On the Rate of Convergence of the Levenberg-Marquardt Method[1]

Nobuo Yamashita and Masao Fukushima

Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
Kyoto 606-8501, Japan

October 28, 2000

Abstract

We consider a rate of convergence of the Levenberg-Marquardt method (LMM) for solving a system of nonlinear equations $F(x) = 0$, where F is a mapping from R^n into R^m. It is well-known that LMM has a quadratic rate of convergence when $m = n$, the Jacobian matrix of F is nonsingular at a solution x and an initial point is chosen sufficiently close to x. In this paper, we show that if $\|F(x)\|$ provides a local error bound for the system of nonlinear equations, then a sequence generated by LMM converges to the solution set quadratically.

[1]This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan.

1 Introduction

We consider the system of nonlinear equations

$$F(x) = 0, \tag{1}$$

where F is a mapping from R^n to R^m. Throughout the paper, we assume that F is continuously differentiable and the system (1) has a nonempty solution set, which we denote by X^*.

The Levenberg-Marquardt method (LMM) is classical but still one of the most popular solution methods for (1) [1, 3, 6]. LMM is a kind of Newton-type method and it adopts the solution d^k of the following system of linear equations as a search direction:

$$(F'(x^k)^T F'(x^k) + \mu_k I)d = -F'(x^k)^T F(x^k), \tag{2}$$

where μ_k is a positive parameter. Since $(F'(x^k)^T F'(x^k) + \mu_k I)$ is a positive definite matrix, the system (2) has the unique solution d^k, which is a nice property of LMM compared with the pure Newton's method or the Gauss-Newton method. Moreover, the search direction d^k is a descent direction of the merit function $\phi : R^n \to R$ defined by

$$\phi(x) := \frac{1}{2}\|F(x)\|^2.$$

Therefore, combining LMM with Armijo step size based on the merit function ϕ, we may expect that such LMM converges globally. In particular we can show that any accumulation point x^* of a sequence generated by such LMM is a stationary point of ϕ, i.e.,

$$0 = \nabla\phi(x^*) = F'(x^*)^T F(x^*),$$

and hence x^* is a solution of (1) if $F'(x^*)$ is nonsingular. As to a rate of convergence, it is well-known that LMM has a superlinear convergence property if $F'(x^*)$ is nonsingular and μ_k is updated by an appropriate rule.

However, the nonsingularity of $F'(x^*)$ seems too stringent a requirement for the purpose of ensuring superlinear convergence of LMM. Recently, algorithms that enjoy superlinear convergence without the nonsingularity condition on $F'(x^*)$ have been proposed [8, 9] for the nonlinear complementarity problem, which is in some sense more general than the system of equations. To guarantee the superlinear convergence, those algorithms require a local error bound assumption instead of the nonsingularity condition. A local error bound for the system (1) is defined as follows.

Definition 1.1 *Let N be a subset of R^n such that $X^* \cap N \neq \emptyset$. We say that $\|F(x)\|$ provides a local error bound on N for the system (1) if there exists a positive constant c such that*

$$c\,\mathrm{dist}(x, X^*) \leq \|F(x)\| \quad \forall x \in N. \tag{3}$$

We note that, if $F'(x^*)$ is nonsingular at a solution x^* of (1), then x^* is an isolated solution and $\|F(x)\|$ provides a local error bound on some neighborhood of x^*. Note that the converse is not necessarily true. For example, let us consider the mapping $F : R^2 \to R^2$ defined by $F(x_1, x_2) = (e^{x_1} - 1, 0)^T$. The solution set of (1) is given by $X^* = \{x \in R^2 \,|\, x_1 = 0\}$, and hence $\mathrm{dist}(x, X^*) = |x_1|$. Then it is not difficult to see that (3) holds for any $c \in (0, 1)$ provided N is chosen as $N = \{x \in R^2 \,|\, -a < x_1 < a\}$ with sufficiently small $a > 0$. However $F'(x)$ is clearly singular for any $x \in X^*$. Hence we see that a local error bound condition is weaker than the nonsingularity of $F'(x^*)$. The purpose of this paper is to reveal that the key condition for superlinear convergence of a Newton-type method is the local error bound rather than the nonsingularity. More specifically, we show that LMM has a quadratic rate of convergence under a local error bound condition. Furthermore, applying this result to the system of nonlinear equations equivalent to the linear complementarity problem, we show that the LMM for the linear complementarity problem enjoys more favorable convergence property compared with other existing algorithms.

