A THEORETICAL AND NUMERICAL COMPARISON OF SOME SEMISMOOTH ALGORITHMS FOR COMPLEMENTARITY PROBLEMS

Tecla De Luca[1,4], Francisco Facchinei[2,5] and Christian Kanzow[3,6,7]

[1] Fondazione Ugo Bordoni
Via Baldassarre Castiglione 59, I-00142 Roma, Italy
e-mail: tdeluca@fub.it

[2] Università di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica
Via Buonarroti 12, I-00185 Roma, Italy
e-mail: soler@dis.uniroma1.it

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

December 30, 1997

[4]The work of this author was partially supported by NATO under grant CRG 960137.

[5]The work of this author was partially supported by MURST National Program MOST and by NATO under grant CRG 960137.

[6]Current address (October 1, 1997 — September 30, 1998): Computer Sciences Department, University of Wisconsin — Madison, 1210 West Dayton Street, Madison, WI 53706, USA; e-mail: kanzow@cs.wisc.edu.

[7]The work of this author was supported by the DFG (Deutsche Forschungsgemeinschaft) and by NATO under grant CRG 960137.

Abstract: In this paper we introduce a general line search scheme which easily allows us to define and analyze known and new semismooth algorithms for the solution of nonlinear complementarity problems. We enucleate the basic assumptions that a search direction to be used in the general scheme has to enjoy in order to guarantee global convergence, local superlinear/quadratic convergence or finite convergence. We examine in detail several different semismooth algorithms and compare their theoretical features and their practical behavior on a set of large-scale problems.

Key Words: Nonlinear complementarity problem, semismoothness, Newton's method, projected gradient method, large-scale problem.

AMS Subject Classification: 90C33, 90C06, 65K05.

1 Introduction

The main aim of this paper is twofold: On the one hand, we present a general scheme which allows us to easily analyze and compare, in a unified framework, the characteristics of several semismooth algorithms for the solution of the nonlinear complementarity problem

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0, \qquad \text{(NCP}(F))$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function and all inequalities are taken componentwise. On the other hand, we give a numerical comparison of several realizations of our general framework. Although some of these realizations are known from the literature, some others lead to new algorithms. Interestingly, it turns out that these new semismooth algorithms have stronger theoretical properties than the known ones and that their numerical behaviour is at least comparable.

The nonlinear complementarity problem has important applications in operations research, economic equilibrium models and in the engineering sciences, see, e.g., [18, 12]. For this reason, there is a growing interest in finding efficient and robust algorithms for solving NCP(F). This reflects in an increasing number of proposals of solution schemes for NCP(F) in recent years. In these recent developments an important role has been played by the *semismooth methods*, i.e., by those methods that attempt to solve the complementarity problem by first reformulating it as a semismooth system of equations and then by applying a generalized Newton method to solve this system. We refer the reader to Subsection 2.1 for a short overview on semismooth Newton methods and to Subsection 2.3 for the presentation of some basic properties of two different reformulations which will be considered in this paper.

The first semismooth methods were based on the Fischer function [13]:

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

The function φ_F has the property that

$$\varphi_F(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0, \qquad (1)$$

so that the NCP(F) can be reformulated as a system of nonsmooth equations

$$\Phi_F(x) = 0,$$

where $\Phi_F : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

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

The Fischer function, or some suitable modifications of it [25, 23, 22, 2], has maintained a central role in the development of semismooth methods up to now; in fact it appears to possess many favourable properties both from the theoretical and numerical point of view.

In this paper we introduce a general line search scheme based on the semismooth Newton method and on the Fischer function. This scheme encompasses several known methods and easily allows the development of new, interesting ones. More precisely, our general algorithmic scheme allows the use of a variety of different search directions; then it monitors convergence of the iterates by using the merit function

$$\Psi_F(x) := \frac{1}{2}\Phi_F(x)^T\Phi_F(x), \tag{2}$$

i.e., the natural merit function of the operator Φ_F. We shall consider in detail several different specializations of the general scheme and compare the resulting algorithms both from the theoretical and numerical point of view.

Another motivation for this work comes from the following observation: Almost all existing semismooth methods reformulate the complementarity problem as a system of equations

$$\Phi(x) = 0$$

in order to obtain a locally fast convergent algorithm, and then use the corresponding merit function

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

in order to globalize this method. This is a very natural approach which, usually, leads to very good algorithms. However, and this will also be discussed in more detail in Sections 2 and 3, sometimes one reformulation of the complementarity problem gives rise to a nicer merit function (e.g., weaker conditions for a stationary point to be a global minimum), whereas another reformulation gives the better search direction and therefore, in particular, the better local algorithm (e.g., finite termination or quadratic convergence under weaker assumptions). So it seems interesting to combine "the best" merit function with several schemes to compute a search direction. In fact, the probably most interesting algorithms among those studied in this paper are based on this combination idea.

The organization of the paper is as follows: In Section 2, we restate some known properties of nonsmooth Newton methods, prove some preliminary results on certain regularity concepts and apply these results to two specific reformulations of the complementarity problem NCP(F). Our general algorithmic scheme is presented in Section 3. There, based on fairly general assumptions on the search directions computed within the algorithmic scheme, we also prove global and local convergence results. In Section 4, we consider six different realizations of our class of algorithms. We show that these realizations satisfy, under suitable conditions, all or most of the assumptions required for the search directions in the previous section. In Section 5, we then present a summary of our extensive numerical testing when using four of the six realizations of our general scheme. We conclude this paper with some final remarks in Section 6.

Some words about our notation: We say that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a C^1-function if F is continuously differentiable, and an LC^1-function if F is differentiable with a locally Lipschitzian Jacobian. The Jacobian of a C^1-function $F : \mathbb{R}^n \to \mathbb{R}^m$ at a point $x \in \mathbb{R}^n$ is denoted by $F'(x)$. If $M \in \mathbb{R}^{n\times n}$ is any given matrix with elements m_{ij} and $J, K \subseteq \{1, \ldots, n\}$ are two nonempty subsets, we write M_{JK} for the submatrix consisting of the elements m_{ij} with $i \in J$ and $j \in K$. A similar notation will be used for subvectors.

2 Preliminaries

In this section we collect some material that will be used subsequently. In particular, in Subsection 2.1 we recall some known schemes for the solution of systems of semismooth equations along with their main convergence properties; these schemes will constitute the core step in the algorithms considered in this paper. In Subsection 2.2, in order to be able to better appreciate the theoretical differences between the algorithms we shall study, we elucidate the relations between b-regularity and R-regularity, the two assumptions most commonly used in analyzing convergence rates of algorithms for NCP(F). In particular, we point out the rather unexpected fact that, for P_0-problems, R-regularity and b-regularity coincide. Finally, in Subsection 2.3 we apply these regularity concepts to prove nonsingularity results for two specific reformulations of the complementarity problem.

2.1 Local Algorithms for Semismooth Systems

Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be a locally Lipschitzian function. Then, by Rademacher's theorem, G is almost everywhere differentiable. If we denote by D_G the set of points $x \in \mathbb{R}^n$ at which G is differentiable, we can define the *B-subdifferential* of G at x as

$$\partial_B G(x) := \{H \in \mathbb{R}^{n\times n} \,|\, H = \lim_{x^k \in D_G, x^k \to x} G'(x^k)\},$$

see [32]. It is easy to see that this set is nonempty and compact. The convex hull of this set,

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

is the *generalized Jacobian* of G at x, see [4]. G is said to be *semismooth* at x [27, 33, 32] if it is directionally differentiable at x and

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

for any $d \to 0$ and any $H \in \partial G(x + d)$, where $G'(x; d)$ denotes the directional derivative of G at x in the direction d. Analogously, G is called *strongly semismooth* at x if

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

for any $d \to 0$ and any $H \in \partial G(x + d)$. We call G (strongly) semismooth if it is (strongly) semismooth at any point $x \in \mathbb{R}^n$.

If G is a C^1-function, then G is semismooth and

$$\partial_B G(x) = \partial G(x) = \{G'(x)\}$$

for every $x \in \mathbb{R}^n$; moreover, if G is an LC^1-function, then G is strongly semismooth. This motivates the following generalization of the classical Newton method for the solution of the semismooth system of equations

$$G(x) = 0. \tag{3}$$

Generalized Newton Method

0. Choose $x^0 \in \mathbb{R}^n$ and set $k := 0$.
1. Select an element $H_k \in \partial_B G(x^k)$. Compute a solution $d^k \in \mathbb{R}^n$ of the generalized Newton equation
$$H_k d = -G(x^k). \tag{4}$$
2. Let $x^{k+1} = x^k + d^k, k = k + 1$ and go to Step 1.

Note that we do not incorporate any termination criterion in this and in the other algorithms presented in this section, since, in order to simplify the statement of their convergence properties, we always assume that they generate an infinite sequence of points.

We call a solution $x^* \in \mathbb{R}^n$ of (3) *BD-regular* if all elements $H \in \partial_B G(x^*)$ are nonsingular, see [32]. Now we can restate from [32] the following local convergence result for the above algorithm.

Theorem 2.1 *Assume that $G : \mathbb{R}^n \to \mathbb{R}^n$ is semismooth (strongly semismooth) and $x^* \in \mathbb{R}^n$ is a BD-regular solution of (3). Then, for any $x^0 \in \mathbb{R}^n$ sufficiently close to x^*, the above algorithm is well-defined and generates a sequence $\{x^k\}$ which converges to x^* Q-superlinearly (Q-quadratically).*

It is possible to consider variants of the Generalized Newton Method that are more suited to large-scale problems. The most obvious variant is an *inexact* version in which the equation (4) is only solved inexactly.

Generalized Inexact Newton Method

0. Choose $x^0 \in \mathbb{R}^n$ and set $k := 0$.
1. Select an element $H_k \in \partial_B G(x^k)$. Compute a solution $d^k \in \mathbb{R}^n$ of the equation
$$H_k d = -G(x^k) + r^k \tag{5}$$
for a suitable vector $r^k \in \mathbb{R}^n$.
2. Let $x^{k+1} = x^k + d^k, k = k + 1$ and go to Step 1.

The vector r^k in (5) is called the *residual* and measures the inaccuracy with which the Newton equation (4) is solved. Note that usually, in actual computations, the vector r^k is not fixed beforehands. Instead, an iterative method is used to solve the linear system (4), and this method is stopped when the norm of the residual r^k is smaller than a prefixed accuracy. The following result can be established, see [26, 9].

Theorem 2.2 *Let G be semismooth and let x^* be a BD-regular solution of the system $G(x) = 0$. Suppose that $\{x^k\}$ is a sequence generated by the Generalized Inexact Newton Method. Then the following statements hold:*

(a) There are numbers $\bar{\eta} > 0$ and $\varepsilon > 0$ such that, if $\|x^0 - x^*\| \leq \varepsilon$ and $\|r^k\| \leq \bar{\eta}\|G(x^k)\|$ for all k, the sequence $\{x^k\}$ is well-defined and converges Q-linearly to the solution x^*.

(b) If the sequence $\{x^k\}$ converges to the solution x^*, then the rate of convergence is Q-superlinear if and only if $\|r^k\| = o(\|G(x^k)\|)$.

(c) If the sequence $\{x^k\}$ converges to the solution x^* and if G is strongly semismooth, then the rate of convergence is Q-quadratic if and only if $\|r^k\| = O(\|G(x^k)\|^2)$.

Paralleling well-known results in the smooth case, we can also consider a Levenberg-Marquardt type version of the Generalized Newton Method.

Generalized Inexact Levenberg-Marquardt Method

0. Choose $x^0 \in \mathbb{R}^n$ and set $k := 0$.
1. Select an element $H_k \in \partial_B G(x^k)$ and a nonnegative σ_k. Compute a solution $d^k \in \mathbb{R}^n$ of the equation
$$(H_k^T H_k + \sigma_k I)d = -H_k^T G(x^k) + r^k \tag{6}$$
for a suitable vector $r^k \in \mathbb{R}^n$.
2. Let $x^{k+1} = x^k + d^k, k = k + 1$ and go to Step 1.

