb{N}$ sufficiently large. Using $\{\mu_k\} \to 0$ once again and taking into account that the sequence $\{\|\Phi(x^k)\|\}$ is bounded by Proposition 5.4, we also have

$$2\kappa\sqrt{\mu_k}\|\Phi(x^k)\| + 2\kappa^2\mu_k \leq \frac{c}{4} \tag{39}$$

for all $k \in \mathbb{N}$ large enough. Let L consist of $l_0, l_1, l_2, \ldots$. Then, for all l_j sufficiently large, we obtain in a similar way as in the proof of Proposition 5.4 (see (30) and recall that K is finite):

$$\begin{array}{rcl}
\Psi(x^{l_{j+1}}) & = & \frac{1}{2}\|\Phi(x^{l_{j+1}})\|^2 \\
& \leq & \frac{1}{2}\left(\|\Phi(x^{l_j+1})\| + 2\kappa\sqrt{\mu_{l_j+1}}\right)^2 \\
& = & \Psi(x^{l_j+1}) + 2\kappa\sqrt{\mu_{l_j+1}}\|\Phi(x^{l_j+1})\| + 2\kappa^2\mu_{l_j+1} \\
& \leq & \Psi(x^{l_j+1}) + \frac{c}{4},
\end{array} \tag{40}$$

where the last inequality follows from (39). Using (38) and (40), we obtain

$$\Psi(x^{l_{j+1}}) - \Psi(x^{l_j}) = \underbrace{\Psi(x^{l_{j+1}}) - \Psi(x^{l_j+1})}_{\leq \frac{c}{4}} + \underbrace{\Psi(x^{l_j+1}) - \Psi(x^{l_j})}_{\leq -\frac{c}{2}} \leq -\frac{c}{4}$$

for all l_j large enough. Hence $\{\Psi(x^{l_j})\} \to -\infty$ for $j \to \infty$, but this contradicts the fact that $\Psi(x) \geq 0$ for all $x \in \mathbb{R}^n$. Hence we have $\liminf_{k \in L} t_k = 0$.

Subsequencing if necessary, we can assume that $\lim_{k \in L} t_k = 0$. We now want to derive a contradiction to our assumption that $\nabla \Psi(x^*) \neq 0$. Since $\lim_{k \in L} t_k = 0$, the full stepsize is never accepted for all $k \in L$ sufficiently large. Hence we obtain from the Armijo-rule (18)

$$\Psi(x^k + \lambda^{m_k - 1} d^k) > \Psi(x^k) - \sigma \lambda^{m_k - 1} \|d^k\|^2$$

or, equivalently,

$$\frac{\Psi(x^k + \lambda^{m_k - 1} d^k) - \Psi(x^k)}{\lambda^{m_k - 1}} > -\sigma \|d^k\|^2. \tag{41}$$

By taking the limit $k \to \infty$ on L, we obtain from (41), the continuous differentiability of Ψ, $d^k = -\nabla \Psi(x^k)$ for all $k \in L$ and the fact that $\lambda^{m_k - 1} \to 0$ for $k \to_L \infty$:

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

This yields $1 \leq \sigma$, a contradiction to our choice of the parameter σ. Hence we must have $\nabla \Psi(x^*) = 0$, and this completes the proof of Proposition 5.7. □

We are now able to prove the main global convergence result for Algorithm 4.1.

Theorem 5.8. *Let $\{x^k\}$ be a sequence generated by Algorithm 4.1. Then each accumulation point of the sequence $\{x^k\}$ is a stationary point of Ψ.*

Proof. If K is infinite, the conclusion follows immediately from Proposition 5.6. Hence we can assume that K contains only finitely many indices.

Similar to the proof of Proposition 5.7, we denote by $\hat{k}$ the largest index in K. Then (34), (35), (36) and (37) hold for all $k > \hat{k}$.

Let x^* be an accumulation point of the sequence $\{x^k\}$, and let $\{x^k\}_L$ be a subsequence converging to x^*. If $d^k = -\nabla \Psi(x^k)$ for infinitely many $k \in L$, then x^* is a stationary point of Ψ by Proposition 5.7. Hence we can assume without loss of generality that d^k is the Newton direction computed as a solution of the linear system (14) for all $k \in L$, so that

$$\|\Phi(x^k)\| = \|\Phi'_{\mu_k}(x^k) d^k\| \leq \|\Phi'_{\mu_k}(x^k)\| \, \|d^k\| \tag{42}$$

holds for all $k \in L$. Since K is finite, we can further assume without loss of generality that $k \notin K$ for all $k \in L$, i.e., neither the updating rule (20) nor the updating rule (21) is active for $k \in L$.

The proof is by contradiction: Assume that x^* is not a stationary point of Ψ. Since the sequence $\{\mu_k\}$ is monotonically decreasing and bounded from below, it converges to some $\mu_* \geq 0$. If $\mu_* > 0$, then it follows from the updating rules of Step (S.4) in Algorithm 4.1 that μ_k is actually constant for all k sufficiently large.

The remaining part of this proof is divided into three steps.

(a) We first show that there exist positive constants m and M such that

$$0 < m \leq \|d^k\| \leq M \text{ for all } k \in L. \tag{43}$$

In fact, if $\{\|d^k\|\}_{\tilde{L}} \to 0$ on a subset $\tilde{L} \subseteq L$, we would have from (42) that $\{\|\Phi(x^k)\|\}_{\tilde{L}} \to 0$ because the sequence $\{\Phi'_{\mu_k}(x^k)\}_{\tilde{L}}$ is obviously bounded on the convergent sequence $\{x^k\}_{\tilde{L}}$. But then the continuity of Φ would imply that $\Phi(x^*) = 0$, so that K would be infinite by Lemma 5.1. This, however, would contradict our assumption that K is finite.