The paper is organized as follows. In Section 2, we show that LMM with unit step size has the quadratic rate of convergence under a local error bound condition. In Section 3, we consider the LMM with Armijo step size rule based on the merit function ϕ and show its global convergence. In Section 4, we apply the results obtained in Sections 2 and 3 to the linear complementarity problem.

2 The Quadratic Rate of Convergence of LMM

In this section, we discuss local convergence properties of LMM without line search, i.e., the next iterate x^{k+1} is given by

$$x^{k+1} := x^k + d^k,$$

where d^k is the solution of the system (2) of linear equations.

Most of the previous local convergence analysis of LMM was based on the fact that (2) is almost equivalent to the Newton equation near a solution where the Jacobian of F is nonsingular. Here we will take another viewpoint, i.e., we consider an unconstrained minimization problem equivalent to (2). By using a property of the minimization problem, we will investigate a rate of convergence of LMM. To this end, define the quadratic function $\theta^k : R^n \to R$ by

$$\theta^k(d) = \|F'(x^k)d + F(x^k)\|^2 + \mu_k \|d\|^2.$$

The function θ^k is strictly convex. Therefore, the unconstrained minimization problem

$$\min \;\; \theta^k(d) \tag{4}$$

is equivalent to (2), which can be easily verified from the first-order optimality condition for (4).

Next, we make some assumptions that will guarantee the quadratic rate of convergence of LMM.

Assumption 2.1 *For some $x^* \in X^*$, the following conditions (a) and (b) hold:*

(a) *There exist constants $b \in (0,1)$ and $c_1 \in (0,\infty)$ such that*

$$\|F'(y)(x-y) - (F(x) - F(y))\| \le c_1\|x-y\|^2 \quad \forall x, y \in N(x^*, b) := \{x \mid \|x - x^*\| \le b\}.$$

(b) *$\|F(x)\|$ provides a local error bound on $N(x^*, b)$ for the system (1), i.e., there exists a constant $c_2 \in (0,\infty)$ such that*

$$c_2 \operatorname{dist}(x, X^*) \le \|F(x)\| \quad \forall x \in N(x^*, b).$$

Assumption 2.1 (a) holds when F is continuously differentiable and F' is Lipschitz continuous. Note that, by Assumption 2.1 (a), there exists a positive constant L such that

$$\|F(x) - F(y)\| \le L\|x-y\| \quad \forall x, y \in N(x^*, b). \tag{5}$$

Assumption 2.1 (b) plays an important role in the paper. It is well-known [7] that the local error bound condition holds when F is a piecewise linear function.

We need a further assumption concerning the choice of parameters $\{\mu_k\}$.

Assumption 2.2 *$\mu_k := \|F(x^k)\|^2$ for all k.*

Under these assumptions, we show the following key lemmas. For all k, $\bar{x}^k$ denotes a vector in X^* such that

$$\|x^k - \bar{x}^k\| = \operatorname{dist}(x^k, X^*).$$

Below, x^* refers to the solution of (1) as specified in Assumption 2.1.

Lemma 2.1 *Suppose that Assumptions 2.1 and 2.2 hold. If $x^k \in N(x^*, b/2)$, then the solution d^k of the linear system (2) satisfies*

$$\begin{aligned} \|d^k\| &\le c_3 \operatorname{dist}(x^k, X^*), \\ \|F'(x^k)d^k + F(x^k)\| &\le c_4 \operatorname{dist}(x^k, X^*)^2, \end{aligned}$$

where $c_3 = \sqrt{c_1^2 + c_2^2}/c_2$ and $c_4 = \sqrt{c_1^2 + L^2}$.

Proof. Since d^k solves the minimization problem (4), we get

$$\theta^k(d^k) \le \theta^k(\bar{x}^k - x^k). \tag{6}$$

Furthermore, since $x^k \in N(x^*, b/2)$, we get

$$\|\bar{x}^k - x^*\| \le \|\bar{x}^k - x^k\| + \|x^* - x^k\| \le \|x^* - x^k\| + \|x^* - x^k\| \le b,$$

and hence $\bar{x}^k \in N(x^*, b)$. It then follows from the definition of θ^k, the inequality (6), and Assumption 2.1 (a) that

$$\begin{aligned}
\|d^k\|^2 &\le \frac{1}{\mu_k}\theta^k(d^k) \\
&\le \frac{1}{\mu_k}\theta^k(\bar{x}^k - x^k) \\
&= \frac{1}{\mu_k}\left(\|F'(x^k)(\bar{x}^k - x^k) + F(x^k)\|^2 + \mu_k\|\bar{x}^k - x^k\|^2\right) \\
&\le \frac{1}{\mu_k}\left(c_1^2\|x^k - \bar{x}^k\|^4 + \mu_k\|x^k - \bar{x}^k\|^2\right) \\
&\le \left(\frac{c_1^2\|x^k - \bar{x}^k\|^2}{\mu_k} + 1\right)\|x^k - \bar{x}^k\|^2.
\end{aligned} \tag{7}$$