The vector r^k is, again, the residual. The local properties of the Generalized Inexact Levenberg-Marquardt Method are also very similar to its smooth counterpart, and have been established in [10].

Theorem 2.3 *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth and let $x^* \in \mathbb{R}^n$ be a BD-regular solution of $G(x) = 0$. Suppose that $\{x^k\}$ is a sequence generated by the Generalized Inexact Levenberg-Marquardt Method. Then the following statements hold:*

(a) There exist constants $\varepsilon > 0, \bar{\eta} > 0$ and $\bar{\sigma} > 0$ such that, if $\|x^0 - x^*\| \leq \varepsilon, \|r^k\| \leq \bar{\eta}\|H_k^T G(x^k)\|$ and $\sigma_k \leq \bar{\sigma}$ for all k, then the sequence $\{x^k\}$ converges Q-linearly to x^*.

(b) If, in addition, $\|r^k\| = o(\|H_k^T G(x^k)\|)$ and $\sigma_k \to 0$, then $\{x^k\} \to x^*$ Q-superlinearly.

(c) If, in addition, G is strongly semismooth, and if $\|r^k\| = O(\|H_k^T G(x^k)\|^2)$ and $\sigma_k = O(\|H_k^T G(x^k)\|)$, then $\{x^k\} \to x^*$ Q-quadratically.

2.2 On b- and R-Regularity

Given a solution $x^* \in \mathbb{R}^n$ of NCP(F), let us introduce the three index sets

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

The following two regularity concepts play an important role in the (theoretical and numerical) analysis of complementarity problems.

Definition 2.1 *A solution $x^* \in \mathbb{R}^n$ of NCP(F) is called*

(a) b-regular *if the principal submatrices $F'(x^*)_{\alpha\cup\delta,\alpha\cup\delta}$ are nonsingular for all subsets δ such that $\emptyset \subseteq \delta \subseteq \beta$;*

(b) R-regular *if the submatrix $F'(x^*)_{\alpha\alpha}$ is nonsingular and the Schur-complement of this matrix in*

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

is a P-matrix.

We recall that the above mentioned Schur-complement is defined by

$$F'(x^*)_{\beta\beta} - F'(x^*)_{\beta\alpha} F'(x^*)_{\alpha\alpha}^{-1} F'(x^*)_{\alpha\beta} \in \mathbb{R}^{|\beta|\times|\beta|}. \tag{7}$$

Obviously, there is no difference between b- and R-regularity at a nondegenerate solution of NCP(F). In general, however, it is known that b-regularity is a weaker assumption than R-regularity, see also Example 2.1 at the end of the next subsection. Here, we want to give a complete relationship between b- and R-regularity. To this end, we first prove a simple result.

Lemma 2.4 *Let $M \in \mathbb{R}^{n\times n}$ and $\sigma \subseteq \{1,\dots,n\}$ be such that the submatrix $M_{\sigma\sigma}$ is nonsingular. Then the principal submatrices of the Schur-complement $M/M_{\sigma\sigma}$ are given by $M_{\sigma\cup\nu,\sigma\cup\nu}/M_{\sigma\sigma}$, where ν is any subset such that $\emptyset \subseteq \nu \subseteq \bar{\sigma}$ and $\bar{\sigma} := \{1,\dots,n\} \setminus \sigma$ denotes the complementary subset of σ.*

Proof. For any subset ν such that $\emptyset \subseteq \nu \subseteq \bar{\sigma}$, we obtain

$$\begin{aligned}
[M/M_{\sigma\sigma}]_{\nu\nu} &= [M_{\bar{\sigma}\bar{\sigma}} - M_{\bar{\sigma}\sigma} M_{\sigma\sigma}^{-1} M_{\sigma\bar{\sigma}}]_{\nu\nu} \\
&= M_{\bar{\sigma}\cap\nu,\bar{\sigma}\cap\nu} - M_{\bar{\sigma}\cap\nu,\sigma} M_{\sigma\sigma}^{-1} M_{\sigma,\bar{\sigma}\cap\nu} \\
&= M_{\nu\nu} - M_{\nu\sigma} M_{\sigma\sigma}^{-1} M_{\sigma\nu} \\
&= M_{\sigma\cup\nu,\sigma\cup\nu}/M_{\sigma\sigma},
\end{aligned}$$

where the first and last equality follow directly from the definition of the Schur-complement and the third equation is a consequence of the fact that $\bar{\sigma} \cap \nu = \nu$. $\square$

We are now in the position to prove the main result of this subsection.

Theorem 2.5 *Let $x^* \in \mathbb{R}^n$ be a solution of NCP(F). Then the following statements are equivalent:*

(a) x^* is an R-regular solution of NCP(F).

(b) x^* is a b-regular solution of NCP(F) and the Schur-complement (7) is a P-matrix.

(c) x^* is a b-regular solution of NCP(F) and the Schur-complement (7) is a P_0-matrix.

Proof. For simplicity, let us write $M := F'(x^*)$.

(1) First assume that x^* is an R-regular solution. Then the Schur-complement (7) is a P-matrix by definition. The fact that $M_{\alpha\cup\delta,\alpha\cup\delta}$ is nonsingular for all δ with $\emptyset \subseteq \delta \subseteq \beta$ follows from the assumed R-regularity and [30, Lemma 1].

(2) Obviously, part (b) implies part (c) since any P-matrix is a P_0-matrix.

(3) Now assume that x^* is a b-regular solution of NCP(F) and that the Schur-complement $S := M_{\alpha\cup\beta,\alpha\cup\beta}/M_{\alpha\alpha}$ is a P_0-matrix. We have to show that this Schur-complement is actually a P-matrix, i.e., that $det(S_{\delta\delta}) > 0$ for all subsets $\emptyset \subseteq \delta \subseteq \beta$. Since, by the assumed P_0-property, we already know that $det(S_{\delta\delta}) \geq 0$, it suffices to show that $det(S_{\delta\delta}) \neq 0$ for all $\emptyset \subseteq \delta \subseteq \beta$. In view of Lemma 2.4, we have

$$S_{\delta\delta} = M_{\alpha\cup\delta,\alpha\cup\delta}/M_{\alpha\alpha} \text{ for any } \emptyset \subseteq \delta \subseteq \beta.$$

Hence, from a well-known result on the determinant of a Schur-complement (see, e.g., [5, Proposition 2.3.5]), we get

$$det(S_{\delta\delta}) = det(M_{\alpha\cup\delta,\alpha\cup\delta}/M_{\alpha\alpha}) = det(M_{\alpha\cup\delta,\alpha\cup\delta})/det(M_{\alpha\alpha}) \neq 0,$$

where the last part follows from the fact that $M_{\alpha\cup\delta,\alpha\cup\delta}$ is nonsingular for any $\emptyset \subseteq \delta \subseteq \beta$ by the assumed b-regularity of the solution x^*. Hence x^* is an R-regular solution of NCP(F). $\square$

The above result immediately implies the following corollary.

Corollary 2.6 *Let $x^* \in \mathbb{R}^n$ be a solution of NCP(F) such that the Jacobian $F'(x^*)$ is a P_0-matrix. Then x^* is an R-regular solution if and only if it is a b-regular solution.*

Proof. The statement follows directly from Theorem 2.5 and the fact that the Schur-complement of a nonsingular submatrix of a P_0-matrix is again a P_0-matrix, see [3, Lemma 2.3]. $\square$

If $F : \mathbb{R}^n \to \mathbb{R}^n$ is a P_0-function, then all Jacobian matrices are P_0-matrices by [28, Theorem 5.8]. Hence there is no difference between R- and b-regular solutions for P_0-function complementarity problems. In particular, there is no difference between these two regularity concepts for monotone problems.

2.3 The Operators Φ_F and Φ_P

In this section we recall some basic (semi)smoothness properties of two operators which are often used in order to reformulate the nonlinear complementarity problem as a nonsmooth system of equations, namely the Minimum and the Fischer operator. The Fischer operator has already been defined in the introduction. The Minimum operator is similarly defined by

$$\Phi_P(x) := \begin{pmatrix} \min\{x_1, F_1(x)\} \\ \vdots \\ \min\{x_n, F_n(x)\} \end{pmatrix};$$

here we use the subscript "P" for the Minimum operator in order to stress the importance of Pang's [29, 30] seminal work in this area.

We believe that $\Phi_F(x) = 0$ and $\Phi_P(x) = 0$ are the most used reformulations of the complementarity problem as system of equations, and we shall use them extensively in this paper. In view of the results of the previous sections we are obviously interested in the semismoothness properties of these reformulations. The first result that can easily be established is that, if F is (strongly) semismooth, then also Φ_F and Φ_P are strongly semismooth, see [14, 20] for details. In particular, we have the following

Proposition 2.7 (a) *If F is a C^1-function, then Φ_F and Φ_P are semismooth.*

(b) *If F is an LC^1-function, then Φ_F and Φ_P are strongly semismooth.*

Another point which deserves attention is the structure of the B-subdifferentials of Φ_F and Φ_P and the conditions which guarantee BD-regularity at a solution. The following result was shown in [11].

Proposition 2.8 *Suppose that F is a C^1-function and x^* is an R-regular solution of NCP(F). Then x^* is a BD-regular solution of the system $\Phi_F(x) = 0$.*

A crucial property in the design of globally convergent algorithms, that has been extensively exploited in recent years, is that the corresponding merit function Ψ_F (see (2)) is continuously differentiable [11, 16, 19]).

Proposition 2.9 *Suppose that F is a C^1-function. Then the merit function Ψ_F is continuously differentiable with gradient $\nabla\Psi_F(x) = H^T\Phi_F(x)$ for an arbitrary matrix $H \in \partial\Phi_F(x)$.*

In [6] it has been shown how to calculate elements in the B-subdifferential of Φ_F, and this, in view of Proposition 2.9, also allows to easily evaluate the gradient of Ψ_F.

We now consider analogous issues for Φ_P. To this end let x^* be a solution of NCP(F) and let $H \in \partial_B\Phi_P(x^*)$ be arbitrarily chosen. Let $H_{i.}$ denote the ith row of this matrix. Furthermore, let α, β and γ denote the index sets defined in the previous subsection. From the very definition of the operator Φ_P and the B-subdifferential $\partial_B\Phi_P(x^*)$, we have the following:

(a) If $i \in \alpha$, then $H_{i.} = \nabla F_i(x^*)^T$.

(b) If $i \in \gamma$, then $H_{i.} = e_i^T$.

(c) If $i \in \beta$, then $H_{i.} = \nabla F_i(x^*)^T$ or $H_{i.} = e_i^T$.

Hence there is an index set δ such that $\emptyset \subseteq \delta \subseteq \beta$ and

$$H = \begin{pmatrix} F'(x^*)_{\alpha\cup\delta,\alpha\cup\delta} & F'(x^*)_{\alpha\cup\delta,\gamma\cup\bar{\delta}} \\ 0_{\gamma\cup\bar{\delta},\alpha\cup\delta} & I_{\gamma\cup\bar{\delta},\gamma\cup\bar{\delta}} \end{pmatrix}, \tag{8}$$

where $\bar{\delta} := \beta \setminus \delta$ denotes the complementary subset of δ in β. By using this expression of $\partial_B \Phi_P(x^*)$ and the very definition of b-regularity, it is then immediate to prove the following proposition.

Proposition 2.10 *Suppose that F is a C^1-function and x^* is a b-regular solution of NCP(F). Then x^* is a BD-regular solution of the system $\Phi_P(x) = 0$.*

It is interesting to note that Φ_P is BD-regular at a solution of the complementarity problem under an assumption which is weaker than that needed to establish an analogous result for Φ_F (even is we saw in the previous section that these two assumptions coincide in the case of P_0-functions). One may wonder whether this gap is only due to a lack in the analysis of the properties of Φ_F; the following example shows that the gap is actually intrinsic.

