Convergence Properties of the Inexact Levenberg-Marquardt Method under Local Error Bound Conditions

Guidance

Professor	Masao FUKUSHIMA
Assistant Professor	Nobuo YAMASHITA

Hiroshige DAN

1999 Graduate Course

in

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2001

Acknowledgments

I would like to thank Professor Masao Fukushima for his helpful comments. I respect him as a great researcher and a gentle educator. I will not forget his advice. I would like to express much appreciation for Associate Professor Tetsuya Takine. His precise comments makes this paper better. Moreover, I would like to thank Assistant Professor Nobuo Yamashita for daily communication. Without it, I would not be able to write this paper at all.

Lastly, I dedicates this paper to my family, personally.

Abstract

In this paper, we consider convergence properties of the Levenberg-Marquardt method for solving nonlinear equations. It is well-known that the nonsingularity of Jacobian at a solution guarantees that the Levenberg-Marquardt method has a quadratic rate of convergence. Recently, Yamashita and Fukushima showed that the Levenberg-Marquardt method has a quadratic rate of convergence under the assumption of local error bound, which is milder than the nonsingularity of Jacobian. In this paper, we show that the inexact Levenberg-Marquardt method (ILMM), which does not require computing exact search directions, has a superlinear rate of convergence under the same assumption of local error bound. Moreover, we propose the ILMM with Armijo's stepsize rule that has global convergence under mild conditions.

Contents

1 Introduction

In this paper, we consider solving the system of nonlinear equations

$$F(x) = 0, \tag{1.1}$$

where $F : \Re^n \to \Re^m$ is a continuously differentiable function. To solve (1.1) is one of the most fundamental themes in engineering, economics and so on.

When $m = n$, we can use Newton's method to solve (1.1). As a search direction at a current point x^k, Newton's method uses a solution d^k of the system of linear equations

$$F'\left(x^k\right) d = -F\left(x^k\right), \tag{1.2}$$

where $F'\left(x^k\right)$ is the Jacobian of F at x^k. Newton's method has a rapid convergence property, and hence it is used extensively in many applications. But, to obtain a solution efficiently by Newton's method, an initial point has to be sufficiently close to a solution x^* and the Jacobian of F at the solution x^* has to be nonsingular. Besides these drawbacks, (1.2) may have no solution.

On the other hand, even if $m \neq n$ or there is no solution of (1.2), the system of linear equations

$$F'\left(x^k\right)^T F'\left(x^k\right) d = -F'\left(x^k\right)^T F\left(x^k\right) \tag{1.3}$$

always has a solution $\bar{d}^k$. The Gauss-Newton method uses the solution $\bar{d}^k$ as a search direction. However, like Newton's method, the Gauss-Newton method requires an initial point to be sufficiently close to a solution in order to ensure convergence to a solution. Note that we can use Moore-Penrose's generalized inverse [9] to compute a solution of (1.3) when $F'(x^k)$ is singular. However, Moore-Penrose's generalized inverse is somewhat intractable theoretically and numerically.

The Levenberg-Marquardt method (LMM) [1, 5, 8] is a modified Gauss-Newton method that is designed to overcome the above mentioned drawbacks of the Gauss-Newton method. The LMM uses a solution of the system of linear equations

$$\left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right) d = -F'\left(x^k\right)^T F\left(x^k\right) \tag{1.4}$$

as a search direction $\hat{d}^k$, where μ_k is a positive parameter and I is the identity matrix. Since $F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I$ is always positive definite, (1.4) has a unique solution. Moreover, $\hat{d}^k$ is a descent direction of ϕ at x^k, where $\phi : \Re^n \to \Re$ is a merit function defined by

$$\phi(x) = \frac{1}{2}\|F(x)\|^2. \tag{1.5}$$

In fact, if $\nabla\phi\left(x^k\right) \neq 0$, we obtain

$$\nabla\phi\left(x^k\right)^T \hat{d}^k = -\left(F'\left(x^k\right)^T F\left(x^k\right)\right)^T \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right)^{-1} \left(F'\left(x^k\right)^T F\left(x^k\right)\right) < 0.$$