On the other hand, we have from (15) for all $k \in L$:

$$-\|\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\| \, \|d^k\| \leq \Phi(x^k)^T\Phi'_{\mu_k}(x^k)d^k \leq -\rho\|d^k\|^p. \tag{44}$$

Since $\{\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\}_L$ is convergent (either by Lemma 2.6 or because μ_k is eventually constant) and therefore bounded, there exists a constant $C > 0$ such that

$$\|\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\| \leq C$$

for all $k \in L$. With (44), we have

$$\rho\|d^k\|^p \leq \|\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\|\|d^k\| \leq C\|d^k\|$$

for all $k \in L$. Since $p > 1$, this shows that $\{\|d^k\|\}_L$ is bounded. This completes the proof of Part (a).

(b) We now show that $\liminf_{k\in L} t_k = 0$. Suppose that $\liminf_{k\in L} t_k =: t_* > 0$. Then from (37) and the line search rule (17), we have for all $k \in L$ sufficiently large:

$$\Psi_{\mu_k}(x^{k+1}) - \Psi_{\mu_k}(x^k) \leq -2\sigma t_k\Psi(x^k) \leq -\sigma t_*\eta^2\Psi(x^{\hat{k}}) < 0. \tag{45}$$

We define $c := \sigma t_*\eta^2\Psi(x^{\hat{k}}) > 0$ and consider two cases.

Case 1: $\{\mu_k\} \to \mu_* > 0$.
Then we have $\mu_k = \mu_*$ is constant for $k \in \mathbb{N}$ sufficiently large. Hence we obtain from (45) for all $k \in L$ large enough:

$$\Psi_{\mu_*}(x^{k+1}) - \Psi_{\mu_*}(x^k) = \Psi_{\mu_k}(x^{k+1}) - \Psi_{\mu_k}(x^k) \leq -c. \tag{46}$$

Since μ_k is eventually constant, the updating rule (21) excludes the existence of gradient steps for $k \in \mathbb{N}$ sufficiently large. Hence, if we assume that L consists of $l_0, l_1, l_2, \ldots$, we obtain from Lemma 5.2 (a) for all l_j sufficiently large:

$$\Psi_{\mu_*}(x^{l_{j+1}}) - \Psi_{\mu_*}(x^{l_j}) \leq \Psi_{\mu_*}(x^{l_j+1}) - \Psi_{\mu_*}(x^{l_j}) \leq -c.$$

This implies

$$\Psi_{\mu_*}(x^{l_j}) \to -\infty$$

for $j \to \infty$, a contradiction to $\Psi_{\mu_*}(x) \geq 0$ for all $x \in \mathbb{R}^n$.

Case 2: $\{\mu_k\} \to 0$.
Then we obtain from Corollary 2.4 that

$$|\Psi_{\mu_k}(x^{k+1}) - \Psi(x^{k+1})| \le \frac{c}{4} \quad \text{and} \quad |\Psi_{\mu_k}(x^k) - \Psi(x^k)| \le \frac{c}{4} \tag{47}$$

for all $k \in \mathbb{N}$ sufficiently large. Again, let the sequence L consists of $l_0, l_1, l_2, \dots$. Then the following inequality holds for all l_j large enough:

$$\begin{aligned} \Psi(x^{l_j+1}) - \Psi(x^{l_j}) = &- (\Psi_{\mu_{l_j}}(x^{l_j+1}) - \Psi(x^{l_j+1})) + (\Psi_{\mu_{l_j}}(x^{l_j}) - \Psi(x^{l_j})) \\ &+ \Psi_{\mu_{l_j}}(x^{l_j+1}) - \Psi_{\mu_{l_j}}(x^{l_j}) \\ \le &\underbrace{|\Psi_{\mu_{l_j}}(x^{l_j+1}) - \Psi(x^{l_j+1})|}_{\le \frac{c}{4} \text{ by } (47)} + \underbrace{|\Psi_{\mu_{l_j}}(x^{l_j}) - \Psi(x^{l_j})|}_{\le \frac{c}{4} \text{ by } (47)} \\ &+ \underbrace{\Psi_{\mu_{l_j}}(x^{l_j+1}) - \Psi_{\mu_{l_j}}(x^{l_j})}_{\le -c \text{ by } (45)} \\ \le &- \frac{c}{2}. \end{aligned} \tag{48}$$

The remaining part of the proof for Case 2 is now similar to the one for Proposition 5.7: In particular, for l_j large enough, we can prove the following inequality in essentially the same way as in the proof of Proposition 5.7 (see (40) and recall that K is finite):

$$\Psi(x^{l_{j+1}}) \le \Psi(x^{l_j+1}) + \frac{c}{4}. \tag{49}$$

Combining (48) and (49), we obtain

$$\Psi(x^{l_{j+1}}) - \Psi(x^{l_j}) = \Psi(x^{l_{j+1}}) - \Psi(x^{l_j+1}) + \Psi(x^{l_j+1}) - \Psi(x^{l_j}) \le \frac{c}{4} - \frac{c}{2} = -\frac{c}{4}.$$

This implies $\Psi(x^{l_j}) \to -\infty$ for $j \to \infty$, contradicting the fact that $\Psi(x) \ge 0$ for all $x \in \mathbb{R}^n$.

Since both Case 1 and Case 2 lead to a contradiction, the proof of Part (b) is also completed.

(c) We now turn back to the main part of our proof, i.e., we will now derive a contradiction to our assumption that $\nabla\Psi(x^*) \neq 0$.

Because of Part (b), we have $\liminf_{k \in L} t_k = 0$. Let L_0 be a subsequence of L such that $\{t_k\}_{L_0}$ converges to 0. Then $m_k > 0$ for all $k \in L_0$ sufficiently large, where $m_k \in \mathbb{N}$ denotes the exponent from the line search rule (17). By this line search rule, we therefore have

$$-2\sigma\lambda^{m_k-1}\Psi(x^k) < \Psi_{\mu_k}(x^k + \lambda^{m_k-1}d^k) - \Psi_{\mu_k}(x^k)$$

for all $k \in L_0$ large enough. Dividing both sides by λ^{m_k-1}, we obtain

$$-2\sigma\Psi(x^k) < \frac{\Psi_{\mu_k}(x^k + \lambda^{m_k-1}d^k) - \Psi_{\mu_k}(x^k)}{\lambda^{m_k-1}}.$$

Let μ_* be the limit of $\{\mu_k\}$, and if $\mu_* = 0$, we write $\nabla\Psi_{\mu_*}(x^*)$ for the gradient of the unperturbed function Ψ at the limit point x^*. By (43) we can assume, subsequencing if necessary, that $\{d^k\}_{L_0} \to d^* \neq 0$, so that, passing to the limit, we get

$$-2\sigma\Psi(x^*) \leq \nabla\Psi_{\mu_*}(x^*)^T d^*. \tag{50}$$

For $\mu_* = 0$ this follows from Lemma 2.7, and if $\mu_* > 0$, then $\mu_k = \mu_*$ for sufficiently large k, so that (50) follows from the Mean Value Theorem.