Example 2.1 Let $n = 2$ and $F : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$F(x) := \begin{pmatrix} -x_1 + x_2 \\ -x_2 \end{pmatrix}.$$

Then it is easy to see that the complementarity problem NCP(F) has the unique solution $x^* := (0, 0)$ which is b-regular (since $F'(x^*)$ is a nondegenerate matrix) but not R-regular. Now consider the sequence $\{x^k\}$ defined by

$$x^k := \left(\frac{1}{k}, \frac{2}{k}\right)^T.$$

Then $x^k \to x^*$ and Φ_F is continuously differentiable at all x^k with Jacobian

$$\Phi_F'(x^k) = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} - 1 \\ 0 & \sqrt{2} \end{pmatrix}$$

for all k. Obviously, the sequence $\{\Phi_F'(x^k)\}$ converges to the singular matrix

$$\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} - 1 \\ 0 & \sqrt{2} \end{pmatrix} \in \partial_B \Phi_F(x^*),$$

i.e., x^* is not a BD-regular solution of the system $\Phi_F(x) = 0$.

On the other hand it is very important to note that, contrary to Φ_F, the natural merit function associated to the operator Φ_P is not differentiable everywhere in general, making the design of globally convergent algorithms based on Φ_P fairly cumbersome.

3 Algorithm and Convergence

In this section we propose a general algorithmic scheme which is similar to the algorithms considered, e.g., in [19, 6]. The only difference is that, instead of using a specific direction, we use a generic direction on which we impose certain assumptions. In the next section, we shall illustrate how directions satisfying these assumptions can be calculated. This allows us to analyze in a unified way different algorithms, corresponding to different directions.

In the sequel, in order to prove global convergence, we shall always assume that the search direction used in the algorithmic scheme satisfies Assumption 1 below. Assumptions 2_{sup}, 2_{quad} and 2_{fin} will be invoked only to establish convergence rates.

Assumption 1

(a) $d^k = 0 \implies \nabla \Psi_F(x^k) = 0$.

(b) If $x^k \to x^*$ and $d^k \to 0$, then $\nabla \Psi_F(x^*) = 0$.

Assumption $\mathbf{2}_{sup}$

Let x^* be a solution of NCP(F) and assume that $\{x^k\}$ is a sequence converging to x^*. Then

$$\lim_{k\to\infty} \frac{\|x^k + d^k - x^*\|}{\|x^k - x^*\|} = 0.$$

Assumption $\mathbf{2}_{quad}$

Let x^* be a solution of NCP(F) and assume that $\{x^k\}$ is a sequence converging to x^*. Then

$$\limsup_{k\to\infty} \frac{\|x^k + d^k - x^*\|}{\|x^k - x^*\|^2} < +\infty.$$

Assumption $\mathbf{2}_{fin}$

Let x^* be a solution of NCP(F). There exists a neighborhood Ω of x^* such that, if x^k belongs to Ω, then

$$x^* = x^k + d^k.$$

Note that in our three different Assumptions 2, no connection is assumed between the way the sequences $\{x^k\}$ and $\{d^k\}$ are generated. The three Assumptions 2_{sup}, 2_{quad} and 2_{fin} say that the direction has to be locally superlinearly convergent, locally quadratically convergent or has a finite termination property, respectively.

In what follows, a general scheme of an algorithm for the solution of the complementarity problem using the merit function Ψ_F is given.

General Line Search Algorithm

0. Choose $x^0 \in \mathbb{R}^n$, $\rho > 0$, $s > 1$, $\beta \in (0, 1/2)$, $\sigma \in (0, 1)$. Set $k := 0$.
1. If x^k satisfies a suitable termination criterion: STOP.
2. Calculate a direction d^k satisfying Assumption 1 and set $p^k = d^k$.
3. If
$$\Psi_F(x^k + p^k) \leq \sigma \Psi_F(x^k), \tag{9}$$
set $x^{k+1} := x^k + p^k$, set $k := k + 1$ and go to Step 1.
4. If p^k does not satisfy the following test
$$\nabla \Psi_F(x^k)^T p^k \leq -\rho \|p^k\|^s, \tag{10}$$
set $p^k = -\nabla \Psi_F(x^k)$. Find the smallest $i^k \in \{0, 1, 2, \ldots\}$ such that
$$\Psi_F(x^k + 2^{-i^k} p^k) \leq \Psi_F(x^k) + \beta 2^{-i^k} \nabla \Psi_F(x^k)^T p^k, \tag{11}$$
set $x^{k+1} := x^k + 2^{-i^k} p^k$, set $k := k + 1$ and go to Step 1.

The algorithm itself, and the corresponding convergence analysis, are heavily based on the continuous differentiability of Ψ_F described in Section 2.3.

In the sequel, in order to facilitate the statement of the convergence properties of the algorithm, we shall always assume that the termination criterion at Step 1 is never satisfied, so that an infinite sequence of points is generated

Theorem 3.1 *If Assumption 1 holds, then every limit point of a sequence $\{x^k\}$ generated by the General Line Search Algorithm is a stationary point of Ψ_F.*

Proof. The proof is by contradiction. Suppose, renumbering if necessary, that $\{x^k\} \to x^*$ and that $\nabla \Psi_F(x^*) \neq 0$; then we can assume without loss of generality that the test (9) is never passed and that
$$0 < \delta \leq \|p^k\| \leq D \tag{12}$$
for suitable constants $0 < \delta \leq D$. In fact, if the test (9) is satisfied infinitely many times, this would imply, recalling that at each step $\Psi_F(x^{k+1}) \leq \Psi_F(x^k)$, that $\{\Psi_F(x^k)\} \to 0$, so

that x^* is a global minimum point of Ψ_F and hence $\nabla\Psi_F(x^*) = 0$. On the other hand if, for some subsequence K, $\{\|p^k\|\}_K \to 0$, we have that $\nabla\Psi_F(x^*) = 0$ by Assumption 1 (b), while $\{\|p^k\|\}$ cannot be unbounded because, taking into account that $\nabla\Psi_F(x^k)$ is bounded and $s > 1$, this would contradict (10).

Then, since at each iteration (11) holds and Ψ_F is bounded from below, we have that $\{\Psi_F(x^{k+1}) - \Psi_F(x^k)\} \to 0$ which implies, by the linesearch test,

$$\{2^{-i^k}\nabla\Psi_F(x^k)^T d^k\} \to 0. \tag{13}$$

We want to show that 2^{-i^k} is bounded away from 0. Suppose the contrary. Then, subsequencing if necessary, we have that $\{2^{-i^k}\} \to 0$ so that at each iteration the stepsize is reduced at least once and (11) gives

$$\frac{\Psi_F(x^k + 2^{-(i^k-1)}p^k) - \Psi_F(x^k)}{2^{-(i^k-1)}} > \beta\nabla\Psi_F(x^k)^T p^k. \tag{14}$$

By (12) we can assume, subsequencing if necessary, that $\{p^k\} \to p^* \neq 0$, so that, passing to the limit in (14), we get

$$\nabla\Psi_F(x^*)^T p^* \geq \beta\nabla\Psi_F(x^*)^T p^*$$

and therefore

$$\nabla\Psi_F(x^*)^T p^* \geq 0 \tag{15}$$

since $\beta \in (0,1)$. On the other hand, we also have, by (10), that $\nabla\Psi_F(x^*)^T p^* \leq -\rho\|p^*\|^s < 0$, which contradicts (15); hence 2^{-i^k} is bounded away from 0. But then (13) and (10) imply that $\{p^k\} \to 0$ so that $\nabla\Psi_F(x^*) = 0$ by Assumption 1 (b). □

Assumptions under which a stationary point of the function Ψ_F is a solution of the complementarity problem are given in [6] to which we refer the interested reader. Here we only note that if F is a P_0-function then every stationary point of Ψ_F is a solution of the complementarity problem. This condition is probably the weakest known condition guaranteeing the correspondence between stationary points of a merit function and solutions of the complementarity problem, and is one of the main motivations for using Ψ_F instead of other merit functions in our General Line Search Algorithm.

In the sequel we examine the convergence rate of the algorithm.

Theorem 3.2 *Suppose that x^* is a b-regular solution of NCP(F), that $\{x^k\}$ converges to x^* and that Assumption 2_{sup} is satisfied in x^*. Then, eventually $p^k = d^k$, $x^{k+1} = x^k + d^k$ and $\{x^k\}$ converges Q-superlinearly to x^*.*

The proof uses the following two lemmas. The first one was shown in [11] under a slightly stronger assumption. However, it is easy to see that the proof goes through with the weaker assumption stated here. The second lemma readily follows from [34, Lemma 3.1].

Lemma 3.3 *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth and $x^* \in \mathbb{R}^n$ be a BD-regular solution of the system $G(x) = 0$. Suppose that there are two sequences $\{x^k\}$ and $\{d^k\}$ such that*

$$x^k \to x^* \text{ and } \lim_{k\to\infty} \frac{\|x^k + d^k - x^*\|}{\|x^k - x^*\|} = 0.$$

Then

$$\lim_{k\to\infty} \frac{\|G(x^k + d^k)\|}{\|G(x^k)\|} = 0.$$

Lemma 3.4 *There exist constants $c_1 > 0$ and $c_2 > 0$ such that*

$$c_1\Psi_P(x) \le \Psi_F(x) \le c_2\Psi_P(x)$$

holds for all $x \in \mathbb{R}^n$, where $\Psi_P(x) := \frac{1}{2}\|\Phi_P(x)\|^2$ denotes the merit function belonging to the Minimum operator Φ_P.

Proof of Theorem 3.2. By Lemmas 3.3, 3.4, Proposition 2.10 and Assumption 2_{sup}, the b-regularity condition implies that eventually the test (9) is satisfied. So the assertion follows by the instructions of the General Line Search Algorithm and Assumption 2_{sup}. $\square$

The following two results can be easily proved along the same lines used for the proof of Theorem 3.2.

Theorem 3.5 *Suppose that x^* is a b-regular solution of NCP(F), that $\{x^k\}$ converges to x^* and that Assumption 2_{quad} is satisfied in x^*. Then, eventually $p^k = d^k$, $x^{k+1} = x^k + d^k$ and $\{x^k\}$ converges Q-quadratically to x^*.*

Theorem 3.6 *Suppose that x^* is a b-regular solution of NCP(F), that $\{x^k\}$ converges to x^* and that Assumption 2_{fin} is satisfied in x^*. Then there exists an index k_0 such that $x^{k_0} = x^*$.*

4 Search Directions

By the results of the previous section we see that we can define algorithms which enjoy favourable properties if we are able to define search directions which satisfy Assumptions 1 and 2_{sup} (or 2_{quad} or 2_{fin}). Obviously, from a theoretical point of view, we wish these assumptions to hold under conditions on the complementarity problem which are as weak as possible. In this section we consider several possible directions d^k and compare their theoretical properties. A numerical comparison is the subject of Section 5.

4.1 The Fischer-Qi Direction

The first direction we consider is basically the semismooth Newton direction for the system $\Phi_F(x) = 0$.

$$d^k = \begin{cases} \text{solution of } H_k d = -\Phi_F(x^k), \ \text{with } H_k \in \partial_B\Phi_F(x^k), & \text{if system solvable} \\ -\nabla\Psi_F(x^k) & \text{otherwise.} \end{cases} \tag{16}$$

Theorem 4.1 *(a) The direction defined by (16) satisfies Assumption 1.*

(b) If x^* is an R-regular solution of NCP(F), then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function, then also Assumption 2_{quad} holds.

Proof. (a) The proof of this point is obvious if $d^k = -\nabla\Psi_F(x^k)$. So consider the case in which d^k is the solution of $H_k d = -\Phi_F(x^k)$. If $x^k \to x^*$ (in particular, if $x^k = x^*$ for every k) and if $d^k \to 0$, we have, taking into account the boundedness of the B-subdifferential on bounded sets, that $\Phi_F(x^*) = 0$. But then x^* is a global minimum of Ψ_F so that $\nabla\Psi_F(x^*) = 0$.