Therefore, the LMM with Armijo's stepsize rule enjoys global convergence to a stationary point of ϕ. Especially, when $m = n$ and $F'(x^*)$ is nonsingular at a stationary point x^*, x^* is a solution of (1.1) because

$$0 = \nabla\phi\left(x^*\right) = F'\left(x^*\right)^T F\left(x^*\right).$$

Recently, Yamashita and Fukushima [20] showed that the LMM has a quadratic rate of convergence under the assumption that $\|F(x)\|$ provides a local error bound for (1.1), instead of the nonsingularity of $F'(x)$ at a solution. Recall that $\|F(x)\|$ is said to provide a local error bound on a neighborhood N of a solution of (1.1) if there exist positive constants b such that

$$b \operatorname{dist}(x, X^*) \leq \|F(x)\| \qquad \forall x \in N, \tag{1.6}$$

where X^* is the solution set of (1.1). Concrete examples of local error bounds can be found in [13]. When $m = n$ and $F'(x^*)$ is nonsingular at a solution x^*, x^* is a locally unique solution of (1.1) and $\|F(x)\|$ provides a local error bound for (1.1) in a neighborhood of x^* [4]. Moreover, $\|F(x)\|$ may provide a local error bound even if $F'(x)$ is singular at a solution. For example, let us consider the mapping $F : \Re^2 \to \Re^2$ defined by

$$F(x_1, x_2) = \left(e^{x_1 - x_2} - 1, (x_1 - x_2)\,(x_1 - x_2 - 2)\right)^T.$$

The solution set of (1.1) is given by $X^* = \left\{x \in \Re^2 \mid x_1 - x_2 = 0\right\}$, and hence $\operatorname{dist}(x, X^*) = \frac{\sqrt{2}}{2}\,|x_1 - x_2|$. It is easy to see that

$$F'(x_1, x_2) = \begin{pmatrix} e^{x_1 - x_2} & -e^{x_1 - x_2} \\ 2\,(x_1 - x_2 - 1) & -2\,(x_1 - x_2 - 1) \end{pmatrix},$$

and hence $F'(x_1, x_2)$ is singular at any solution. Moreover, we obtain

$$\begin{aligned} \|F(x)\| &= \sqrt{\left(e^{x_1 - x_2} - 1\right)^2 + (x_1 - x_2)^2\,(x_1 - x_2 - 2)^2} \\ &\geq |x_1 - x_2|\,|x_1 - x_2 - 2|\,. \end{aligned}$$

Consequently, when N is chosen as $N = \left\{x \in \Re^2 \;\middle|\; |x_1 - x_2| \leq 2 - \frac{\sqrt{2}}{2}\right\}$, the condition (1.6) holds for any $b \in (0, 1)$. Therefore, the condition that $\|F(x)\|$ provides a local error bound in a neighborhood of x^* is milder than the condition that $F'(x^*)$ is nonsingular.

In the LMM considered in [20], it is assumed that (1.4) is solved exactly at every iteration. For large-scale problems, however, it is expensive to solve (1.4) exactly, and hence it is often effective to use inexact methods that find an approximate solution satisfying some appropriate conditions. Therefore, in this paper, we consider the inexact Levenberg-Marquardt method (ILMM) that uses an approximate solution d^k of (1.4). Let the vector r^k be defined by

$$r^k := \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right) d^k + F'\left(x^k\right)^T F\left(x^k\right). \tag{1.7}$$

The vector r^k is a residual vector associated with an approximate solution d^k. Facchinei and Kanzow [5] showed that the ILMM converges superlinearly under the assumption that $\left\|r^k\right\|$ is sufficiently small and $F'\left(x^k\right)^T F'\left(x^k\right)$ is uniformly nonsingular.

In this paper, using techniques similar to [20], we show that the distance between the solution set and a sequence generated by the ILMM converges to 0 superlinearly under the assumption that $\|F(x)\|$ provides a local error bound in a neighborhood of a solution. Moreover, we propose the ILMM with Armijo's stepsize rule and show that the proposed algorithm enjoys global convergence.

This paper is organized as follows: In Section 2, we establish local convergence of the ILMM with unit step size under the local error bound assumption. In Section 3, we propose the ILMM with Armijo's stepsize rule and show that the algorithm has global convergence. In Section 4, we report numerical results. In Section 5, we make some concluding remarks and discuss future research topics.

2 Local Convergence

In this section, we discuss local convergence properties of the ILMM. Yamashita and Fukushima [20] consider a minimization problem equivalent to (1.4) and analyze local convergence of the LMM by using the properties of the minimization problem. In this paper, we use similar techniques to analyze the rate of convergence of the ILMM.

For each k, we define $\theta^k : \Re^n \to \Re$ by

$$\theta^k(d) = \left\|F'\left(x^k\right) d + F\left(x^k\right)\right\|^2 + \mu_k \|d\|^2, \tag{2.1}$$

and consider the minimization problem

$$\min_{d \in \Re^n} \quad \theta^k(d). \tag{2.2}$$

Since the first order optimality condition for (2.2) is given by (1.4) and θ^k is a strictly convex quadratic function, (1.4) is in fact equivalent to (2.2).