Since

$$\mu_k = \|F(x^k)\|^2 \ge c_2^2\|x^k - \bar{x}^k\|^2$$

by Assumption 2.1 (b) and Assumption 2.2, the inequality (7) yields

$$\|d^k\| \le \sqrt{\frac{c_1^2 + c_2^2}{c_2^2}}\|x^k - \bar{x}^k\|.$$

Next we show the second inequality. In a way similar to proving the first inequality, we can show

$$\begin{aligned}
\|F'(x^k)d^k + F(x^k)\|^2 &\le \theta^k(d^k) \\
&\le \theta^k(\bar{x}^k - x^k) \\
&\le c_1^2\|x^k - \bar{x}^k\|^4 + \mu_k\|x^k - \bar{x}^k\|^2.
\end{aligned}$$

Since

$$\mu_k = \|F(x^k)\|^2 = \|F(x^k) - F(\bar{x}^k)\|^2 \le L^2\|x^k - \bar{x}^k\|^2$$

by the Lipschitz property (5) and Assumption 2.2, we then have

$$\|F'(x^k)d^k + F(x^k)\| \le \sqrt{c_1^2 + L^2}\|x^k - \bar{x}^k\|^2.$$

□

The next lemma shows that $\mathrm{dist}(x^k, X^*)$ converges to 0 quadratically, as long as the iterates $\{x^k\}$ lie sufficiently near x^*.

Lemma 2.2 *If $x^k, x^{k-1} \in N(x^*, b/2)$, then*

$$\mathrm{dist}(x^k, X^*) \leq c_5 \mathrm{dist}(x^{k-1}, X^*)^2 \tag{8}$$

holds, where $c_5 = (c_1 c_3^2 + c_4)/c_2$.

Proof. Since $x^k, x^{k-1} \in N(x^*, b/2)$ and $x^k = x^{k-1} + d^{k-1}$, it follows from Assumption 2.1 (a), (b) and Lemma 2.1 that

$$\begin{aligned}
c_2\,\mathrm{dist}(x^k, X^*) &= c_2\,\mathrm{dist}(x^{k-1} + d^{k-1}, X^*) \\
&\leq \|F(x^{k-1} + d^{k-1})\| \\
&\leq \|F'(x^{k-1})d^{k-1} + F(x^{k-1})\| + c_1\|d^{k-1}\|^2 \\
&\leq c_4\,\mathrm{dist}(x^{k-1}, X^*)^2 + c_1 c_3^2\,\mathrm{dist}(x^{k-1}, X^*)^2 \\
&= (c_1 c_3^2 + c_4)\,\mathrm{dist}(x^{k-1}, X^*)^2.
\end{aligned}$$

Therefore, setting $c_5 = (c_1 c_3^2 + c_4)/c_2$ gives (8). □

From this lemma, we see that $\{\mathrm{dist}(x^k, X^*)\}$ converges to 0 quadratically if $x^k \in N(x^*, b/2)$ for all k. Next, we show that the latter condition holds, provided the initial point x^0 is chosen sufficiently close to x^*. Let

$$r := \min\left\{\frac{b}{2 + 4c_3}, \frac{1}{2c_5}\right\}. \tag{9}$$

Lemma 2.3 *If $x^0 \in N(x^*, r)$, then we have $x^k \in N(x^*, b/2)$ for all k.*

Proof. We show that, for each k, if $x^l \in N(x^*, b/2), l = 1, \ldots, k$, then $x^{k+1} \in N(x^*, b/2)$. Then, since $x^0 \in N(x^*, r) \subseteq N(x^*, b/2)$, it shows the lemma.