Using (14), (36) and Corollary 2.4, we further have for $k \in L_0$:

$$\begin{aligned}
\nabla\Psi_{\mu_k}(x^k)^T d^k &= -\,\Phi(x^k)^T\Phi_{\mu_k}(x^k) \\
&= -\,2\Psi(x^k) + \Phi(x^k)^T(\Phi(x^k) - \Phi_{\mu_k}(x^k)) \\
&\leq -\,2\Psi(x^k) + \|\Phi(x^k)\|\,\|\Phi(x^k) - \Phi_{\mu_{k-1}}(x^k)\| \\
&\quad + \|\Phi(x^k)\|\,\|\Phi_{\mu_{k-1}}(x^k) - \Phi_{\mu_k}(x^k)\| \\
&\leq -\,2\Psi(x^k) + 2\alpha\Psi(x^k) + \kappa\|\Phi(x^k)\|(\sqrt{\mu_{k-1}} - \sqrt{\mu_k}) \\
&= -\,2(1-\alpha)\Psi(x^k) + \kappa\|\Phi(x^k)\|(\sqrt{\mu_{k-1}} - \sqrt{\mu_k}).
\end{aligned} \tag{51}$$

By taking the limit $k \to_{L_0} \infty$ in (51), we obtain from (50) (and Lemma 2.6 if $\mu_* = 0$)

$$-2\sigma\Psi(x^*) \leq \nabla\Psi_{\mu_*}(x^*)^T d^* \leq -2(1-\alpha)\Psi(x^*), \tag{52}$$

since $\{\|\Phi(x^k)\|\}$ is bounded (by Proposition 5.4), and $(\sqrt{\mu_{k-1}} - \sqrt{\mu_k}) \to 0$ (because $\{\mu_k\}$ converges). We have $\Psi(x^*) > 0$, because otherwise K would be infinite. Therefore (52) gives $\sigma \geq (1-\alpha)$ which is a contradiction to $\sigma < \frac{1}{2}(1-\alpha)$. This, finally, completes the proof of Theorem 5.8. □

Note that Theorem 5.8 is a subsequential convergence result to stationary points of Ψ only. However, it is well-known that such a stationary point x^* is already a solution of NCP(F) if, e.g., the Jacobian $F'(x^*)$ is a P_0-matrix, see [16, 13]. Moreover, Proposition 5.6 provides another sufficient condition for an accumulation point to be a solution of the complementarity problem.

6 Local Convergence

In this section, we want to show that Algorithm 4.1 is locally Q-superlinearly/Q-quadratically convergent under certain assumptions. As a first step in this direction, we show that the whole sequence $\{x^k\}$ generated by Algorithm 4.1 converges to a unique point x^* if certain conditions hold. The proof of this result is based on the following Proposition by Moré and Sorensen [31] (note that their result is fairly general and completely independent of any specific algorithm).

Proposition 6.1. *Assume that $x^* \in \mathbb{R}^n$ is an isolated accumulation point of a sequence $\{x^k\} \subseteq \mathbb{R}^n$ (not necessarily generated by Algorithm 4.1) such that $\{\|x^{k+1} - x^k\|\}_L \to 0$ for any subsequence $\{x^k\}_L$ converging to x^*. Then the whole sequence $\{x^k\}$ converges to x^*.*

Proposition 6.1 enables us to establish the following result.

Theorem 6.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 4.1. If one of the accumulation points of the sequence $\{x^k\}$, let us say x^*, is an isolated solution of* NCP(F)*, then $\{x^k\} \to x^*$.*

Proof. Let x^* be an isolated solution of NCP(F). We want to verify the assumptions of Proposition 6.1. To this end, we first show that x^* is also an isolated accumulation point of the sequence $\{x^k\}$.

Since x^* solves NCP(F), Lemma 5.1 shows that the index set K is infinite and $\{\mu_k\}$ converges to 0. Hence Proposition 5.6 shows that each accumulation point of the sequence $\{x^k\}$ is already a solution of NCP(F). Thus x^* is necessarily an isolated accumulation point of the sequence $\{x^k\}$.

Now let $\{x^k\}_L$ be an arbitrary subsequence of $\{x^k\}$ converging to x^*. From the updating rule in Step (S.3) of Algorithm 4.1, we have

$$\|x^{k+1} - x^k\| = \lambda^{m_k} \|d^k\| \leq \|d^k\|. \tag{53}$$

Therefore it suffices to show that $\{\|d^k\|\}_L \to 0$. Since Ψ is continuously differentiable and since the solution x^* of NCP(F) is, in particular, a stationary point of Ψ, we have

$$\{\nabla\Psi(x^k)\}_L \to \nabla\Psi(x^*) = 0. \tag{54}$$

Suppose the sequence $\{d^k\}_L$ contains only a finite number of Newton directions. Then $\{\|d^k\|\}_L \to 0$ follows immediately. Assume therefore that there is a subsequence $\{d^k\}_{L_0}$ of $\{d^k\}_L$ such that d^k is the solution of the linear system (14) for all $k \in L_0$.