First, we make the following assumption on F, under which we will show a superlinear rate of convergence of the ILMM.

Assumption 2.1

(i) There exists a solution x^* of (1.1).

(ii) There exist constants $b_1 \in (0,1)$ and $c_1 \in (0,\infty)$ such that

$$\|F'(y)(x-y) - (F(x) - F(y))\| \le c_1 \|x-y\|^2 \quad \forall x, y \in B(x^*, b_1),$$

where $B(x^*, b_1) := \{x \in \Re^n \mid \|x - x^*\| \le b_1\}$.

(iii) $\|F(x)\|$ provides an error bound for (1.1) on $B(x^*, b_1)$, i.e., there exists a constant $c_2 > 0$ such that

$$c_2 \operatorname{dist}(x, X^*) \le \|F(x)\| \quad \forall x \in B(x^*, b_1),$$

where X^* is the solution set of (1.1).

In this paper, Assumption 2.1 (iii) plays the most important role instead of the nonsingularity of Jacobian. Note that Assumption 2.1 (ii) is satisfied if F' is Lipschitzian [12, Theorem 3.2.12]. Moreover, since F is continuously differentiable, $\|F'(y)\|$ is bounded on $B(x^*, b_1)$ and F is Lipschitzian on $B(x^*, b_1)$, i.e., there exists a constant $L > 0$ such that

$$\|F(x) - F(y)\| \le L\|x - y\| \quad \forall x, y \in B(x^*, b_1). \tag{2.3}$$

In what follows, $\bar{x}^k$ denotes an arbitrary vector such that

$$\left\|\bar{x}^k - x^k\right\| = \operatorname{dist}\left(x^k, X^*\right) \text{ and } \bar{x}^k \in X^*.$$

Note that such $\bar{x}^k$ always exists even though the set X^* need not be convex.

The following assumption is concerned with the choice of the parameters μ_k used in the ILMM.

Assumption 2.2 *For each k, the parameter μ_k is chosen to satisfy*

$$\mu_k = \left\|F(x^k)\right\|^\delta, \tag{2.4}$$

where δ is a constant such that $0 < \delta \le 2$.

Throughout this section, we suppose that the ILMM generates a sequence $\{x^k\}$ by

$$x^{k+1} := x^k + d^k, \tag{2.5}$$

where d^k is an approximate solution of (1.4).

Now we show the next lemma.

Lemma 2.1 *Suppose that Assumptions 2.1 and 2.2 hold. If $x^k \in B\left(x^*, \frac{b_1}{2}\right)$, then*

$$\left\|d^k\right\| \le c_3 \operatorname{dist}\left(x^k, X^*\right) + \frac{\left\|r^k\right\|}{\mu_k}, \tag{2.6}$$

$$\left\|F'\left(x^k\right) d^k + F\left(x^k\right)\right\| \le c_4 \operatorname{dist}\left(x^k, X^*\right)^{1+\frac{\delta}{2}} + \left\|F'\left(x^k\right)\right\| \frac{\left\|r^k\right\|}{\mu_k}, \tag{2.7}$$

where r^k is the residual vector given by (1.7), $c_3 = \sqrt{\frac{c_1^2}{c_2^\delta}\left(\frac{b_1}{2}\right)^{2-\delta} + 1}$ and $c_4 = \sqrt{c_1^2\left(\frac{b_1}{2}\right)^{2-\delta} + L^\delta}$.

Proof: Let $\hat{d}^k$ be the exact solution of (1.4). Because (1.4) is equivalent to (2.2), we have

$$\theta^k(\hat{d}^k) \le \theta^k\left(\bar{x}^k - x^k\right). \tag{2.8}$$

Moreover, since $x^k \in B\left(x^*, \frac{b_1}{2}\right)$, we have

$$\left\|\bar{x}^k - x^*\right\| \le \left\|\bar{x}^k - x^k\right\| + \left\|x^* - x^k\right\| \le \left\|x^* - x^k\right\| + \left\|x^* - x^k\right\| \le b_1,$$

and hence $\bar{x}^k \in B(x^*, b_1)$. It then follows from Assumptions 2.1 and 2.2 together with the condition (2.3) that

$$\mu_k = \left\|F\left(x^k\right)\right\|^\delta \ge c_2^\delta \left\|\bar{x}^k - x^k\right\|^\delta, \tag{2.9}$$

$$\mu_k = \left\|F\left(x^k\right)\right\|^\delta = \left\|F(\bar{x}^k) - F\left(x^k\right)\right\|^\delta \le L^\delta \left\|\bar{x}^k - x^k\right\|^\delta. \tag{2.10}$$