(b) If x^* is an R-regular solution of NCP(F) then, by Proposition 2.8, x^* is a BD-regular solution of the system $\Phi_F(x) = 0$, which is semismooth by Proposition 2.7 (a). Hence Assumption 2_{sup} holds because of Theorem 2.1. If, in addition, F is an LC^1-function, then Φ_F is strongly semismooth by Proposition 2.7 (b), and, once again, the assertion follows from Theorem 2.1. $\square$

4.2 The Inexact Fischer-Qi Direction

This is nothing else but the previous direction in which, however, the linear system is solved inexactly.

$$d^k = \begin{cases} \text{solution of } H_k d = -\Phi_F(x^k) + r^k, \text{ with } H_k \in \partial_B \Phi_F(x^k), & \text{if system solvable} \\ -\nabla\Psi_F(x^k) & \text{otherwise.} \end{cases} \tag{17}$$

Theorem 4.2 *Assume that the sequence of residuals $\{r^k\}$ satisfies $\|r^k\| \leq \eta_k \|\Phi_F(x^k)\|$, where $\{\eta_k\}$ is a sequence of numbers such that, for every k, $0 \leq \eta_k \leq \bar{\eta} < 1$. Then:*

(a) The direction defined by (17) satisfies Assumption 1.

(b) If x^* is an R-regular solution of NCP(F) and $\|r^k\| = o(\|\Phi_F(x^k)\|)$, then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function and $\|r^k\| = O(\|\Phi_F(x^k)\|^2)$, then also Assumption 2_{quad} holds.

Proof. (a) The proof of this point is obvious if $d^k = -\nabla\Psi_F(x^k)$. So consider the case in which d^k is the solution of $H_k d = -\Phi_F(x^k) + r^k$ for a suitable residual vector $r^k \in \mathbb{R}^n$. If $x^k \to x^*$ (in particular, if $x^k = x^*$ for every k), we have, renumbering if necessary and taking into account the boundedness of the sequence $\{r^k\}$, that $r^k \to r^*$ for some vector $r^* \in \mathbb{R}^n$. Hence, if $d^k \to 0$, we also have, taking into account the boundedness of the B-subdifferential on bounded sets, that $\Phi_F(x^*) = r^*$. In view of our assumptions, however, we have $\|r^*\| \leq \bar{\eta}\|\Phi_F(x^*)\|$. Since $\bar{\eta} < 1$, this is only possible if $\Phi_F(x^*) = r^* = 0$. In turn, this implies that x^* is a global minimum of Ψ_F so that $\nabla\Psi_F(x^*) = 0$.

(b) If x^* is an R-regular solution of NCP(F) then, by Proposition 2.8, x^* is a BD-regular solution of the system $\Phi_F(x) = 0$, which is semismooth by Proposition 2.7 (a). Hence

Assumption 2_{sup} holds because of Theorem 2.2 (a) and (b). If, in addition, F is an LC^1-function, then Φ_F is strongly semismooth by Proposition 2.7 (b), and the assertion follows from Theorem 2.2 (c). □

4.3 The Inexact LM Fischer-Qi Direction

The direction considered in this subsection is an inexact perturbed Levenberg-Marquardt version of direction (16). It was already considered in [10] and is defined as follows:

$$\boxed{(H_k^T H_k + \sigma_k I)d = -H_k^T \Phi_F(x^k) + r^k, \qquad (18)}$$

where σ_k is a nonnegative number and r^k is the residual, which measures the inaccuracy with which the system is solved. Recall that $(H_k)^T \Phi_F(x^k) = \nabla \Psi_F(x^k)$ and note that the linear system (18) always admits an exact (and hence inexact) solution since, for $\sigma_k > 0$, the matrix on the left-hand side is positive definite, whereas for $\sigma_k = 0$ we try to solve the normal equations of the system $H_k d = -\Phi_F(x^k)$.

Theorem 4.3 *Assume that the sequence of residuals $\{r^k\}$ satisfies $\|r^k\| \leq \eta_k \|\nabla \Psi_F(x^k)\|$, where $\{\eta_k\}$ is a sequence of numbers such that, for every k, $0 \leq \eta_k \leq \bar{\eta} < 1$. Assume also that the sequence $\{\sigma_k\}$ is bounded from above. Then:*

(a) The direction defined by (17) satisfies Assumption 1.

(b) If x^* is an R-regular solution of NCP(F), $\|r^k\| = o(\|\nabla \Psi_F(x^k)\|)$ and $\{\sigma_k\} \to 0$, then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function, $\|r^k\| = O(\|\nabla \Psi_F(x^k)\|^2)$ and $\sigma_k = O(\|\nabla \Psi_F(x^k)\|)$, then also Assumption 2_{quad} holds.

Proof. (a) If $x^k \to x^*$ (in particular, if $x^k = x^*$ for every k), we have, renumbering if necessary and taking into account the boundedness of the sequence $\{r^k\}$, that $r^k \to r^*$ for some vector $r^* \in \mathbb{R}^n$. Hence, if $d^k \to 0$, we also have, using the boundedness of the B-subdifferential on bounded sets and the assumed boundedness of the sequence $\{\sigma_k\}$, that $\nabla \Psi_F(x^*) = r^*$. But, since $\|r^*\| \leq \bar{\eta} \|\nabla \Psi_F(x^*)\|$ with $\bar{\eta} < 1$ in view of our assumptions, this is only possible if $\nabla \Psi_F(x^*) = r^* = 0$.

(b) If x^* is an R-regular solution of NCP(F), then, by Proposition 2.8, x^* is a BD-regular solution of the system $\Phi_F(x) = 0$, which is semismooth by Proposition 2.7 (a). Hence Assumption 2_{sup} holds because of Theorem 2.3 (a) and (b). If, in addition, F is an LC^1-function, then Φ_F is strongly semismooth by Proposition 2.7 (b), and the assertion follows from Theorem 2.3 (c). □

4.4 The Pang-Qi Direction

This search direction is basically the semismooth Newton direction for the system $\Phi_P(x) = 0$.

$$d^k = \begin{cases} \text{solution of } H_k d = -\Phi_P(x^k), \ \text{ with } H_k \in \partial_B \Phi_P(x^k), & \text{if system solvable} \\ -\nabla\Psi_F(x^k) & \text{otherwise.} \end{cases} \tag{19}$$

Theorem 4.4 *(a) The direction defined by (19) satisfies Assumption 1.*

(b) If x^* is a b-regular solution of NCP(F), then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function, then also Assumption 2_{quad} holds.

(c) If F is affine and x^* is a b-regular solution, then Assumption 2_{fin} holds.

Proof. (a) The proof of this point is obvious if $d^k = -\nabla\Psi_F(x^k)$. So consider the case in which d^k is the solution of $H_k d = -\Phi_P(x^k)$. If $x^k \to x^*$ (in particular, if $x^k = x^*$ for every k) and if $d^k \to 0$, we have, taking into account the boundedness of the B-subdifferential on bounded sets, that $\Phi_P(x^*) = 0$. But then x^* is a solution of NCP(F) so that $\nabla\Psi_F(x^*) = 0$.

(b) If x^* is a b-regular solution of NCP(F) then, by Proposition 2.10, x^* is a BD-regular solution of the system $\Phi_P(x) = 0$, which is semismooth by Proposition 2.7 (a). Hence Assumption 2_{sup} holds because of Theorem 2.1 (a) and (b). If, in addition, F is an LC^1-function, then Φ_P is strongly semismooth by Proposition 2.7 (b), and the assertion follows from Theorem 2.1 (c).

(c) This follows immediately from [15, Theorem 4], Proposition 2.10 and the upper semicontinuity of the B-subdifferential. □

Point (c) in the theorem above also appears to be strongly related to the finite termination result in [11].

From a computational point of view, it may be interesting to observe that the system $H_k d = -\Phi_P(x^k)$, used in (19), is very structured, and this can be exploited numerically. In fact, defining the index sets

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

and recalling the definition of Φ_P, we see, similarly to (8), that we can write the system $H_k d = -\Phi_P(x^k)$ as

$$\begin{pmatrix} F'(x^k)_{\alpha\cup\delta,\alpha\cup\delta} & F'(x^k)_{\alpha\cup\delta,\gamma\cup\bar\delta} \\ 0_{\gamma\cup\bar\delta,\alpha\cup\delta} & I_{\gamma\cup\bar\delta,\gamma\cup\bar\delta} \end{pmatrix} \begin{pmatrix} d_{\alpha\cup\delta} \\ d_{\gamma\cup\bar\delta} \end{pmatrix} = - \begin{pmatrix} F_{\alpha\cup\delta}(x^k) \\ x^k_{\gamma\cup\bar\delta} \end{pmatrix} \tag{20}$$

for some index set δ with $\emptyset \subseteq \delta \subseteq \beta$ and $\bar\delta := \beta \setminus \delta$. This readily gives $d_{\gamma\cup\bar\delta} = -x^k_{\gamma\cup\bar\delta}$. It is then obvious that it is not very sensible to apply a linear solver to the whole linear system in order to get a solution of (20). Is is more convenient to set $d^k_{\gamma\cup\bar\delta} = -x^k_{\gamma\cup\bar\delta}$, substitute this value in the first set of equations in (20) and solve the resulting *reduced* linear system, which reads

$$F'_{\alpha\cup\delta,\alpha\cup\delta}(x^k) d_{\alpha\cup\delta} = -F_{\alpha\cup\delta}(x^k) + F'(x^k)_{\alpha\cup\delta,\gamma\cup\bar\delta} x^k_{\gamma\cup\bar\delta}. \tag{21}$$

4.5 The Inexact Pang-Qi Direction

This is nothing else but the previous direction in which, however, the linear system is solved approximately.

$$d^k = \begin{cases} \text{solution of } H_k d = -\Phi_P(x^k) + r^k, \;\; \text{with } H_k \in \partial_B \Phi_P(x^k), & \text{if system solvable} \\ -\nabla \Psi_F(x^k) & \text{otherwise.} \end{cases} \tag{22}$$

The proof of the following theorem, which we omit, can be carried out exactly along the same lines used in Theorem 4.3, using Proposition 2.10 instead of Proposition 2.8.

Theorem 4.5 *Assume that the sequence of residuals $\{r^k\}$ satisfies $\|r^k\| \leq \eta_k \|\Phi_P(x^k)\|$, where $\{\eta_k\}$ is a sequence of numbers such that, for every k, $0 \leq \eta_k \leq \bar{\eta} < 1$. Then:*

(a) The direction defined by (17) satisfies Assumption 1.

(b) If x^* is a b-regular solution of NCP(F) and $\|r^k\| = o(\|\Phi_P(x^k)\|)$, then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function and $\|r^k\| = O(\|\Phi_P(x^k)\|^2)$, then also Assumption 2_{quad} holds.

Obviously, also in this case, similarly to what has been seen in the previous section, it is possible to approximately solve just the reduced system (21).

4.6 The Inexact LM Pang-Qi Direction

The direction considered in this subsection is an inexact perturbed Levenberg-Marquardt version of direction (19).

$$(H_k^T H_k + \sigma_k I)d = -H_k^T \Phi_P(x^k) + r^k, \tag{23}$$

where $H_k \in \partial_B \Phi_P(x^k), \sigma_k$ is a nonnegative number and r^k is the residual which measures the inaccuracy with which the system is solved.

Theorem 4.6 *Assume that the sequence of residuals $\{r^k\}$ satisfies $\|r^k\| \leq \eta_k \|H_k^T \Phi_P(x^k)\|$, where $\{\eta_k\}$ is a sequence of numbers such that, for every k, $0 \leq \eta_k \leq \bar{\eta} < 1$, and assume that the sequence $\{\sigma_k\}$ is bounded from above. Assume also that for every k the matrix $F'_{\alpha\cup\delta,\alpha\cup\delta}(x^k)$ is nonsingular (where α and δ denote the index sets used in Subsection 4.4). Then:*

(a) The direction defined by (17) satisfies Assumption 1.

(b) If x^* is a b-regular solution of NCP(F), $\|r^k\| = o(\|H_k^T \Phi_P(x^k)\|)$ and $\{\sigma_k\} \to 0$, then Assumption 2_{sup} holds. Furthermore, if F is an LC^1-function, $\|r^k\| = O(\|H_k^T \Phi_P(x^k)\|^2)$ and $\sigma_k = O(\|H_k^T \Phi_P(x^k)\})$, then also Assumption 2_{quad} holds.