From (15), we obtain

$$\rho\|d^k\|^p \leq -(\Phi'_{\mu_k}(x^k)^T\Phi(x^k))^T\, d^k \leq \|\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\|\, \|d^k\|$$

for all $k \in L_0$, from which we get

$$\|d^k\| \leq \left(\frac{\|\Phi'_{\mu_k}(x^k)^T\Phi(x^k)\|}{\rho}\right)^{\frac{1}{p-1}} \tag{55}$$

because $p > 1$. Since $\{\mu_k\} \to 0$, we obtain

$$\lim_{k\to\infty, k\in L_0} \Phi'_{\mu_k}(x^k)^T\Phi(x^k) \to \nabla\Psi(x^*) = 0$$

from Lemma 2.6. Hence the right-hand side of (55) converges to 0, so that $\{d^k\}_{L_0} \to 0$. We obviously also have $\{d^k\}_{L\setminus L_0} \to 0$ from (54) (if the set $L \setminus L_0$ is infinite). Hence (53) shows that

$$\{\|x^{k+1} - x^k\|\}_L \to 0.$$

The assertion now follows from Proposition 6.1. $\square$

Remark 6.3. *We explicitly point out that, in the proof of Theorem 6.2, we have actually shown that if the sequence $\{x^k\}$ generated by Algorithm 4.1 converges to a solution of* NCP(F)*, then $\{\|d^k\|\} \to 0$. This fact will be important in the proof of Theorem 6.6 below.*

In order to verify that Algorithm 4.1 eventually takes the full stepsize $t_k = 1$, we state the following Lemma which was shown by Chen, Qi and Sun [10, Lemma 3.2].

Lemma 6.4. *If there exists a scalar*

$$\omega \in \left[\frac{1}{2} - \frac{(1-\alpha-2\sigma)^2}{2(2+\alpha)^2}, \frac{1}{2}\right]$$

such that

$$\Psi(y) \leq \Psi(x^k) - 2\omega\Psi(x^k) \tag{56}$$

for some $k \in K$ and $y \in \mathbb{R}^n$, then it holds

$$\Psi_{\mu_k}(y) \leq \Psi_{\mu_k}(x^k) - 2\sigma\Psi(x^k), \tag{57}$$

where μ_k is the smoothing parameter in the kth step.

In the proof of our main local convergence result, we will also utilize the following Proposition which was originally shown by Facchinei and Soares [16]. An alternative proof of this result was given by Kanzow and Qi [29] under slightly different assumptions. Here we restate the result from [29].

Proposition 6.5. *Let $G\colon \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitzian and $x^* \in \mathbb{R}^n$ with $G(x^*) = 0$ such that all elements in $\partial G(x^*)$ are nonsingular, and assume that there are two subsequences $\{x^k\} \subseteq \mathbb{R}^n$ and $\{d^k\} \subseteq \mathbb{R}^n$ with*

$$\lim_{k\to\infty} x^k = x^* \quad \text{and} \quad \|x^k + d^k - x^*\| = o(\|x^k - x^*\|).$$

Then

$$\|G(x^k + d^k)\| = o(\|G(x^k)\|).$$

Before stating our local convergence result, we recall that a solution x^* of NCP(F) is called *R-regular* if the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular and the Schur complement

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

is a P-matrix, see Robinson [38]; here, we have used the standard index set notation

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

Theorem 6.6. *Let $\{x^k\}$ be a sequence generated by Algorithm 4.1. If one of the limit points of the sequence $\{x^k\}$, let us say x^*, is an R-regular solution of* NCP(F)*, then $\{x^k\} \to x^*$, and the convergence rate is at least Q-superlinear. If $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable with a locally Lipschitzian Jacobian, then the convergence rate is Q-quadratic.*

Proof. We first note that the assumed R-regularity of the solution x^* implies that all elements of the C-subdifferential $\partial_C \Phi(x^*)$ are nonsingular, see [16]. Hence Proposition 2.5 in [33] together with Proposition 2.2 shows that x^* is an isolated solution of $\Phi(x) = 0$ and therefore also of NCP(F). Hence, by Theorem 6.2, the whole sequence $\{x^k\}$ converges to x^*. Let K be again the set defined by (22), which, by Lemma 5.1, is infinite since the sequence $\{x^k\}$ converges to a solution of NCP(F). In particular, we have $\{x^k\}_K \to x^*$.

We now divide the proof into four steps.

(a) In this part, we show that, for all $k \in K$ sufficiently large, the matrix $\Phi'_{\mu_k}(x^k)$ is nonsingular and satisfies the inequality

$$\|\Phi'_{\mu_k}(x^k)^{-1}\| \leq 2c$$

for a certain constant $c > 0$.

Since $\{x^k\}$ converges to x^*, the assumed R-regularity together with the upper semicontinuity of the C-subdifferential implies that, for all $k \in \mathbb{N}$ sufficienly large, all matrices $V_k \in \partial_C \Phi(x^k)$ are nonsingular with $\|V_k^{-1}\| \leq c$ for some constant $c > 0$. We now want to show that the same is true for $\Phi'_{\mu_k}(x^k)$. Let $H_k \in \partial_C \Phi(x^k)$ such that

$$\operatorname{dist}_F(\Phi'_{\mu_k}(x^k), \partial_C \Phi(x^k)) = \|\Phi'_{\mu_k}(x^k) - H_k\|_F$$

(note that such an element exists since the set $\partial_C \Phi(x^k)$ is nonempty and compact). With (24) we have

$$\|H_k - \Phi'_{\mu_k}(x^k)\| \leq \|H_k - \Phi'_{\mu_k}(x^k)\|_F \leq \gamma \beta_k \tag{58}$$

for all $k \in K$. Hence it follows that

$$\begin{aligned} \|I - H_k^{-1}\Phi'_{\mu_k}(x^k)\| &= \|H_k^{-1}(H_k - \Phi'_{\mu_k}(x^k))\| \\ &\leq \|H_k^{-1}\| \, \|H_k - \Phi'_{\mu_k}(x^k)\| \\ &\leq \gamma \beta_k c. \end{aligned} \tag{59}$$

Since K is infinite, we have $\beta_k \to 0$ in view of the updating rules in Step (S.4) of Algorithm 4.1. Therefore, for $k \in K$ large enough such that $\beta_k \leq \frac{1}{2\gamma c}$, we have

$$\|I - H_k^{-1}\Phi'_{\mu_k}(x^k)\| \leq \frac{1}{2}.$$

From the Perturbation Lemma [14, Theorem 3.1.4], we obtain that $\Phi'_{\mu_k}(x^k)$ is nonsingular for all $k \in K$ large enough with

$$\|\Phi'_{\mu_k}(x^k)^{-1}\| \leq 2\|H_k^{-1}\| \leq 2c. \tag{60}$$

Hence system (14) admits a solution for all $k \in K$ sufficiently large, and the proof of Part (a) is completed.