By the definition (2.1) of θ^k, we have

$$\begin{aligned} \|\hat{d}^k\|^2 &\le \frac{1}{\mu_k}\theta^k(\hat{d}^k) \\ &\le \frac{1}{\mu_k}\theta^k\left(\bar{x}^k - x^k\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mu_k}\left(\left\|F'\left(x^k\right)\left(\bar{x}^k - x^k\right) + F\left(x^k\right)\right\|^2 + \mu_k \left\|\bar{x}^k - x^k\right\|^2\right) \\
&\leq \frac{1}{\mu_k}\left(c_1^2 \left\|\bar{x}^k - x^k\right\|^4 + \mu_k \left\|\bar{x}^k - x^k\right\|^2\right) \\
&= \left(\frac{c_1^2 \left\|\bar{x}^k - x^k\right\|^2}{\mu_k} + 1\right)\left\|\bar{x}^k - x^k\right\|^2 \\
&\leq \left(\frac{c_1^2}{c_2^\delta}\left\|\bar{x}^k - x^k\right\|^{2-\delta} + 1\right)\left\|\bar{x}^k - x^k\right\|^2 \\
&\leq \left\{\frac{c_1^2}{c_2^\delta}\left(\frac{b_1}{2}\right)^{2-\delta} + 1\right\}\left\|\bar{x}^k - x^k\right\|^2,
\end{aligned}
$$

where the second inequality follows from (2.8), the third inequality follows from Assumption 2.1 (ii), the fourth inequality follows from (2.9), and the last inequality follows from the fact that

$$
\left\|\bar{x}^k - x^k\right\| \leq \left\|x^* - x^k\right\| \leq \frac{b_1}{2}. \tag{2.11}
$$

So we obtain

$$
\|\hat{d}^k\| \leq \sqrt{\frac{c_1^2}{c_2^\delta}\left(\frac{b_1}{2}\right)^{2-\delta} + 1}\;\left\|\bar{x}^k - x^k\right\|. \tag{2.12}
$$

Moreover, from (1.7), we have

$$
d^k = \hat{d}^k + \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right)^{-1} r^k.
$$

It then follows that

$$
\begin{aligned}
\left\|d^k\right\| &\leq \|\hat{d}^k\| + \left\|\left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right)^{-1}\right\| \left\|r^k\right\| \\
&\leq \|\hat{d}^k\| + \frac{\left\|r^k\right\|}{\mu_k}.
\end{aligned} \tag{2.13}
$$

Consequently, we obtain (2.6) with $c_3 = \sqrt{\frac{c_1^2}{c_2^\delta}\left(\frac{b_1}{2}\right)^{2-\delta} + 1}$ from (2.12) and (2.13).

Next we show (2.7). The left-hand side of (2.7) can be estimated as

$$
\begin{aligned}
\left\|F'\left(x^k\right) d^k + F\left(x^k\right)\right\| &= \left\|F'\left(x^k\right)\left(\hat{d}^k + \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right)^{-1} r^k\right) + F\left(x^k\right)\right\| \\
&\leq \left\|F'\left(x^k\right)\hat{d}^k + F\left(x^k\right)\right\| + \left\|F'\left(x^k\right)\right\| \left\|\left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right)^{-1}\right\| \left\|r^k\right\| \\
&\leq \left\|F'\left(x^k\right)\hat{d}^k + F\left(x^k\right)\right\| + \left\|F'\left(x^k\right)\right\| \frac{\left\|r^k\right\|}{\mu_k}.
\end{aligned} \tag{2.14}
$$

It then follows from (2.1), (2.8), Assumption 2.1 (ii), (2.10), and (2.11) that

$$
\begin{aligned}
\|F'\left(x^k\right)\hat{d}^k + F\left(x^k\right)\|^2 &\leq \theta^k(\hat{d}^k) \;\leq\; \theta^k\left(\bar{x}^k - x^k\right) \\
&\leq c_1^2 \left\|\bar{x}^k - x^k\right\|^4 + \mu_k \left\|\bar{x}^k - x^k\right\|^2 \\
&\leq c_1^2\left(\frac{b_1}{2}\right)^{2-\delta}\left\|\bar{x}^k - x^k\right\|^{2+\delta} + L^\delta \left\|\bar{x}^k - x^k\right\|^{2+\delta} \\
&= \left\{c_1^2\left(\frac{b_1}{2}\right)^{2-\delta} + L^\delta\right\}\left\|\bar{x}^k - x^k\right\|^{2+\delta}.
\end{aligned}
$$

Therefore, the first term of (2.14) can be estimated as

$$\left\|F'\left(x^k\right)\hat{d}^k+F\left(x^k\right)\right\| \le \sqrt{c_1^2\left(\frac{b_1}{2}\right)^{2-\delta}+L^\delta}\ \left\|\bar{x}^k-x^k\right\|^{1+\frac{\delta}{2}},$$

which yields the desired inequality (2.7). □

Now we give a condition on $\|r^k\|$ for superlinear convergence of the ILMM.