Proof. (a) If $x^k \to x^*$ (in particular, if $x^k = x^*$ for every k), we have, renumbering if necessary and taking into account the boundedness of the sequence $\{r^k\}$, that $r^k \to r^*$ for some vector $r^* \in \mathbb{R}^n$. Hence, if $d^k \to 0$, we also have, using the boundedness of the B-subdifferential on bounded sets and the boundedness of the sequence $\{\sigma_k\}$, that $H_k^T \Phi_P(x^*) = r^*$. In view of our assumptions, however, we also have $\|r^*\| \leq \bar{\eta} \|H_k^T \Phi_P(x^*)\|$ with $\bar{\eta} < 1$. But this is only possible if $H_k^T \Phi_P(x^*) = r^* = 0$. In turn, recalling (20) and the nonsingularity assumption, this implies $\Phi_P(x^*) = 0$. Hence x^* is a solution of NCP(F). This implies $\nabla \Psi_F(x^*) = 0$.

(b) If x^* is a b-regular solution of NCP(F) then, by Proposition 2.10, x^* is a BD-regular solution of the system $\Phi_P(x) = 0$, which is semismooth by Proposition 2.7 (a). Hence Assumption 2_{sup} holds because of Theorem 2.3 (a) and (b). If, in addition, F is an LC^1-function, then Φ_P is strongly semismooth by Proposition 2.7 (b), and the assertion follows from Theorem 2.3 (c). $\square$

Also in this case, as seen in the previous two cases, things can be arranged so that only a reduced system has to be solved approximately at each iteration. Note also that, in order to satisfy Assumption 1, we had to impose a nonsingularity assumption in Theorem 4.6. This assumption, which is satisfied in a sufficiently small neighbourhood of a b-regular solution of NCP(F), was not necessary for any of the other search directions discussed in this section.

4.7 Comparison of the Theoretical Characteristics of the Directions

From a theoretical point of view, the Pang-Qi directions (exact, inexact and Levenberg-Marquardt) need weaker assumptions than the corresponding Fischer-Qi directions in order to satisfy Assumption 2_{sup} or Assumption 2_{quad}, namely b-regularity instead of R-regularity. However it should also be added that, as shown in Section 2.3, these two conditions are equivalent for the class of P_0-functions. Another advantage of the Pang-Qi directions is that they only need the (possibly inexact) solution of a linear system of a dimension that, asymptotically, is equal to the number of positive variables at the solution. This contrasts favourably with the Fischer-Qi directions that always need the solution of an n-dimensional square system. Finally, it should also be noted that there is only one direction for which Assumption 2_{fin} (i.e., finite termination for linear complementarity problems) can be proved: the exact Pang-Qi direction.

The Fischer-Qi directions seem to have only one, but important, advantage over the corresponding Pang-Qi directions: We saw that, in the General Line Search Algorithm, we first try to use a Newton-like direction, but, if neither (9) nor (10) are satisfied, we employ $-\nabla \Psi_F(x^k)$ as search direction. Obviously, it does not appear to be very desirable to have to resort to the gradient often. From this point of view we may note that it is easy to see that the Fischer-Qi directions are always directions of descent for the merit function Ψ_F, see, e.g., [11, 6], while we cannot expect a similar property to hold for the Pang-Qi directions since these directions are calculated on the basis of a reformulation of the complementarity problem based on Φ_P and not on Φ_F. Thus, a priori, we should expect that, if we use a Pang-Qi direction in the General Line Search Algorithm, the Newton direction will be discarded more often than when we use a Fischer-Qi direction.

We conclude this section by noting that some of the algorithms obtained by using some of the directions introduced in this section in the General Line Search Algorithm have already been presented in the literature. In particular, if we use the exact Fischer-Qi direction we have the algorithm considered in [6], whereas if we use the inexact LM Fischer-Qi direction we obtain the algorithm considered in [10].

5 Numerical Results

5.1 Description of the Test Problems and Implementation Details

We tested our algorithm on test problems generated by using a technique first suggested in [17] and subsequently used, e.g., in [10]. Let $g(x) = 0$ be a (large-scale) differentiable system of nonlinear equations and let $x^* \in \mathbb{R}^n$ be defined by $x^* = (1, 0, 1, 0, \ldots)^T$. For all $i = 1, \ldots, n$ set

$$F_i(x) = \begin{cases} g_i(x) - g_i(x^*) & \text{if } i \text{ odd or } i > r, \\ g_i(x) - g_i(x^*) + 1 & \text{otherwise,} \end{cases}$$

where $r \geq 0$ is a given integer. Then x^* is a solution of the nonlinear complementarity problem NCP(F) (but not necessarily its unique solution). Note that, if r is smaller than $n-1$, the problem is degenerate at x^*. As done in [17], we used the collection of 17 large-scale problems from Lukšan [24] and the starting points indicated there. However, we did not consider problems 3 and 6, since these were never solved by any of the algorithms considered; actually we are not aware of any algorithm capable of solving these problems (see also [17]).

We considered problems with dimensions $n = 100$, $n = 1000$ and $n = 10000$. For each dimension we considered two cases: $r = n/2$ and $r = n$. The former case corresponds to a degenerate problem with $n/4$ degenerate components, while the latter case corresponds to a nondegenerate problem. Finally, besides the starting points x^0 suggested in [24], we also considered, for each problem, an additional starting point defined by

$$\tilde{x}_i^0 = \begin{cases} 10x_i^0 & \text{if } x_i^0 \neq 0, \\ 10 & \text{otherwise.} \end{cases}$$

Therefore each algorithm was tested on 90 different problems and on each of these problems two different starting points were used so that we have 180 runs for each algorithm.

In the previous section, we introduced 6 directions which, coupled with the General Line Search Algorithm, give 6 different algorithms. However, in this section we only report the results for four directions: the exact Fischer-Qi (EXFQ) direction, the exact Pang-Qi (EXPQ) direction, the Levenberg-Marquardt inexact Fischer-Qi (LMFQ) direction and the Levenberg-Marquardt inexact Pang-Qi (LMPQ) direction. We do not report results on the inexact Fischer-Qi and inexact Pang-Qi directions. In fact, in order to implement the corresponding algorithms we need an iterative solver for unsymmetric linear systems. Although many of them are available, our experiments seem to indicate that these solvers (at least if they are not combined with a sophisticated preconditioner) are not very reliable in practice (at least on our problems) leading either to extremely long running times or to failure due to the incapability of finding a sufficiently good search direction. We were genuinely surprised by these results. We think that the problem is that these solvers are usually thought for and tested on linear systems arising from the discretization of PDE problems, and they are not very reliable when used on other problems. For the other 4 methods, instead, we had no difficulty in selecting an appropriate solver. In particular, we used HARWELL routine MA50 to calculate the EXFQ and EXPQ directions and the conjugate gradient method to calculate the LMFQ and LMPQ directions.

The conjugate gradient algorithm was stopped when the norm of the residual is smaller than $(0.1/(k+1))\|\nabla\Psi_F(x^k)\|$ in the case of the LMFQ algorithm and when the residual is smaller than

$$\frac{0.1}{k+1}(H^k_{\alpha\cup\delta,\alpha\cup\delta})^T(F_{\alpha\cup\delta}(x^k) - F'(x^k)_{\alpha\cup\delta,\gamma\cup\bar{\delta}}x^k_{\gamma\cup\bar{\delta}})$$

in the case of the LMPQ algorithm. We also set a limit of 200 iterations to the conjugate gradient phase. Finally we turn to the choice of σ_k. We first note that the classical sophisticated choices indicated, e.g., in [7], are not suitable for large-scale problems. We therefore used a very simple, and yet seemingly effective strategy, already employed in [10]. If in the previous iteration the quotient $\|\nabla\Psi_F(x^{k-1})\|/\|d^{k-1}\|$ is greater than 250 and the norm of the natural residual $\|\min\{x^k, F(x^k)\}\|$ is greater than $k(0.1\sqrt{n})$ then we set $\sigma_k = 1$, otherwise we set $\sigma_k = 0$. The first test is a rough indicator that "something is going wrong" while the second test makes the possibility of the perturbation (i.e., $\sigma_k > 0$) more and more unlikely the more the process progresses, so that the final fast convergence rate is preserved.

We coded the algorithm in Fortran 77 and run it on an IBM RISC 6000/375 machine in double precision arithmetic.

The main stopping criterion is $\|\min\{x^k, F(x^k)\}\| \leq 10^{-5}\sqrt{n}$ or $\|\nabla\Psi_F(x^k)\| \leq 10^{-5}\sqrt{n}$, but we also stopped the algorithm after 100 iterations if the former stopping criterion was not met. We used a single set of parameters for all the runs of the problems, more precisely we set: $\rho = 10^{-8}$, $s = 2.1$, $\beta = 10^{-4}$ and $\sigma = 0.9$.

Finally, we also considered the option of initiating the computation by performing at most 10 iterations of a projected gradient method to minimize Ψ_F over the nonnegative orthant [1]. The rationale behind this option is to try to move in a more "promising" zone by performing a few steps of a robust and not expensive method. Similar ideas were successfully used, for example, in [8], where, however, the situation is slightly more complicated since the merit function is not differentiable. It should be remarked that in this gradient phase, we introduce the constraints $x \geq 0$. This seems reasonable, since, on the one hand, we know that the solution of the problem is in the nonnegative orthant and, on the other hand, the projected gradient method can effectively handle the bound constraints. Since it is known that the projected gradient method usually behaves well in the first iterations and then becomes slow, we impose a limit of ten iterations to this phase. However, we stopped earlier if one of the following criteria is satisfied:

— $(\Psi_F(x^{k-1}) - \Psi_F(x^k))/\Psi_F(x^k) \leq 0.05$;

— the zero variables are the same in x^{k-1} and x^k and $(\Psi_F(x^{k-1}) - \Psi_F(x^k))/\Psi_F(x^k) \leq 0.1$;

— $\Psi_F(x^k) \leq 10^{-5}\sqrt{n}$.

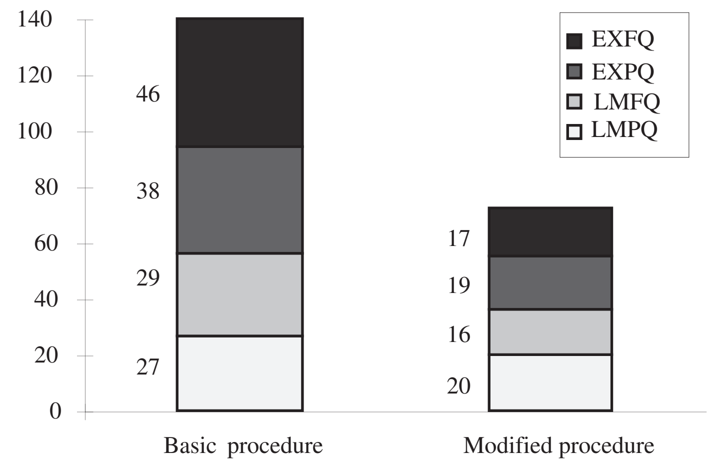

Figure 1: Failures.

5.2 Analysis of the Results

Our first aim is to determine whether the initial steps of projected gradient are beneficial. In Figure 1 we report the number of failures of each algorithm without using the projected gradient phase (basic procedure) and with the projected gradient phase (modified procedure). It is apparent that the projected gradient phase greatly enhances the robustness of all the algorithms. The overall failures pass from 140 to 72, with a reduction of 50%. We also see that the algorithms which seem to benefit more by the gradient steps are the exact ones. This probably indicates that a pure Newton direction is less robust than a mixture of Newton and gradient directions. In fact, the Levenberg-Marquardt directions may be seen as a way of combining the Newton and gradient direction. Nevertheless, even this seems to be not enough and a few initial gradient steps improve the performance also of the Levenberg-Marquardt algorithms. It may also be noted that if the gradient projection strategy is used, there seems to be no significant difference in the robustness of the four algorithms, thus suggesting that the gradient projection phase is actually effective in moving the starting point in a good region, from which all the four directions considered behave well.