(b) We next want to show that, for all $k \in K$ sufficiently large, the solution d^k of the linear system (14) satisfies the descent condition (15).

To this end, we first note that the linear system (14) has a unique solution for all $k \in K$ sufficiently large by Part (a). We now show that these d^k satisfy the inequality

$$\Phi(x^k)^T \Phi'_{\mu_k}(x^k) d^k \leq -\rho_1 \|d^k\|^2 \tag{61}$$

for a certain positive constant ρ_1. Indeed, this follows from the fact that

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

by (14), so that (60) implies

$$\Phi(x^k)^T \Phi'_{\mu_k}(x^k) d^k = -\|\Phi(x^k)\|^2 \leq -\frac{\|d^k\|^2}{4c^2} \tag{62}$$

for all $k \in K$ large enough. Hence (61) follows from (62) by taking $\rho_1 = 1/(4c^2)$. Since $\{\|d^k\|\} \to 0$ by Remark 6.3, it is now easy to see that (61) eventually implies (15) for any $\rho > 0$ und $p > 2$. Hence, for all $k \in K$ sufficiently large, the search direction d^k is always given by (14).

(c) In view of Parts (a) and (b), the search direction d^k is given by (14) for all $k \in K$ large enough. In this step, we want to show that there is an index $\bar{k} \in K$ such that if $k \in K$ is any index with $k \geq \bar{k}$, then the index $k+1$ also belongs to the set K and $x^{k+1} = x^k + d^k$. Repeating this argument, it then follows that eventually all iterations k belong to the set K, and that the full step $t_k = 1$ is always accepted.

In order to prove this statement, we recall from Part (a) that there is a constant $c > 0$ such that $\|\Phi'_{\mu_k}(x^k)^{-1}\| \leq 2c$ for all $k \in K$ sufficiently large. From Algorithm 4.1 and (58), we therefore obtain for all $k \in K$ large enough:

$$\begin{aligned}
& \|x^k + d^k - x^*\| \\
= \; & \|x^k - x^* - \Phi'_{\mu_k}(x^k)^{-1}\Phi(x^k)\| \\
= \; & \|\Phi'_{\mu_k}(x^k)^{-1}(\Phi'_{\mu_k}(x^k)(x^k - x^*) - \Phi(x^k) + \Phi(x^*))\| \\
\leq \; & \|\Phi'_{\mu_k}(x^k)^{-1}\| \left(\|(\Phi'_{\mu_k}(x^k) - H_k)(x^k - x^*)\| + \|H_k(x^k - x^*) - \Phi(x^k) + \Phi(x^*)\|\right) \\
\leq \; & 2c(\gamma\beta_k\|x^k - x^*\| + \|H_k(x^k - x^*) - \Phi(x^k) + \Phi(x^*)\|),
\end{aligned} \tag{63}$$

where, again, $H_k \in \partial_C \Phi(x^k)$ is chosen in such a way that

$$\mathrm{dist}_F(\Phi'_{\mu_k}(x^k), \partial_C\Phi(x^k)) = \|\Phi'_{\mu_k}(x^k) - H_k\|_F,$$

see Part (a) of this proof. Using Proposition 2.2 (a) and taking into account that $\beta_k \to 0$, we have

$$\|x^k + d^k - x^*\| = o(\|x^k - x^*\|) \quad \text{for } k \to \infty, \; k \in K. \tag{64}$$

Hence (64) and Proposition 6.5 show that

$$\|\Phi(x^k + d^k)\| = o(\|\Phi(x^k)\|) \quad \text{for } k \to \infty,\ k \in K. \tag{65}$$

Let $\omega := \max\left\{\frac{1}{2} - \frac{(1-\alpha-2\sigma)^2}{2(2+\alpha)^2}, \frac{1-\eta^2}{2}\right\}$. Then (65) implies that there exists an index $\bar{k} \in K$ such that

$$\Psi(x^k + d^k) \leq \Psi(x^k) - 2\omega\Psi(x^k) \tag{66}$$

for all $k \in K$ with $k \geq \bar{k}$. Hence, by Lemma 6.4, we therefore have

$$\Psi_{\mu_k}(x^k + d^k) \leq \Psi_{\mu_k}(x^k) - 2\sigma\Psi(x^k) \tag{67}$$

for all $k \in K$ with $k \geq \bar{k}$. Hence the full stepsize of 1 will eventually be accepted for all $k \geq \bar{k}$, $k \in K$. In particular, $x^{\bar{k}+1} = x^{\bar{k}} + d^{\bar{k}}$, and from (66) and the definition of ω, we obtain

$$\|\Phi(x^{\bar{k}+1})\| \leq \sqrt{1-2\omega}\|\Phi(x^{\bar{k}})\| \leq \eta\|\Phi(x^{\bar{k}})\| = \eta\beta_{\bar{k}},$$

which implies that $\bar{k}+1 \in K$, cf. (22). Repeating the above process, we may prove that for all $k \geq \bar{k}$, we have

$$k \in K$$

and

$$x^{k+1} = x^k + d^k.$$

This completes the proof of Part (c).

(d) We now turn to the final part of the proof where we want to verify the Q-superlinear/Q-quadratic rate of convergence. Since $k \in K$ and $t_k = 1$ for all $k \in \mathbb{N}$ sufficiently large by Part (c), the Q-superlinear convergence follows immediately from (64).

If $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable with a locally Lipschitzian Jacobian, then Proposition 2.2 (b) shows that

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

Since Φ is obviously locally Lipschitzian, we further have

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

Hence the Q-quadratic rate of convergence of $\{x^k\}$ to x^* follows from (63) by using similar arguments as for the proof of the local Q-superlinear convergence. $\square$

7 Numerical Results

We implemented the Jacobian smoothing method from Algorithm 4.1 in MATLAB and tested it on a SUN SPARC 20 station. As test problems, we use all complementarity problems and all available starting points from the MCPLIB and GAMSLIB collections.