Assumption 2.3 *The residual vector r^k given by (1.7) satisfies*

$$\frac{\|r^k\|}{\mu_k}=o\left(\operatorname{dist}\left(x^k,X^*\right)\right),$$

where $o(\cdot)$ means $\lim_{t\to 0}\frac{o(t)}{t}=0$.

We have the next lemma from Assumption 2.3 directly.

Lemma 2.2 *Suppose that Assumption 2.3 holds. Then there exists a constant $b_2>0$ such that*

$$\operatorname{dist}\left(x^k,X^*\right)\le b_2 \implies \frac{\|r^k\|}{\mu_k}\le \operatorname{dist}\left(x^k,X^*\right).$$

By using Lemma 2.1, we show a key lemma of our analysis.

Lemma 2.3 *Suppose that Assumptions 2.1, 2.2 and 2.3 hold. If $x^k\in B\left(x^*,\frac{b_1}{2}\right)$, then*

$$\operatorname{dist}\left(x^{k+1},X^*\right)=o\left(\operatorname{dist}\left(x^k,X^*\right)\right).$$

In particular, there exists a constant $b_3>0$ such that

$$\operatorname{dist}\left(x^k,X^*\right)\le b_3 \implies \operatorname{dist}\left(x^{k+1},X^*\right)\le\frac{1}{2}\operatorname{dist}\left(x^k,X^*\right).$$

Proof: It follows from Lemma 2.1, Assumption 2.3 and the boundedness of $\|F'(x)\|$ on $B\left(x^*,\frac{b_1}{2}\right)$ that

$$\begin{aligned}\left\|d^k\right\| &\le c_3\operatorname{dist}\left(x^k,X^*\right)+o\left(\operatorname{dist}\left(x^k,X^*\right)\right), &(2.15)\\ \left\|F'\left(x^k\right)d^k+F\left(x^k\right)\right\| &\le c_4\operatorname{dist}\left(x^k,X^*\right)^{1+\frac{\delta}{2}}+\left\|F'\left(x^k\right)\right\|\cdot o\left(\operatorname{dist}\left(x^k,X^*\right)\right) &(2.16)\\ &= o\left(\operatorname{dist}\left(x^k,X^*\right)\right).\end{aligned}$$

Then we obtain

$$\begin{aligned}\operatorname{dist}\left(x^{k+1},X^*\right) &= \left\|\bar{x}^{k+1}-x^{k+1}\right\|\\ &\le \frac{1}{c_2}\left\|F(x^k+d^k)\right\|\\ &\le \frac{1}{c_2}\left\|F'\left(x^k\right)d^k+F\left(x^k\right)\right\|+\frac{c_1}{c_2}\left\|d^k\right\|^2\\ &\le o\left(\operatorname{dist}\left(x^k,X^*\right)\right)+\frac{c_1}{c_2}\left\{c_3\operatorname{dist}\left(x^k,X^*\right)+o\left(\operatorname{dist}\left(x^k,X^*\right)\right)\right\}^2\\ &= o\left(\operatorname{dist}\left(x^k,X^*\right)\right),\end{aligned}$$

where the first inequality follows from Assumption 2.1 (iii) and (2.5), the second inequality follows from Assumption 2.1 (ii), and the last inequality follows from (2.15) and (2.16). This completes the proof. □

Lemma 2.3 shows that $\left\{\mathrm{dist}\left(x^k, X^*\right)\right\}$ is convergent to 0 superlinearly if $x^k \in B\left(x^*, \frac{b_1}{2}\right)$ for all k. Now we give a sufficient condition for $x^k \in B\left(x^*, \frac{b_1}{2}\right)$ for all k.

Lemma 2.4 *Suppose that Assumptions 2.1, 2.2 and 2.3 hold. Let* $\bar{b} := \min\left\{\frac{b_1}{2}, b_2, b_3\right\}$ *and* $e := \frac{\bar{b}}{3+2c_3}$. *If* $x^0 \in B\left(x^*, e\right)$, *then* $x^k \in B\left(x^*, \bar{b}\right) \subseteq B\left(x^*, \frac{b_1}{2}\right)$ *for all* k.

Proof: We prove the lemma by induction.