In Figure 2 we report more in detail the number of failures for each algorithm and for each set of test problems. Some observations can be made. The degeneracy of the solution, that corresponds to the nonsmoothness of Φ_F and Φ_P at the solution, seems to have no impact on the robustness of the algorithms. Furthermore, as could easily be expected, the faraway starting points are consistently more difficult than the original ones. What is more surprising is the relatively high number of failures for the exact algorithms (without projected gradient phase), even for the standard starting points. Again, this seems to suggest that the pure Newton directions are not very robust, and that taking into account the gradient direction, as is done by the Levenberg-Marquardt methods, increases the stability of the algorithms.

We think that the previous results clearly show that the gradient phase is overall very useful in increasing the robustness of all algorithms considered, even if it appears to be more

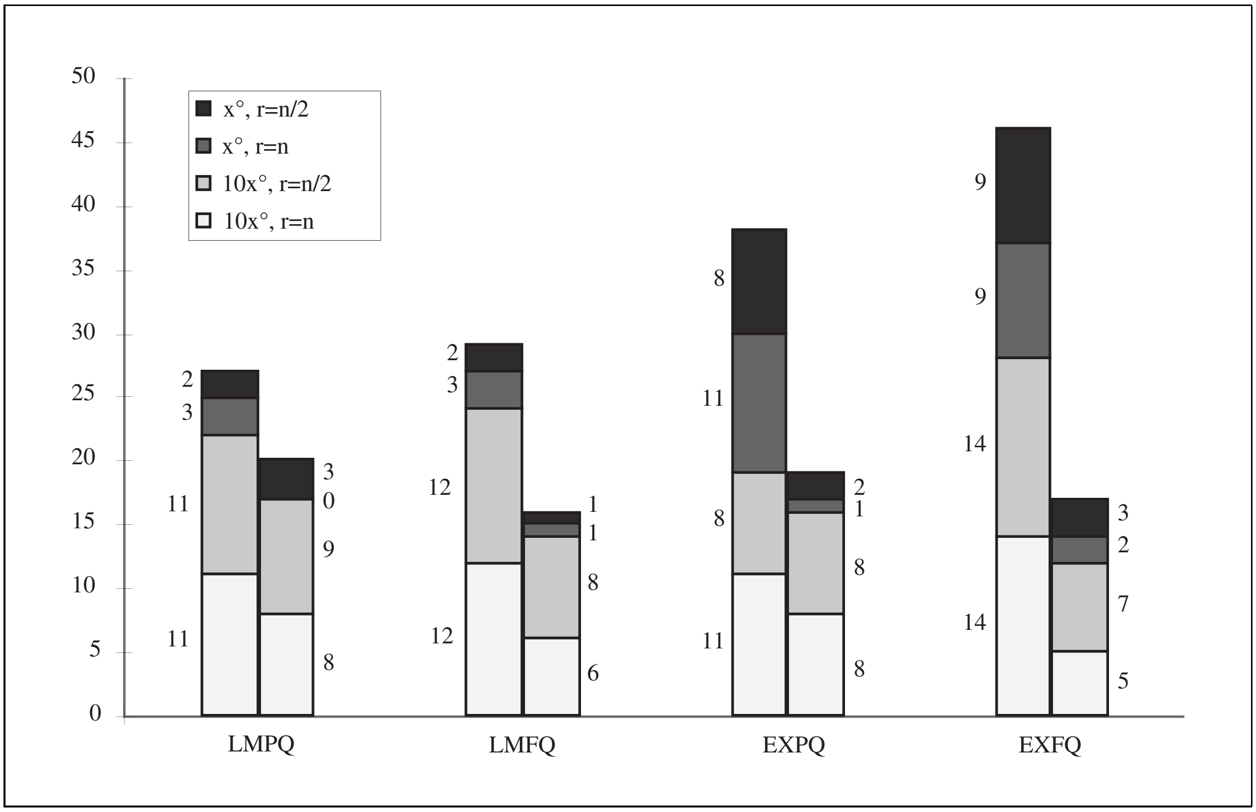

Figure 2: Failures in detail.

useful in the case of the exact algorithms. The next step is then to determine how costly the projected gradient phase is and which algorithm is more efficient from the computational point of view. In order to compare the various algorithms in a homogeneous way, we only considered the subset of test problems for which all algorithms (i.e., the four basic algorithms and the ones including the gradient phase) did not fail and converged to the same solution. This results in a total of 55 test problems: 22 for $n = 100$, 16 for $n = 1000$ and 17 for $n = 10000$.

In Figures 3–5 we report the cumulative times needed by the algorithms to solve the problems of dimension $n = 100$, $n = 1000$ and $n = 10000$, respectively. We reported the times according to the dimensions because the times needed to solve problems, let us say, of dimension 100 are not comparable to those needed to solve those of dimension 10000. Let us first compare the version using exact linear solvers (EXPQ and EXFQ). We see that the modified procedure, i.e., the one using the preliminary steps of projected gradient, never requires more time than the basic procedure. This fact, along with the improvement in the number of failures, seems to indicate that the projected gradient phase is certainly very beneficial on these two algorithms. Another observation is that the algorithms using the Pang-Qi direction are consistently faster than their counterparts using the Fischer-Qi direction. This behavior is due to the fact that to calculate the Pang-Qi direction it is sufficient to solve a linear system of reduced dimension, as already observed in Section 4.4.

If we pass to examine the inexact Levenberg-Marquardt algorithms (LMPQ and LMFQ), we see that some differences are present. In particular, the time needed by the modified versions can be substantially higher than the one needed by the corresponding basic algorithm. In particular, this is apparent in the case $n = 1000$ for both algorithms and in the case $n = 10000$ for the LMPQ one. A closer look at the runs showed that this anomalous behavior is almost completely due to the behavior on three instances of problem 15. When solving these test problems, the projected gradient phase passes an initial point to the Newton phase that gives rise to a Levenberg-Marquardt system which is extremely difficult to solve by the conjugate gradient method. The increase of times reported in Figures 4 and 5 is therefore almost entirely due to the very high number of conjugate gradient inner iterations needed to approximately solve the first systems of the Newton phase when solving these instances of problem 15.

In Figure 6 and in Figure 7 we therefore report the same data of Figures 4 and 5 without considering, however, the three instances of problem 15. If we now look at the Figures 3, 6 and 7, we see that, similarly to what observed for the EXPQ and EXFQ algorithms, the LMPQ algorithms are faster than the LMFQ ones. If we compare the basic and the modified inexact Levenberg-Marquardt versions, we see that the modified versions usually need roughly the same time as the basic versions (except that for the LMFQ when $n = 10000$). This is remarkably different from what we observed for the exact algorithms, where a substantial decrease of the time takes place.

To understand better this behavior, we report in Figure 8 the number of iterations of the various algorithms. Note that the number of iterations of the projected gradient is, obviously, the same for every algorithm, and that the effect of the gradient phase is to reduce the number of subsequent Newton-type iterations (compare the black columns with the dark grey ones). The percentual decrease in the number of Newton iterations, when passing from the basic procedure to the modified one, is roughly similar for all the algorithms. It is then natural that the weight of the gradient phase will be higher for the Levenberg-Marquardt algorithms than for the exact ones since, at least on the problems considered here, the work needed to solve exactly one linear system is higher than the cost to approximately solve the same system by the conjugate gradient method (recall that one system is solved, possibly approximately, at each iteration).

With regard to Figure 8, it is also interesting to note that the number of Newton iterations for the four algorithms (both in the basic or in the modified procedure) is very similar for all the algorithms. Thus we may conclude that the more marked differences in the times reported in the previous tables are mainly due to the different cost of the search direction computations and, possibly, to the number of function evaluations. In fact we recall that the number of Jacobian evaluations is equal to the number of total iterations, and therefore roughly the same for all algorithms.

In Figure 9 we therefore report the number of function evaluations. There are two interesting facts to be observed. First the high number of function evaluations needed by the projected gradient phase. In general the projected gradient method is considered to be a cheap method, however, this is true only if the cost of function evaluations is "low". In our case low should be read as (much) cheaper than the cost of (approximately) solving a linear system. In our test problem set the function evaluations are actually very cheap; however, should the algorithms be applied to problems with very high function evaluation costs, the times of the modified procedures could become much higher than the cost of the basic procedure, contrary to what is reported in Figures 3–7.

A second point to be remarked is that, both in the case of the basic and of the modified versions, the number of function evaluations needed in the Newton phase is sensibly lower for the algorithms based on the Fischer-Qi directions. This corresponds to the fact that much less backtrackings occur when using a Fischer-Qi direction than when using a Pang-Qi direction. This confirms the observation made in Section 4.7 that the Fischer-Qi direction is much more directly related to the merit function Ψ_F while the Pang-Qi directions do not even need to be descent directions for Ψ_F.

Again, since in our test set the cost of function evaluations is low, this behavior does not affect much the times reported in Figures 3–7. However, we should expect that the EXFQ and LMFQ methods become more and more competitive as they are applied to problems with higher and higher function evaluation costs.

5.3 Summary of the Results

The previous observations can be summarized as follows. The projected gradient phase certainly enhances the robustness of all the four algorithms. The additional cost of this phase is low on the test problems considered here, but could become high for those problems with a high function evaluation cost. If the projected gradient phase is used, the reliability of the four algorithms is similar, and they roughly need the same number of iterations. If, however, we consider the times needed by the various algorithms, the behavior may differ widely. If the cost of the function evaluations is low, the Pang-Qi algorithms are probably more convenient; however, if we want to solve problems with functions extremely expensive to evaluate, the Fischer-Qi algorithms can be more attractive.

Another aspect that should be kept in mind is the difficulty in solving the linear systems. All the problems we considered gave rise to sparse, "easy" systems; and this is typical of many applications. If the linear systems are "difficult" and not huge, the exact versions could be preferable; on the other hand, if extremely large instances have to be solved, the inexact algorithms are probably the only available options.

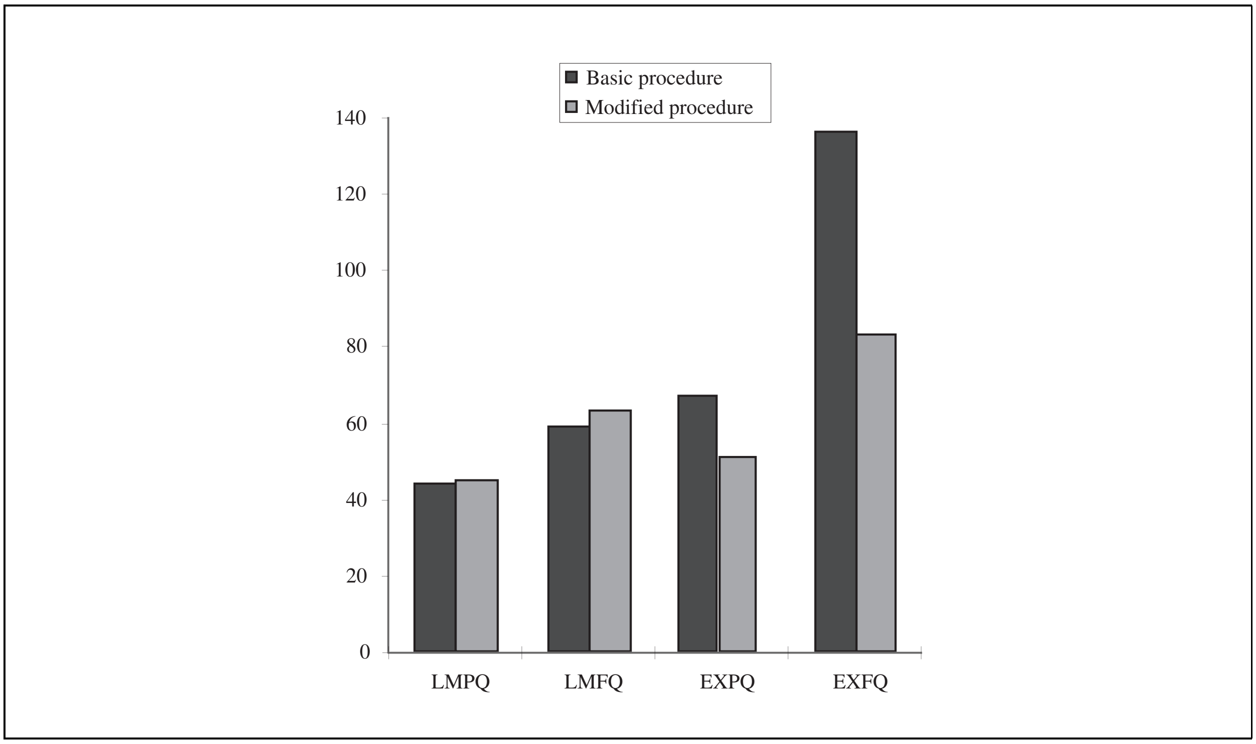

Figure 3: Times in $10^{-2}\times$ seconds, $n = 100$.