The implemented version of the algorithm differs from the one described before in mainly two aspects: On the one hand, we replaced the monotone Armijo-rule by a nonmonotone variant [21]. For the details of the implementation of this nonmonotone Armijo-rule, we refer the interested reader to [32].

On the other hand, we incorporated a heuristic backtracking strategy in our implementation in order to avoid domain violations which occur quite often since the mapping F in many examples of the test libraries is not defined everywhere. To this end, we first compute

$$\hat{t}_k := \max\{\nu_k^l \,|\, l = 0, 1, 2, \dots\}$$

in such a way that $F(x^k + \hat{t}_k d^k)$ is well-defined, and then we take $\hat{t}_k$ as initial steplength with which we go into the nonmonotone line search test. Note that we allow the backtracking factor ν_k to vary in each iteration. In our implementation we choose ν_k between 0.5 and 0.75, depending on the success of the last backtracking step.

The algorithm terminates if one of the following conditions is satisfied:

$$\Psi(x^k) \le \epsilon_1, \;\; \|\nabla\Psi(x^k)\| \le \epsilon_2, \;\; k > k_{\max} \text{ or } t_k < t_{\min}.$$

In the implementation we used the following parameter settings:

$$\rho = 10^{-18}, \;\; p = 2.1, \;\; \lambda = 0.5, \;\; \sigma = 10^{-4}, \;\; \gamma = 30, \;\; \alpha = 0.95, \;\; \eta = 0.9,$$

and

$$\epsilon_1 = 10^{-12}, \;\; \epsilon_2 = 10^{-6}, \;\; k_{\max} = 300, \;\; t_{\min} = 10^{-16}.$$

We report the results for all complementarity problems in the MCPLIB and GAMSLIB libraries and all available starting points in Tables 1 and 2, respectively. The columns in these tables have the following meanings:

problem:	name of the test problem in the specific test library
n:	dimension of the test problem
SP:	number of starting point
k:	number of iterations
F-eval:	number of function evaluations
N:	number of Newton steps taken
G:	number of gradient steps taken
$\Psi(x^f)$:	$\Psi(x)$ at the final iterate $x = x^f$
$\|\nabla\Psi(x^f)\|$:	$\|\nabla\Psi(x)\|$ at the final iterate $x = x^f$
B:	number of backtracking steps.

From the definition of the algorithm it follows that the number of Jacobian evaluations is one more than the number of iterations k.

Table 1: Numerical results for MCPLIB test problems

problem	n	SP	k	F-ev.	N	G	$\Psi(x^f)$	$\|\nabla\Psi(x^f)\|$	B
bertsekas	15	1	34	271	34	0	1.5e-19	3.5e-08	0
bertsekas	15	2	37	353	37	0	1.4e-16	1.0e-06	0
bertsekas	15	3	42	406	42	0	2.5e-19	4.4e-08	0
billups	1	1	27	389	27	0	4.1e-17	1.8e-08	0
colvdual	20	1	15	37	15	0	1.0e-17	4.8e-07	0
colvdual	20	2	26	64	26	0	9.4e-16	4.4e-06	0
colvnlp	15	1	16	39	16	0	2.7e-17	7.8e-07	0
colvnlp	15	2	14	26	14	0	6.8e-15	1.7e-05	0
cycle	1	1	3	5	3	0	8.1e-16	4.0e-08	0
explcp	16	1	5	6	5	0	2.8e-15	7.5e-08	0
hanskoop	14	1	9	14	9	0	2.9e-16	3.3e-08	0
hanskoop	14	2	9	12	9	0	1.4e-17	1.8e-08	0
hanskoop	14	3	8	12	8	0	9.5e-16	1.5e-07	0
hanskoop	14	4	9	13	9	0	4.2e-18	1.0e-08	0
hanskoop	14	5	10	16	10	0	3.3e-18	8.9e-09	1
josephy	4	1	8	11	8	0	1.3e-19	1.7e-09	0
josephy	4	2	7	12	7	0	1.7e-18	1.5e-08	0
josephy	4	3	13	18	13	0	1.0e-14	4.8e-07	0
josephy	4	4	5	6	5	0	2.6e-20	7.6e-10	0
josephy	4	5	5	6	5	0	2.4e-13	2.6e-06	0
josephy	4	6	6	8	6	0	8.1e-21	9.9e-10	0
kojshin	4	1	10	17	10	0	3.3e-24	1.6e-11	0
kojshin	4	2	9	21	9	0	2.9e-15	1.2e-07	0
kojshin	4	3	7	10	7	0	1.8e-15	2.0e-07	0
kojshin	4	4	12	26	12	0	8.0e-17	1.6e-07	0
kojshin	4	5	5	7	5	0	5.0e-18	8.8e-09	0
kojshin	4	6	6	8	6	0	4.7e-25	8.5e-12	0
mathinum	3	1	7	11	7	0	1.7e-24	3.8e-12	0
mathinum	3	2	5	6	5	0	4.4e-15	2.6e-07	0
mathinum	3	3	5	6	5	0	9.2e-18	8.6e-09	0
mathinum	3	4	7	8	7	0	5.1e-23	2.8e-11	0
mathisum	4	1	5	7	5	0	4.1e-19	2.1e-09	0
mathisum	4	2	6	7	6	0	1.5e-13	1.3e-06	0
mathisum	4	3	8	10	8	0	9.0e-17	2.3e-08	0
mathisum	4	4	6	7	6	0	1.5e-22	4.1e-11	0
nash	10	1	8	9	8	0	5.3e-20	2.4e-08	0
nash	10	2	11	25	11	0	1.8e-22	6.9e-10	0