We consider two cases $k = 0$ and $k \geq 1$: When $k = 0$, it follows from Lemma 2.1 that

$$\begin{aligned}
\|x^1 - x^*\| &\leq \|x^0 + d^0 - x^*\| \\
&\leq \|x^0 - x^*\| + \|d^0\| \\
&\leq r + c_3\,\mathrm{dist}(x^0, X^*) \\
&\leq r + c_3\|x^0 - x^*\| \\
&\leq (1 + c_3)r.
\end{aligned} \tag{10}$$

Since $(1 + c_3)r \leq b/2$, we have $x^1 \in N(x^*, b/2)$. Next consider the case $k \geq 1$. Then, since $x^l \in N(x^*, b/2)$ for $l = 0, 1, \ldots, k$, we obtain from Lemma 2.2 that, for $l = 1, \ldots, k$,

$$\mathrm{dist}(x^l, X^*) \leq c_5 \mathrm{dist}(x^{l-1}, X^*)^2 \leq \cdots \leq c_5^{2^l - 1}\|x^0 - x^*\|^{2^l} \leq r\left(\frac{1}{2}\right)^{2^l - 1},$$

where the last inequality follows from $\|x^0 - x^*\| \le r$ and $r \le \frac{1}{2c_5}$. Therefore, from Lemma 2.1, we have for $l = 1, \ldots, k$

$$\|d^l\| \le c_3 \operatorname{dist}(x^l, X^*) \le c_2 r \left(\frac{1}{2}\right)^{2^l - 1} \le c_3 r \left(\frac{1}{2}\right)^{2l-1}. \tag{11}$$

It then follows that

$$\begin{aligned}
\|x^{k+1} - x^*\| &\le \|x^1 - x^*\| + \sum_{l=1}^{k} \|d^l\| \\
&\le (1 + c_3) r + c_3 r \sum_{l=1}^{k} \left(\frac{1}{2}\right)^{2l-1} \\
&\le (1 + c_3) r + c_3 r \sum_{l=1}^{\infty} \left(\frac{1}{2}\right)^{l} \\
&\le (1 + 2c_3) r \\
&\le \frac{b}{2},
\end{aligned}$$

where the second inequality follows from (10) and the last inequality follows from $r \le \frac{b}{2+6c_4}$. Consequently we have $x^{k+1} \in N(x^*, b/2)$. □

By using Lemmas 2.2 and 2.3, we show our main theorem.

Theorem 2.1 *Suppose that Assumptions 2.1 and 2.2 hold. Suppose also that* $\{x^k\}$ *is generated by LMM with* $x^0 \in N(x^*, r)$. *Then* $\{\operatorname{dist}(x^k, X^*)\}$ *converges to* 0 *quadratically. Moreover the sequence* $\{x^k\}$ *converges to a solution* $\hat{x} \in X^* \cap N(x^*, b/2)$.

Proof. The first part of the theorem directly follows from Lemmas 2.2 and 2.3. So, we only show that the sequence $\{x^k\}$ converges to a solution $\hat{x} \in X^* \cap N(x^*, b/2)$. Since $\{\operatorname{dist}(x^k, X)\}$ converges to 0 and $x^k \in N(x^*, b/2)$ for all k, it suffices to show that $\{x^k\}$ converges. Since (11) implies

$$\|d^k\| \le c_3 r \left(\frac{1}{2}\right)^{2k-1}$$

for all $k \ge 1$, we have for any positive integers p, q such that $p \ge q$

$$\|x^p - x^q\| \le \sum_{i=q}^{p-1} \|d^i\| \le \sum_{i=q}^{\infty} \|d^i\| \le c_3 r \sum_{i=q}^{\infty} \left(\frac{1}{2}\right)^{2i-1} = \frac{1}{3} c_3 r \left(\frac{1}{2}\right)^{2q-3}.$$

This means that $\{x^k\}$ is a Cauchy sequence, and hence it converges. □

3 Global Convergence

In the previous section, we have investigated the rate of convergence of LMM with unit step size. In this section, we consider LMM with Armijo step size rule, and discuss its global convergence. To determine a step size, we use the merit function

$$\phi(x) = \frac{1}{2}\|F(x)\|^2.$$

We consider the following algorithm.

LMM with line search

Step 0: Choose parameters $\alpha, \beta, \gamma \in (0,1)$ and an initial point $x^0 \in R^n$. Set $\mu^0 := \|F(x^0)\|^2$ and $k := 0$.

Step 1: If x^k satisfies a termination criterion, then stop.

Step 2: Find the solution d^k of the system of linear equations

$$(F'(x^k)^T F'(x^k) + \mu_k I)d = -F'(x^k)^T F(x^k).$$

If d^k satisfies

$$\|F(x^k + d^k)\| \leq \gamma \|F(x^k)\|, \tag{12}$$

then $x^{k+1} := x^k + d^k$ and go to Step 4. Otherwise go to Step 3.