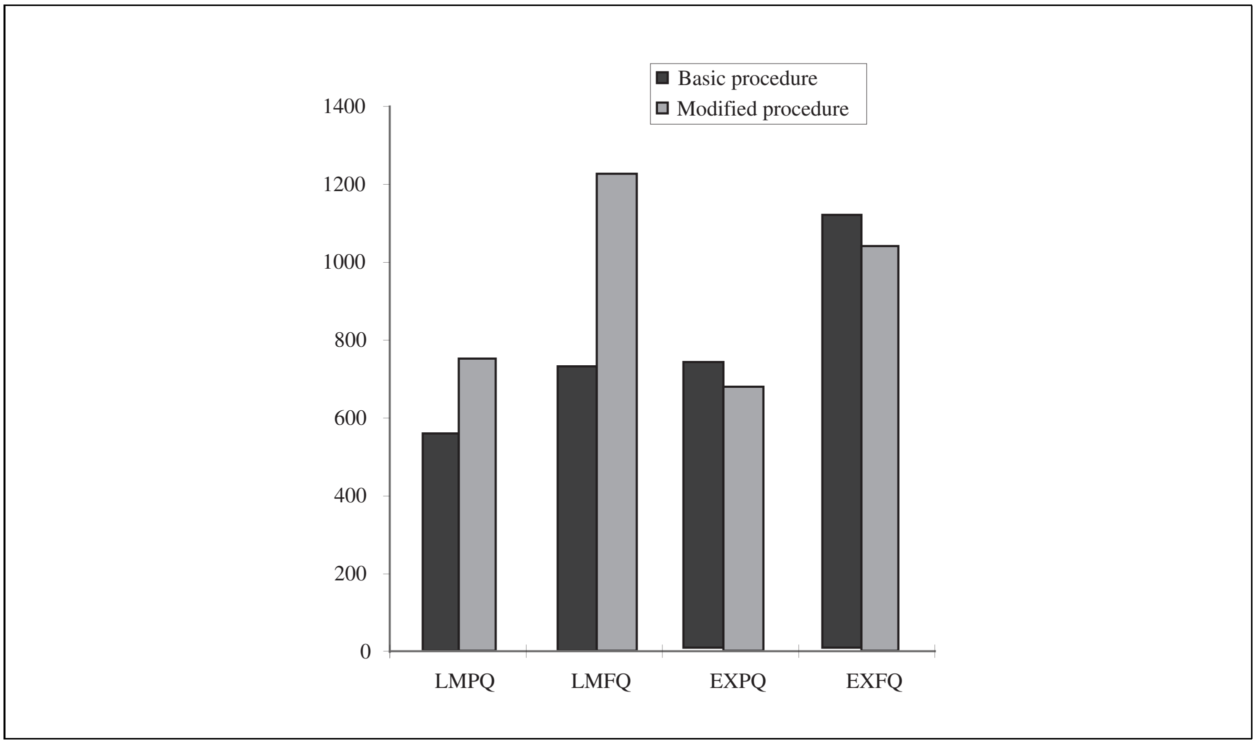

Figure 4: Times in $10^{-2}\times$ seconds, $n = 1000$.

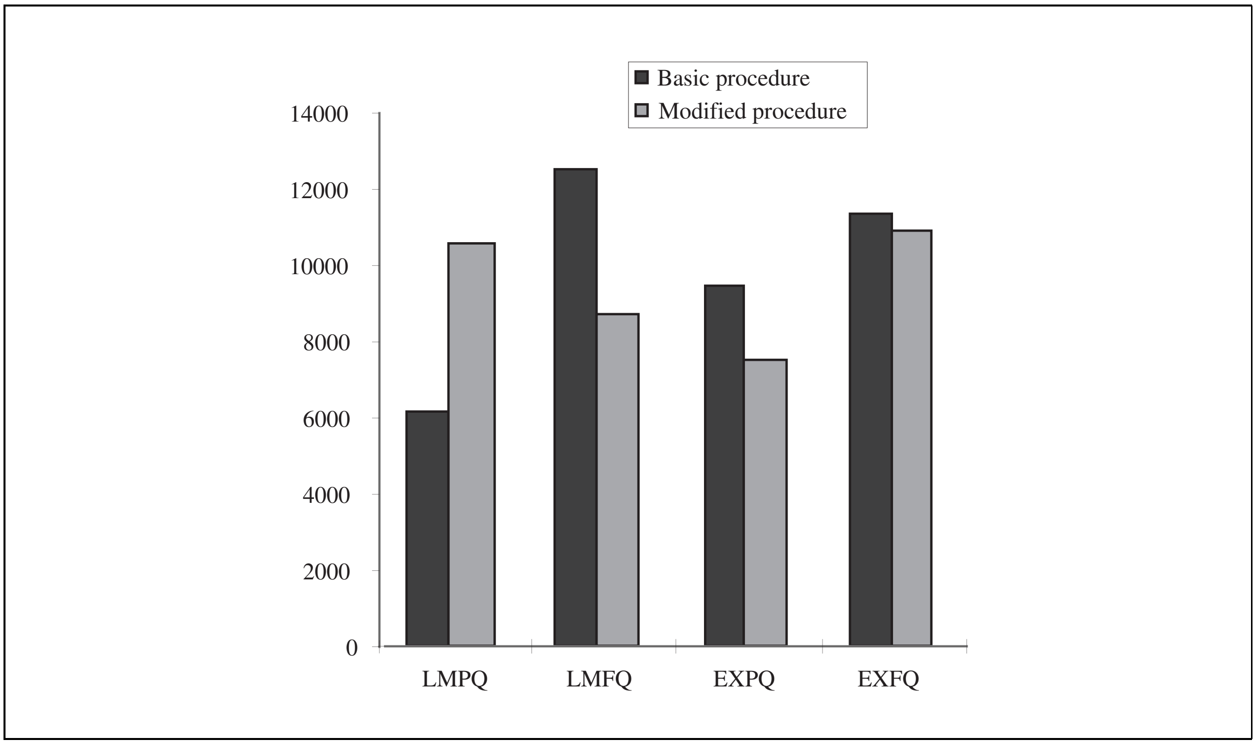

Figure 5: Times in $10^{-2}\times$ seconds, $n = 10000$.

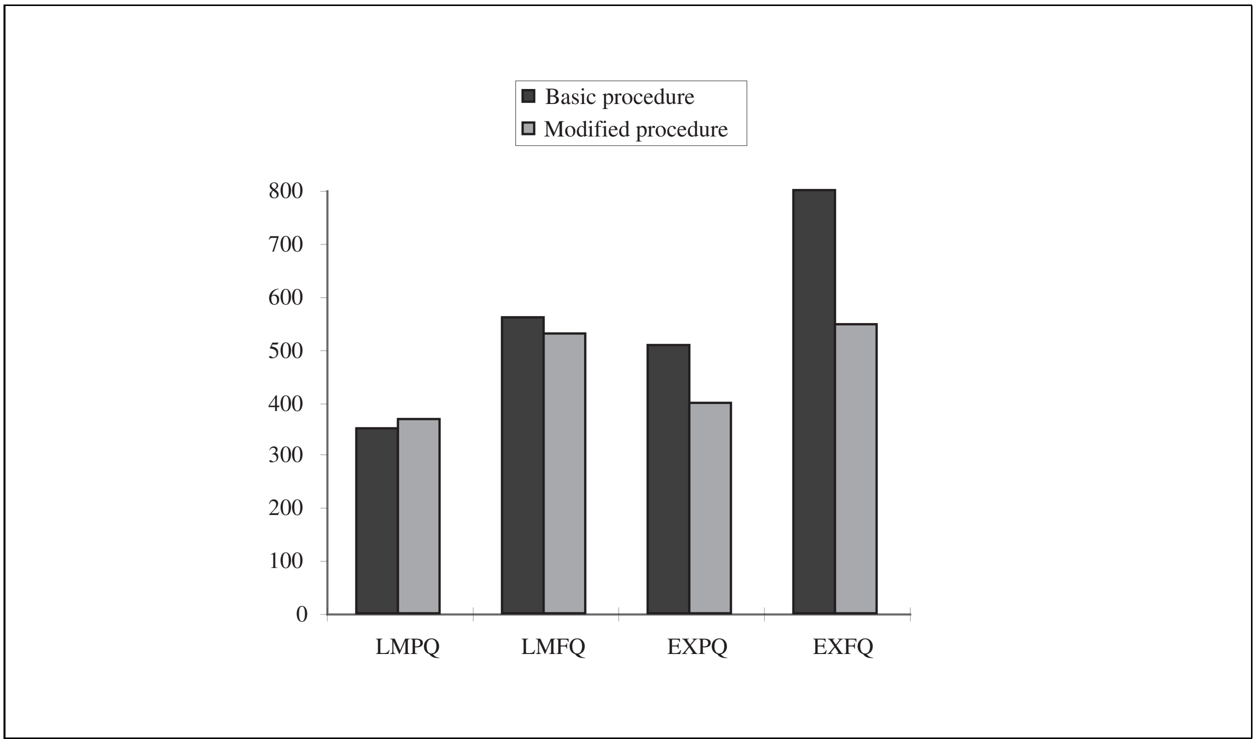

Figure 6: Times in $10^{-2}\times$ seconds, $n = 1000$, without Problems 15.

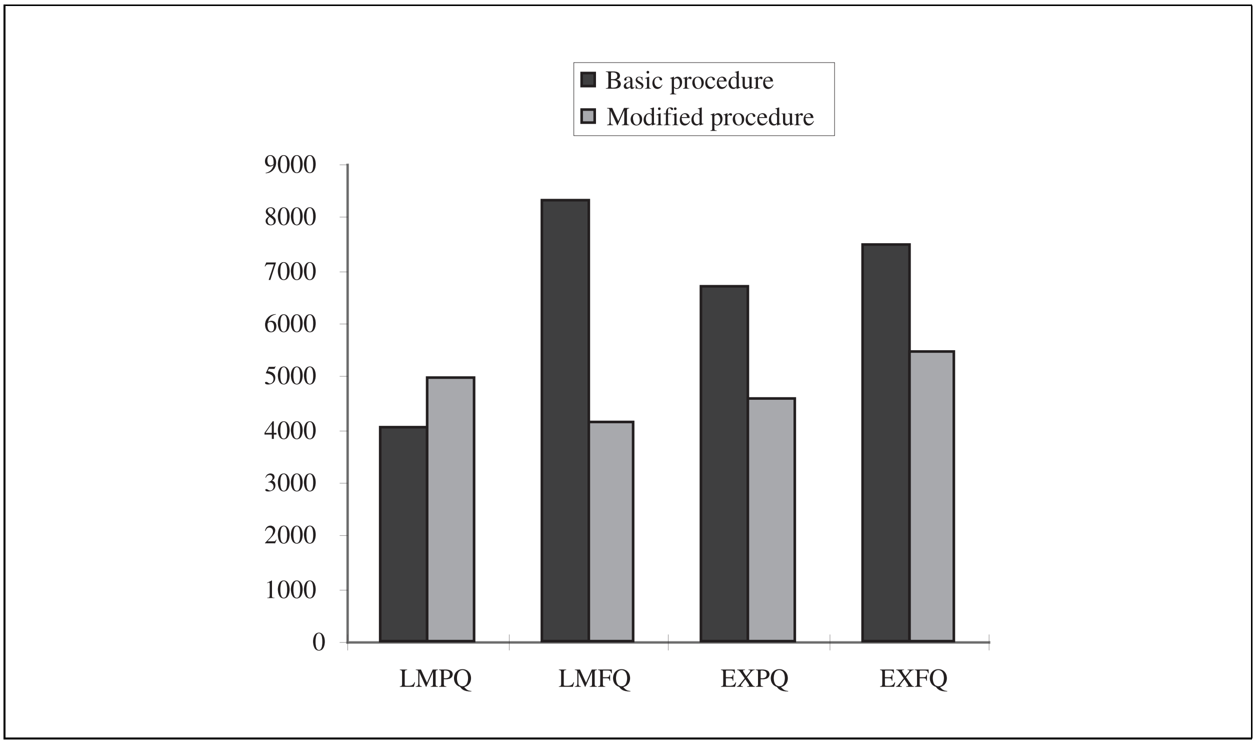

Figure 7: Times in $10^{-2}\times$ seconds, $n = 10000$, without Problems 15.