First we consider the case $k = 0$. Since $e < \bar{b} \le \frac{b_1}{2}$, we have $x^0 \in B\left(x^*, \frac{b_1}{2}\right)$. It then follows from Lemma 2.1 that

$$\begin{aligned}\left\|x^1 - x^*\right\| &= \left\|x^0 + d^0 - x^*\right\| \le \left\|x^* - x^0\right\| + \left\|d^0\right\| \\ &\le \left\|x^* - x^0\right\| + c_3 \,\mathrm{dist}\left(x^0, X^*\right) + \frac{\left\|r^0\right\|}{\mu_0}.\end{aligned}$$

Since

$$\mathrm{dist}\left(x^0, X^*\right) \le \left\|x^* - x^0\right\| \le e \le b_2,$$

we have from Lemma 2.2

$$\begin{aligned}\left\|x^1 - x^*\right\| &\le \left\|x^* - x^0\right\| + c_3 \,\mathrm{dist}\left(x^0, X^*\right) + \mathrm{dist}\left(x^0, X^*\right) \\ &\le \left\|x^* - x^0\right\| + c_3 \left\|x^* - x^0\right\| + \left\|x^* - x^0\right\| \\ &\le (2 + c_3)\, e \le \frac{2 + c_3}{3 + 2c_3}\,\bar{b} \;\le\; \bar{b}.\end{aligned}$$

Next we consider the case $k \ge 1$. Suppose that $x^l \in B\left(x^*, \bar{b}\right),\; l = 1, \ldots, k$. Since $e < \bar{b}$, we have $x^0 \in B\left(x^*, \bar{b}\right)$. It follows from $\mathrm{dist}\left(x^l, X^*\right) \le \left\|x^* - x^l\right\| \le \bar{b} \le b_3$ and Lemma 2.3 that

$$\begin{aligned}\mathrm{dist}\left(x^l, X^*\right) &\le \frac{1}{2}\,\mathrm{dist}\left(x^{l-1}, X^*\right) \le \cdots \le \left(\frac{1}{2}\right)^l \mathrm{dist}\left(x^0, X^*\right) \le \left(\frac{1}{2}\right)^l \left\|x^* - x^0\right\| \\ &\le \left(\frac{1}{2}\right)^l e \qquad\qquad 0 \le \forall l \le k.\end{aligned}$$

Since $x^l \in B\left(x^*, \frac{b_1}{2}\right) \cap B\left(x^*, b_2\right),\; l = 0, \ldots, k$, Lemmas 2.1 and 2.2 yield

$$\left\|d^l\right\| \le (c_3 + 1)\,\mathrm{dist}\left(x^l, X^*\right) \le (c_3 + 1)\left(\frac{1}{2}\right)^l e. \tag{2.17}$$

Consequently, we obtain

$$\begin{aligned}\left\|x^{k+1} - x^*\right\| &\le \left\|x^0 - x^*\right\| + \sum_{l=0}^{k}\left\|d^l\right\| \\ &\le \left\{1 + 2\left(c_3 + 1\right)\right\} e = \bar{b}.\end{aligned}$$

This completes the proof. □

From these lemmas, we can show the next theorem, which is the main result in the paper.

Theorem 2.1 *Suppose that Assumptions 2.1, 2.2 and 2.3 hold. Let* e *and* $\bar{b}$ *be the constants given in Lemma 2.4, and* $\{x^k\}$ *be a sequence generated by the ILMM with* $x^0 \in B(x^*, e)$ *and (2.5). Then,* $\left\{\text{dist}\left(x^k, X^*\right)\right\}$ *converges to* 0 *superlinearly. Moreover,* $\left\{x^k\right\}$ *converges to a solution* $\hat{x} \in B\left(x^*, \bar{b}\right)$.

Proof: We obtain the first half of the theorem from Lemmas 2.3 and 2.4 directly. So, we only show that $\left\{x^k\right\}$ converges to $\hat{x} \in B\left(x^*, \bar{b}\right)$. For this purpose, we only have to show that $\left\{x^k\right\}$ is a convergent sequence because $\left\{x^k\right\} \subset B\left(x^*, \bar{b}\right)$ and $\left\{\text{dist}\left(x^k, X^*\right)\right\}$ converges to 0.

Note that we have

$$\left\|d^l\right\| \le (c_3 + 1)\left(\frac{1}{2}\right)^l e \quad \forall l \ge 0$$

from (2.17) in the proof of Lemma 2.4. Therefore, for all integers $p > q \ge 0$, we obtain

$$\begin{aligned} \|x^p - x^q\| &\le \sum_{l=q}^{p-1} \left\|d^l\right\| \le (c_3+1)\, e \sum_{l=q}^{p-1} \left(\frac{1}{2}\right)^l \le (c_3+1)\, e \sum_{l=q}^{\infty} \left(\frac{1}{2}\right)^l \\ &\le (c_3+1)\, e \left(\frac{1}{2}\right)^{q-1}. \end{aligned}$$

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

Note that Theorem 2.1 does not say that a sequence $\left\{x^k\right\}$ generated by the ILMM converges to $\hat{x}$ superlinearly. Moreover, this theorem does not show the quadratic convergence, though the exact LMM converges quadratically under the same conditions and $r^k = 0$ for all k [20]. However, we can show that the ILMM has a quadratic rate of convergence if we assume a more restrictive condition than Assumption 2.3.