Step 3: Let m_k be the smallest nonnegative integer m such that

$$\phi(x^k + \beta^m d^k) - \phi(x^k) \leq \alpha \beta^m \nabla\phi(x^k)^T d^k. \tag{13}$$

Set $x^{k+1} := x^k + \beta^{m_k} d^k$.

Step 4: Set $\mu^{k+1} := \|F(x^{k+1})\|^2$ and $k := k+1$. Go to Step 1.

For the above algorithm, we establish the following global convergence theorem. To this end, we assume that $\nabla\phi(x^k) \neq 0$ for all k, so the algorithm generates an infinite sequence $\{x^k\}$.

Theorem 3.1 *Let $\{x^k\}$ be generated by LMM with line search. Then any accumulation point of the sequence $\{x^k\}$ is a stationary point of ϕ. Moreover, if an accumulation point x^* of the sequence $\{x^k\}$ is a solution of the system (1), then the sequence $\{\mathrm{dist}(x^k, X^*)\}$ converges to 0 quadratically, provided Assumption 2.1 with this particular x^* and Assumption 2.2 hold.*

Proof. Since $\nabla\phi(x^k) \neq 0$ implies $d^k \neq 0$, we have

$$\nabla\phi(x^k)^T d^k = (F'(x^k)^T F(x^k))^T d^k = -((F'(x^k)^T F'(x^k) + \mu_k I)d^k)^T d^k < 0.$$

It then follows from (12) and (13) that $\{\phi(x^k)\}$ is monotonically decreasing, and hence μ_k is also monotonically decreasing and has a limit. If $\mu_k \to 0$, then $F(x^k) \to 0$, which implies that any accumulation point of $\{x^k\}$ is a solution of (1), and hence it is a stationary point of ϕ. If $\lim_{k\to\infty} \mu_k = \bar{\mu} > 0$, then we have

$$\nabla\phi(x^k)^T d^k = -((F'(x^k)^T F'(x^k) + \mu_k I)d^k)^T d^k < -\bar{\mu}\|d^k\|^2.$$

Hence, we can show that any accumulation point is a stationary point of ϕ by using the standard arguments. This completes the proof of the first part of the theorem.

Next we show the last part. Since the accumulation point x^* is a solution of (1), there exists a $\bar{k}$ such that

$$\|F(x^{\bar{k}})\| \leq \frac{c_2^2\gamma}{c_5 L} \tag{14}$$

and

$$\|x^{\bar{k}} - x^*\| \leq r,$$

where r is the positive constant specified by (9). Now let $\{y^k\}$ be a sequence generated by LMM with unit step size and $y^0 := x^{\bar{k}}$. Then, by Theorem 2.1, $\mathrm{dist}(y^l, X^*)$ converges to 0 quadratically. Therefore, it suffices to show that $x^{\bar{k}+l} = y^l$ for all l, i.e., $\{y^l\}$ satisfies

$$\|F(y^{l+1})\| \leq \gamma\|F(y^l)\|$$

for all l. Note that, from Lemmas 2.2 and 2.3, we have

$$\mathrm{dist}(y^{l+1}, X^*) \leq c_5\, \mathrm{dist}(y^l, X^*)^2$$

for all l. Let $\bar{y}^{l+1}$ be a point in X^* such that $\|y^{l+1} - \bar{y}^{l+1}\| = \mathrm{dist}(y^{l+1}, X^*)$. It then follows from the Lipschitz property (5) and Assumption 2.1 (b) that

$$\begin{aligned}
\|F(y^{l+1})\| &= \|F(y^{l+1}) - F(\bar{y}^{l+1})\| \\
&\leq L\, \mathrm{dist}(y^{l+1}, X^*) \\
&\leq c_5 L\, \mathrm{dist}(y^l, X^*)^2 \\
&\leq \frac{c_5 L\|F(y^l)\|}{c_2^2}\|F(y^l)\|. \qquad (15)
\end{aligned}$$

Since (14) implies

$$\frac{c_5 L\|F(y^0)\|}{c_2^2} \leq \gamma,$$

we have from (15) and $\gamma < 1$ that

$$\frac{c_5 L\|F(y^l)\|}{c_2^2} \leq \gamma,$$

and hence

$$\|F(y^{l+1})\| \leq \gamma \|F(y^l)\|$$

for all l. This completes the proof. $\square$

4 Application to the Linear Complementarity Problem

In this section, we apply the results obtained in the previous sections to the linear complementarity problem LCP(M,q), which is to find a vector $x \in R^n$ such that

$$x \geq 0, \quad Mx + q \geq 0, \quad x^T(Mx+q) = 0,$$

where M is an $n \times n$ matrix and q is an n-dimensional vector.