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

problem	n	SP	k	F-ev.	N	G	$\Psi(x^f)$	$\|\nabla\Psi(x^f)\|$	B
pgvon105	105	1	33	81	33	0	1.1e-13	4.7e-03	33
pgvon105	105	2	33	98	33	0	1.3e-14	7.3e-03	31
pgvon105	105	3	69	251	69	0	6.2e-17	5.0e-04	68
pgvon106	106	1	23	49	23	0	4.6e-14	4.0e-07	23
powell	16	1	13	41	13	0	3.3e-17	9.1e-08	4
powell	16	2	14	36	14	0	2.4e-14	3.3e-06	4
powell	16	3	23	45	23	0	1.3e-13	1.5e-06	4
powell	16	4	16	45	16	0	9.7e-16	5.7e-07	6
scarfanum	13	1	10	13	10	0	1.7e-16	1.7e-07	0
scarfanum	13	2	12	15	12	0	1.7e-16	1.7e-07	0
scarfanum	13	3	12	16	12	0	1.7e-16	1.7e-07	1
scarfasum	14	1	8	11	8	0	1.1e-18	3.1e-08	0
scarfasum	14	2	10	14	10	0	9.6e-17	2.8e-07	0
scarfasum	14	3	11	14	11	0	2.5e-19	1.4e-08	0
scarfbnum	39	1	23	36	23	0	1.7e-14	3.4e-05	0
scarfbnum	39	2	24	42	24	0	2.4e-14	3.7e-05	0
scarfbsum	40	1	20	56	20	0	1.2e-16	1.9e-06	0
scarfbsum	40	2	26	72	26	0	9.1e-20	5.2e-08	0
sppe	27	1	7	8	7	0	4.8e-14	4.4e-07	0
sppe	27	2	6	7	6	0	4.8e-25	2.9e-12	0
tobin	42	1	9	12	9	0	4.8e-13	9.9e-07	0
tobin	42	2	11	15	11	0	4.8e-24	3.1e-12	0

Looking at Tables 1 and 2, the most obvious observation is that we do not have a single failure, i.e., the main termination criterion

$$\Psi(x^k) \leq 10^{-12}$$

is satisfied for all test problems including the difficult ones like `billups`, `colvdual`, `vonthmcp` and `vonthmge`, to mention just a few.

As known to the authors, there is currently only one other algorithm available which also has no failures on these problems, namely the semismooth Newton-type method by Chen, Chen and Kanzow [5]. Compared to that algorithm, it seems that our Jacobian smoothing method sometimes needs less many iterations, whereas the number of function evaluations is usually higher. This may indicate that the step size rule (17) is not "optimal" and may be improved. However, function evaluations are, in general, considerably cheaper than, e.g., the solution of the linear system (14). We also stress that the philosophy of these two methods is different, so it is difficult to compare them with each other.

On the other hand, however, we could try to compare our algorithm with its underlying semismooth Newton method from De Luca et al. [13]. It turns out that our algorithm is more reliable and that we use considerably less many gradient steps. In fact, we have just

Table 2: Numerical results for GAMSLIB test problems

problem	n	SP	k	F-ev.	N	G	$\Psi(x^f)$	$\|\nabla\Psi(x^f)\|$	B
cafemge	101	1	11	19	11	0	7.9e-25	1.7e-09	0
cammge	128	1	0	1	0	0	5.1e-13	3.1e-04	0
co2mge	208	1	1	2	1	0	1.3e-14	1.0e-07	0
dmcmge	170	1	88	523	88	0	1.3e-21	2.1e-07	1
etamge	114	1	20	49	20	0	1.6e-15	3.6e-05	0
finmge	153	1	0	1	0	0	2.2e-14	7.6e-06	0
hansmcp	43	1	17	31	17	0	3.3e-14	7.8e-07	0
hansmge	43	1	14	30	14	0	4.9e-13	9.1e-07	0
harkmcp	32	1	13	16	13	0	2.0e-16	2.9e-08	0
kehomge	9	1	10	12	10	0	1.7e-20	7.9e-09	0
mr5mcp	350	1	10	17	10	0	1.7e-18	2.8e-07	1
nsmge	212	1	12	19	12	0	5.6e-18	2.9e-07	0
oligomcp	6	1	6	7	6	0	7.1e-17	1.5e-07	0
sammge	23	1	0	1	0	0	0.0	0.0	0
scarfmcp	18	1	9	12	9	0	9.2e-17	1.3e-07	1
scarfmge	18	1	11	15	11	0	5.3e-13	1.1e-05	0
shovmge	51	1	1	2	1	0	5.6e-14	5.7e-05	0
threemge	9	1	0	1	0	0	0.0	0.0	0
transmcp	11	1	13	22	13	0	3.1e-16	2.5e-08	0
two3mcp	6	1	8	12	8	0	4.8e-13	2.0e-05	0
unstmge	5	1	8	9	8	0	1.6e-13	7.6e-07	0
vonthmcp	125	1	54	280	54	0	6.1e-15	2.9e-02	37
vonthmge	80	1	31	97	30	1	4.5e-13	1.3e-04	0

one gradient step, namely on example `vonthmge`. We believe that this indicates that the smoothing parameter μ regularizes the Jacobian matrix $\Phi'_\mu(x)$ to some extent. This is also reflected by some known theoretical results, e.g., the Jacobian $\Phi'_\mu(x)$ is nonsingular if $F'(x)$ is a P_0-matrix (see [26]), whereas an element from the C-subdifferential $\partial_C\Phi(x)$ is nonsingular only under a slightly stronger assumption (see [13]).

We finally stress that we also tested some other parameter settings; there, we usually had some more gradient steps, but still less than for the method from [13]. This fact may explain why our Jacobian smoothing method seems to be superior to its underlying semismooth Newton method from [13] since it is well-accepted in the community that taking as many Newton steps as possible usually improves the overall behaviour of the algorithm.

8 Final Remarks

In this paper, we introduced a new algorithm for the solution of a general (i.e., not necessarily monotone) complementarity problem. We call this algorithm a Jacobian smoothing method since, basically, it is a perturbation of a semismooth Newton method being applied to a reformulation of the complementarity problem as a nonsmooth system of equations $\Phi(x) = 0$. In this perturbation, we replace an element from the generalized Jacobian by a standard Jacobian of a smooth operator Φ_μ which approximates Φ for $\mu \to 0$.