Theorem 2.2 *Suppose that Assumptions 2.1 and 2.2 hold. Suppose also that* $\delta = 2$ *and*

$$\frac{\left\|r^k\right\|}{\mu_k} = O\left(\text{dist}\left(x^k, X^*\right)^2\right) \tag{2.18}$$

holds, where $O(\cdot)$ *means* $\lim_{t\to 0} \frac{O(t)}{t} < \infty$. *Let* e *and* $\bar{b}$ *be the constants defined in Lemma 2.4 and* $\{x^k\}$ *be a sequence generated by the ILMM with* $x^0 \in B(x^*, e)$ *and (2.5). Then,* $\left\{\text{dist}\left(x^k, X^*\right)\right\}$ *is convergent to* 0 *quadratically. Moreover,* $\left\{x^k\right\}$ *converges to a solution* $\hat{x} \in B\left(x^*, \bar{b}\right)$.

Proof: From (2.18), we immediately obtain

$$\text{dist}\left(x^{k+1}, X^*\right) = O\left(\text{dist}\left(x^k, X^*\right)^2\right),$$

using proof techniques similar to Lemma 2.3. Then, in a way similar to Theorem 2.1, we can show this theorem. □

3 Global Convergence

In the previous section, we considered the rate of convergence of the ILMM with unit stepsize. In this section, we propose the ILMM with Armijo's stepsize rule and show that it has global convergence.

We propose the following algorithm, which uses the merit function ϕ defined by (1.5).

Algorithm 3.1

Step 0: *Choose parameters* $\alpha \in (0,1), \beta \in (0,1), \gamma \in (0,1), \rho \in (0,1), \delta \in (0,2], p > 0$, *and an initial point* $x^0 \in \Re^n$. *Set* $\mu_0 = \|F(x^0)\|^\delta$ *and* $k := 0$.

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

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

$$\left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right) d = -F'\left(x^k\right)^T F\left(x^k\right). \tag{3.1}$$

If the condition

$$\left\|F\left(x^k + d^k\right)\right\| \le \gamma \left\|F\left(x^k\right)\right\| \tag{3.2}$$

is satisfied, then set $x^{k+1} := x^k + d^k$ *and go to* Step 4.

Step 3: *If*

$$\nabla\phi\left(x^k\right)^T d^k \le -\rho \left\|d^k\right\|^p \tag{3.3}$$

is not satisfied, set $d^k = -\nabla\phi\left(x^k\right)$. *Find the smallest nonnegative integer* m *such that*

$$\phi\left(x^k + \beta^m d^k\right) - \phi\left(x^k\right) \le \alpha\beta^m \nabla\phi\left(x^k\right)^T d^k,$$

and set $x^{k+1} := x^k + \beta^m d^k$.

Step 4: *Set* $\mu_{k+1} = \left\|F(x^{k+1})\right\|^\delta$ *and* $k := k+1$. *Go to* Step 1. □

In Step 3 of Algorithm 3.1, we must check whether a search direction d^k satisfies (3.3) or not. This is because d^k is not an exact solution of (3.1) in general and hence d^k may not be a good descent direction of the merit function ϕ. If d^k is not a good search direction, we reset d^k to be the steepest descent direction of ϕ. Consequently, d^k is always a sufficient descent direction of the merit function, and hence the line search in Step 3 is well-defined.

Now, we can prove the following global convergence theorem for Algorithm 3.1 by using techniques similar to [5].

Theorem 3.1 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 3.1. If the residual vector* r^k *given by (1.7) satisfies the condition*

$$\left\|r^k\right\| \le \min\left\{\eta \left\|F'\left(x^k\right)^T F\left(x^k\right)\right\|,\ \nu_k \left\|F'\left(x^k\right)^T F\left(x^k\right)\right\|^\delta\right\}, \tag{3.4}$$

where $\eta \in (0,1)$ *and* $\nu_k = o\left(\mathrm{dist}\left(x^k, X^*\right)\right)$, *then any accumulation point of* $\left\{x^k\right\}$ *is a stationary point of* ϕ. *Moreover, if an accumulation point* x^* *of* $\left\{x^k\right\}$ *is a solution of (1.1) that satisfies Assumption 2.1, then* $\left\{\mathrm{dist}\left(x^k, X^*\right)\right\}$ *converges to* 0 *superlinearly.*