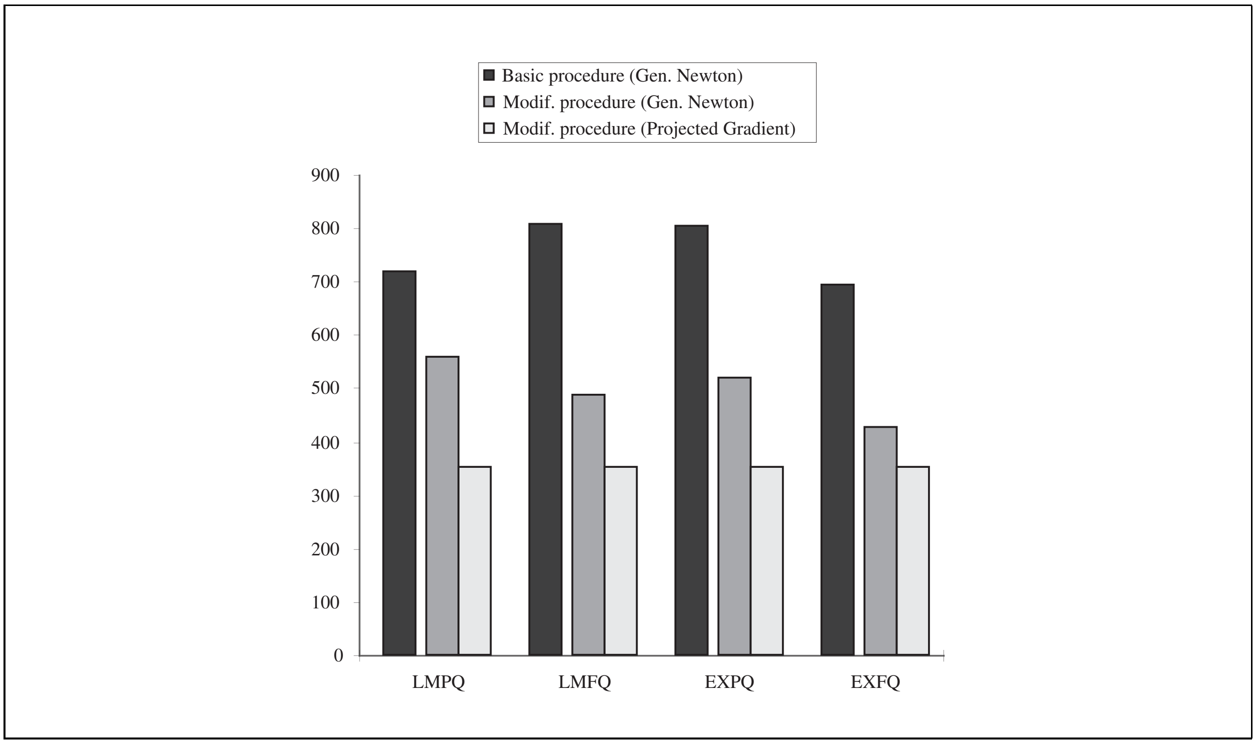

Figure 8: Number of iterations.

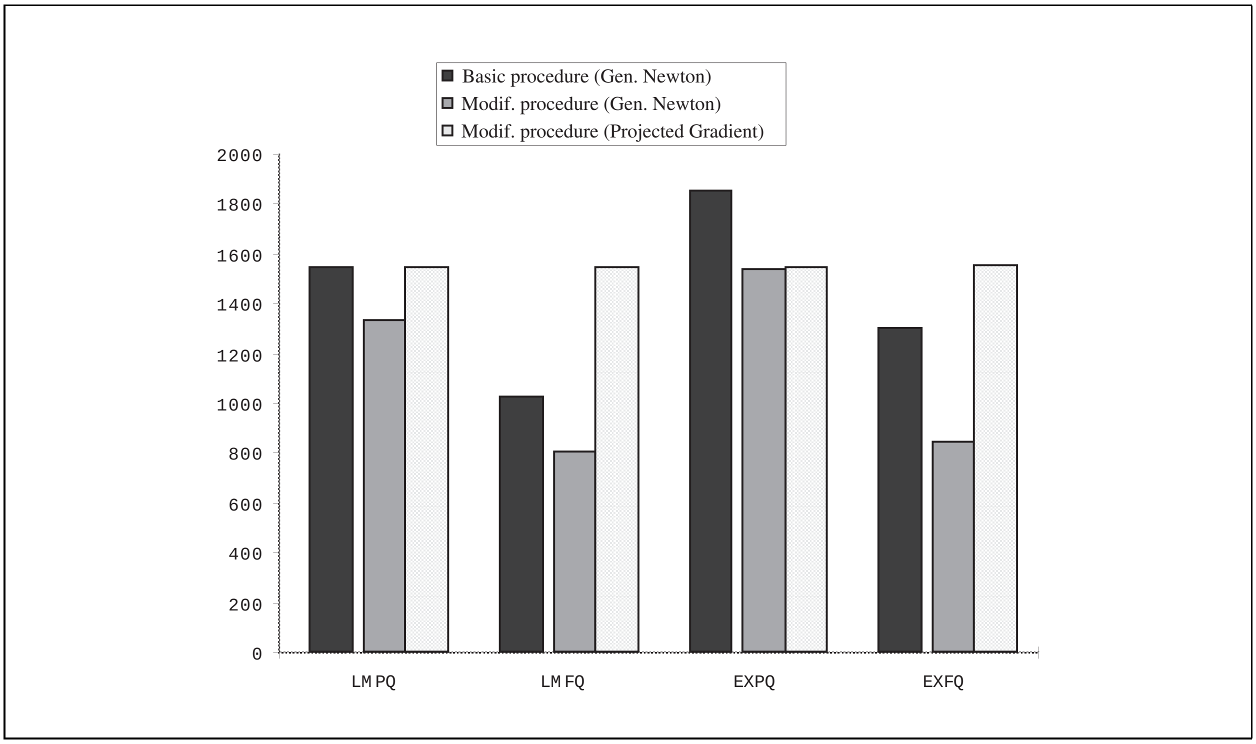

Figure 9: Number of function evaluations.

6 Final Remarks

We have presented and studied a general line search scheme which allows us to analyze in a unified framework several semismooth algorithms for the solution of nonlinear complementarity problems. Several search directions have been studied in detail, both from a theoretical and a numerical point of view. However, we remark that it is easily possible to define and analyze in a similar way other algorithms. Below, we hint at some of these further possibilities.

We first recall that a mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is called an *NCP-function* if

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

cf. (1). If φ is any NCP-function, then NCP(F) is equivalent to

$$\Phi(x) = 0,$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

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

Obviously, the Fischer function φ_F and the Minimum function φ_P are NCP-functions. Other examples include, e.g.,

$$\varphi_{KK}(a, b) := \sqrt{(a-b)^2 + \lambda ab} - a - b, \quad \lambda \in (0, 4) \text{ fixed}$$

(see Kanzow and Kleinmichel [22]), and

$$\varphi_{CCK}(a, b) := \lambda \varphi_F(a, b) - (1-\lambda) a_+ b_+, \quad \lambda \in (0, 1) \text{ fixed}$$

(see Chen, Chen and Kanzow [2]), where $a_+ := \max\{0, a\}$ and, similarly, $b_+ := \max\{0, b\}$. We denote the corresponding operators by Φ_{KK} and Φ_{CCK}, respectively.

The function φ_{KK} is interesting because for $\lambda = 2$ it reduces to the Fischer function, and in the limiting case $\lambda = 0$ it becomes a multiple of the Minimum function. On the other hand, very strong theoretical and numerical results have been reported for φ_{CCK} in [2].

Now, similar to the (exact, inexact and LM inexact) Fischer-Qi directions, we can define (exact, inexact and LM inexact) search directions based on Φ_{KK} and Φ_{CCK}. Since it is known [22, 2] that Φ_{KK} and Φ_{CCK} are (strongly) semismooth if F is a C^1-function (LC^1-function) and that an R-regular solution of NCP(F) is a BD-regular solution for the nonlinear systems of equations $\Phi_{KK}(x) = 0$ and $\Phi_{CCK}(x) = 0$, it is not difficult to see that Theorems 4.1, 4.2 and 4.3 also hold for the corresponding search directions based on Φ_{KK} and Φ_{CCK}.

To conclude, we also mention that the results of this paper can be generalized to variational inequalities with box constraints. In fact, the method introduced very recently in [21] can be viewed as an extension of one of the methods discussed in this paper.

References

[1] D.P. Bertsekas: *Nonlinear Programming.* Athena Scientific, Belmont, MA, 1995.

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

[3] B. Chen and P.T. Harker: *A noninterior continuation method for quadratic and linear programming.* SIAM Journal on Optimization 3, 1993, pp. 503–515.

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

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

[6] 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.

[7] 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).

[8] S.P. Dirkse and M.C. Ferris: *Crash techniques for large-scale complementairty problems.* In: M.C. Ferris and J.-S. Pang (eds.): *Complementarity and Variational Problems: State of the Art.* SIAM, Philadelphia, PA, 1997, pp. 40–53.

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

[10] F. Facchinei and C. Kanzow: *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems.* Mathematical Programming 76, 1997, pp. 493–512.

[11] 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.

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

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

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

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

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

[17] M.A. Gomes-Ruggiero, J.M. Martínez and S.A. Santos: *Solving nonsmooth equations by means of quasi-Newton methods with globalization.* In: D.-Z. Du, L. Qi and R. Womersley (eds.): *Recent Advances in Nonsmooth Optimization.* World Scientific Publisher, Singapore, 1995, pp. 121–140.

[18] 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.

[19] 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.

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

[21] C. Kanzow and M. Fukushima: *Solving box constrained variational inequality problems by using the natural residual with D-gap function globalization.* Mathematical Programming Technical Report 97-14, Computer Sciences Department, University of Wisconsin – Madison, Madison, WI, November 1997.

[22] C. Kanzow and H. Kleinmichel: *A new class of semismooth Newton-type methods for nonlinear complementarity problems.* Computational Optimization and Applications, to appear.

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

[24] L. Lukšan: *Inexact trust region method for large sparse systems of nonlinear equations.* Journal of Optimization Theory and Applications 81, 1994, pp. 569–590.

[25] 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.

[26] J.M. Martínez and L. Qi: *Inexact Newton methods for solving nonsmooth equations.* Journal of Computational and Applied Mathematics 60, 1995, pp. 127–145.

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

[28] J. J. Moré and W. C. Rheinboldt: *On P- and S-functions and related classes of n-dimensional nonlinear mappings.* Linear Algebra and its Applications 6, 1973, pp. 45–68.

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

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

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

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

[33] L. Qi and J. Sun: *A nonsmooth version of Newton's method.* Mathematical Programming 58, 1993, pp. 353–368.

[34] P. Tseng: *Growth behavior of a class of merit functions for the nonlinear complementarity problem.* Journal of Optimization Theory and Applications 89, 1996, pp. 69–94.