Until now, various reformulations of LCP(M,q) to an equivalent system of equations have been proposed. In this section, we consider the following reformulation based on the Fischer-Burmeister function [4]:

$$H(x) := \begin{pmatrix} \psi(x_1, (Mx+q)_1) \\ \vdots \\ \psi(x_n, (Mx+q)_n) \end{pmatrix} = 0, \tag{16}$$

where $\psi : R^2 \to R$ is the Fischer-Burmeister function defied by

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

Since this function is not differentiable, the previous results cannot be applied to the system (16) directly. However, using the generalized Jacobian $\partial H(x)$ in the sense of [2], we can modify LMM so that it can deal with the nonsmooth system (16). More specifically, we adopt, as a search direction, the solution d^k of the linear system

$$((J^k)^T J^k + \mu_k I)d = -\nabla\phi(x^k),$$

where $J^k \in \partial H(x^k)$. Note that $\phi(x) := \frac{1}{2}\|H(x)\|^2$ is differentiable. Though the function H is not differentiable everywhere, H enjoys some nice properties related with Assumption 2.1. In particular, H provides a local error bound on $\{x \in R^n \mid \|H(x)\| \leq \varepsilon\}$ for LCP(M,q), where ε is a sufficiently small positive constant [5], and if $(x_i, F_i(x)) \neq (0,0)$ for all i, then H is twice continuously differentiable at x. Moreover, any stationary point of ϕ is a solution of LCP(M,q) when M is a P_0 matrix.

Using these properties and the results obtained in the previous sections, we can establish the following theorem. We omit the proof, since it is a rather straightforward modification of that of Theorem 3.1.

Theorem 4.1 *Suppose that M is a P_0 matrix. Then any accumulation point of the sequence generated by LMM with line search is a solution of LCP(M, q). Moreover, if an accumulation point x^* of the generated sequence is a nondegenerate solution, i.e., $x_i^* + F_i(x^*) > 0$ for all i, then the sequence converges to the solution set of LCP(M, q) quadratically.* $\square$

For LCP(M, q), some existing algorithms, such as interior point methods and proximal point algorithms, are known to have a quadratic rate of convergence when one of the following conditions holds [8, 9, 10].

(a) M is positive semidefinite and one of the accumulation points is a nondegenerate solution.

(b) M is a P_0 matrix and a kind of nonsingularity holds at one of the accumulation points.

In this section, we have shown that LMM has a quadratic rate of convergence when M is a P_0 matrix and one of the accumulation points is a nondegenerate solution, which is a new result for LCP(M, q).

References

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

[2] Clarke, F.H.: Optimization and Nonsmooth Analysis. New York: John Wiley & Sons 1983.

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

[4] Fischer, A.: An NCP-function and its use for the solution of complementarity problems, in: Recent Advances in Nonsmooth Optimization (Du, D.Z., Qi, L., Womersley, R.S., eds.) pp. 88–105, Singapore: World Scientific 1995.

[5] Fischer, A.: On the local superlinear convergence of a Newton-type method for LCP under weak conditions. Optimization Methods and Software 6, 83–107 (1995).

[6] Fletcher, R.: Practical Methods of Optimization. New York: John Wiley & Sons 1987.

[7] Robinson, S.M.: Some continuity properties of polyhedral multifunctions. Mathematical Programming Study 14, 206–214 (1981).

[8] Tseng, P.: Error bounds and superlinear convergence analysis of some Newton-type methods in optimization, in: Applications and Algorithms of Complementarity (Ferris, M.C., Mangasarian, O.L., Pang, J.S., eds.), Boston: Kluwer Academic Publishers (to appear).

[9] Yamashita, N., Fukushima, M.: The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem. SIAM Journal on Optimization (to appear).

[10] Yamashita, N., Imai, J., Fukushima, M.: The proximal point algorithm for the P_0 complementarity problem, in: Applications and Algorithms of Complementarity (Ferris, M.C., Mangasarian, O.L., Pang, J.S., eds.), Boston: Kluwer Academic Publishers (to appear).