The basic idea of this Jacobian smoothing method is taken from the recent paper [10] by Chen, Qi and Sun. We modified their algorithm in such a way that it becomes applicable to general complementarity problems. Although this modification makes the convergence analysis rather technical (especially the global one), the main convergence results are quite nice. Moreover, the numerical performance is extremely promising. In fact, we are able to solve all complementarity problems from the MCPLIB and GAMSLIB test problem collections. In particular, our Jacobian smoothing method is considerably more reliable than the semismooth method by De Luca et al. [13] which is the underlying semismooth Newton method for our algorithm.

It would be interesting to see how our perturbation technique would work if we apply it to other equation-reformulations of the nonlinear complementarity problem like those presented in [28, 35, 5]. Finally, it would also be interesting to see how the Jacobian smoothing method would work on mixed complementarity problems. An extension to this more general class of problems seems possible by using, e.g., an idea from Billups [1], see also Qi [35] and Sun and Womersley [39]. We leave this as a future research topic.

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] S.C. Billups, S.P. Dirkse and M.C. Ferris: *A comparison of algorithms for large scale mixed complementarity problems.* Computational Optimization and Applications 7, 1997, pp. 3–25.

[3] J. Burke and S. Xu: *The global linear convergence of a non-interior path-following algorithm for linear complementarity problems.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, December 1996 (revised July 1997).

[4] B. Chen and X. Chen: *A global and local superlinear continuation-smoothing method for $P_0 + R_0$ and monotone NCP.* Technical Report, Department of Management and Systems, Washington State University, Pullman, WA, May 1997.

[5] B. Chen, X. Chen and C. Kanzow: *A penalized Fischer-Burmeister NCP-function: Theoretical investigation and numerical results.* Preprint 126, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, September 1997.

[6] B. Chen and P.T. Harker: *A non-interior-point continuation method for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 14, 1993, pp. 1168–1190.

[7] B. Chen and P.T. Harker: *Smooth approximations to nonlinear complementarity problems.* SIAM Journal on Optimization 7, 1997, pp. 403–420.

[8] B. Chen and N. Xiu: *A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing function.* Technical Report, Department of Management and Systems, Washington State University, Pullman, WA, 1997.

[9] C. Chen and O.L. Mangasarian: *A class of smoothing functions for nonlinear and mixed complementarity problems.* Computational Optimization and Applications 5, 1996, pp. 97–138.

[10] X. Chen, L. Qi and D. Sun: *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities.* Mathematics of Computation, to appear.

[11] X. Chen and Y. Ye: *On homotopy-smoothing methods for variational inequalities.* Technical Report AMR 96/39, School of Mathematics, The University of New South Wales, Sydney, Australia, December 1996.

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

[13] 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.

[14] J.E. Dennis, Jr., and R.B. Schnabel: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, NJ, 1983 (reprinted by SIAM, Philadelphia, PA, 1996).

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

[16] 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.

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

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

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

[20] 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, 1997, pp. 105–116.

[21] 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.

[22] P.T. Harker and J.-S. Pang: *Finite dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications.* Mathematical Programming 48, 1990, pp. 161–220.

[23] K. Hotta and A. Yoshise: *Global convergence of a class of non-interior-point algorithms using Chen-Harker-Kanzow functions for nonlinear complementarity problems.* Technical Report 708, Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba, Ibaraki 305, Japan, December 1996.

[24] H. Jiang: *Smoothed Fischer-Burmeister equation methods for the complementarity problem.* Technical Report, Department of Mathematics, The University of Melbourne, Parkville, Victoria, Australia, June 1997.

[25] H. Jiang and L. Qi: *A new nonsmooth equations approach to nonlinear complementarity problems.* SIAM Journal on Control and Optimization 35, 1997, pp. 178–193.

[26] C. Kanzow: *Some noninterior continuation methods for linear complementarity problems.* SIAM Journal on Matrix Analysis and Applications 17, 1996, pp. 851–868.

[27] C. Kanzow: *A new approach to continuation methods for complementarity problems with uniform P-functions.* Operations Research Letters 20, 1997, pp. 85–92.

[28] C. Kanzow and H. Kleinmichel: *A new class of semismooth Newton-type methods for nonlinear complementarity problems.* Preprint 118, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, January 1997 (revised September 1997).

[29] C. Kanzow and H.-D. Qi: *A QP-free constrained Newton-type method for variational inequality problems.* Preprint 121, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, March 1997.

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

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

[32] H. Pieper: *Ein Glättungsverfahren zur Lösung von nichtlinearen Komplementaritätsproblemen.* Diploma Thesis, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, May 1997 (in German).

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

[34] L. Qi: *C-differentiability, C-differential operators and generalized Newton methods.* Technical Report, School of Mathematics, The University of New South Wales, Sydney, Australia, January 1996.

[35] L. Qi: *Regular pseudo-smooth NCP and BVIP functions and globally and quadratically convergent generalized Newton methods for complementarity and variational inequality problems.* Technical Report AMR 97/14, School of Mathematics, The University of New South Wales, Sydney, Australia, July 1997 (revised September 1997).

[36] L. Qi and D. Sun: *Globally linearly, and globally and locally superlinearly convergent versions of the Hotta-Yoshise non-interior point algorithm for nonlinear complementarity problems.* Technical Report, School of Mathematics, The University of New South Wales, Sydney, Australia, May 1997.

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

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

[39] D. Sun and R.S. Womersley: *A new unconstrained differentiable merit function for box constrained variational inequality problems and a damped Gauss-Newton method.* Technical Report AMR 96/37, School of Mathematics, The University of New South Wales, Sydney, Australia, 1996.

[40] P. Tseng: *Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, July 1997.

[41] S. Xu: *The global linear convergence of an infeasible non-interior path-following algorithm for complementarity problems with uniform P-functions.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, December 1996.

[42] S. Xu: *The global linear convergence and complexity of a non-interior path-following algorithm for monotone LCP based on Chen-Harker-Kanzow-Smale smooth functions.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, February 1997.

[43] S. Xu and J.V. Burke: *A polynomial time interior-point path-following algorithm for LCP based on Chen-Harker-Kanzow smoothing techniques.* Technical Report, Department of Mathematics, University of Washington, Seattle, WA, September 1996 (revised July 1997).