Proof: First we show that any accumulation point of $\{x^k\}$ is a stationary point of ϕ. Let $K_1 := \{k \mid \|F(x^k + d^k)\| \le \gamma \|F(x^k)\|\}$. If K_1 is an infinite set, then we have $\|F(x^k)\| \to 0$ as $k \to \infty$ because $\{\|F(x^k)\|\}$ is a monotonically decreasing sequence. This shows that any accumulation point of $\{x^k\}$ is a stationary point of ϕ. If K_1 is finite, then, without loss of generality, we can assume that $\|F(x^k + d^k)\| > \gamma \|F(x^k)\|$ for all k. We will show that $\{d^k\}$ is gradient related to $\{x^k\}$, i.e., for any subsequence $\{x^k\}_{k\in K}$ which converges to a nonstationary point of ϕ, $\{d^k\}_{k\in K}$ satisfies

$$\limsup_{k\to\infty,\ k\in K} \|d^k\| < \infty, \tag{3.5}$$

$$\liminf_{k\to\infty,\ k\in K} \nabla\phi(x^k)^T d^k < 0. \tag{3.6}$$

Then we can conclude that any accumulation point of $\{x^k\}$ is a stationary point of the merit function ϕ [1, Proposition 1.2.1]. Let $\{x^k\}_{k\in K}$ be a subsequence which converges to a nonstationary point of ϕ, and let $K_2 := \{k \in K \mid d^k = -\nabla\phi(x^k)\}$. If K_2 is an infinite set, $\{d^k\}$ is obviously gradient related, and hence $\{x^k\}$ must converge to a stationary point of the merit function ϕ. If K_2 is a finite set, then we may assume, without loss of generality, that d^k are approximate solutions of (1.4) with residuals r^k satisfying (1.7) for all k.

First we show that (3.5) holds. To this end, we consider two cases for the sequence $\{\mu_k\}$: (i) $\liminf_{k\to\infty,k\in K} \mu_k = 0$ and (ii) $\liminf_{k\to\infty,k\in K} \mu_k > 0$.

Case (i): In this case, without loss of generality, we assume that $\mu_k \to 0$ $(k \in K)$. Then, we have $\|F(x^k)\| \to 0$ $(k \in K)$. Since $\{\|F'(x^k)\|\}$ is bounded on any convergent subsequence $\{x^k\}_{k\in K}$, we obtain that $\nabla\phi(x^k) = F'(x^k)^T F(x^k) \to 0$ $(k \in K)$. This contradicts the assumption that $\{x^k\}_{k\in K}$ converges to a nonstationary point. Therefore this case does not occur.

Case (ii): In this case, there exists a constant ξ such that $\mu_k \ge \xi > 0$ for all k. Since $\left\{\left\|F'(x^k)^T F(x^k)\right\|\right\}$ is bounded on any convergent subsequence $\{x^k\}_{k\in K}$, it follows from (1.7) and (3.4) that

$$\begin{aligned} \|d^k\| &\le \left\|\left(F'(x^k)^T F'(x^k) + \mu_k I\right)^{-1}\right\| \left(\left\|F'(x^k)^T F(x^k)\right\| + \|r^k\|\right) \\ &\le \frac{1+\eta}{\xi} \left\|F'(x^k)^T F(x^k)\right\| < \infty. \end{aligned}$$

Therefore (3.5) holds.

Next, we show that (3.6) holds. Suppose to the contrary that (3.6) does not hold, i.e.,

$$\liminf_{k\to\infty,k\in K} \nabla\phi(x^k)^T d^k = 0.$$

It then follows from (3.3) that there exists an infinite set $K_3 \subset K$ such that

$$\lim_{k\to\infty,\ k\in K_3} d^k = 0.$$

On the other hand, we have from (1.7)

$$\|\nabla\phi(x^k) - r^k\| = \left\|F'(x^k)^T F(x^k) - r^k\right\|$$

$$\begin{aligned}
&\le \left\| \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I \right) d^k \right\| \\
&\le \left\| F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I \right\| \left\| d^k \right\|.
\end{aligned}$$

Since $\limsup_{k\to\infty, k\in K_3} \left\| F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I \right\| < \infty$ and $\lim_{k\to\infty,\ k\in K_3} d^k = 0$, we have

$$\lim_{k\to\infty,\ k\in K_3} \left\| \nabla\phi\left(x^k\right) - r^k \right\| = 0. \tag{3.7}$$

Moreover, from (3.4), we have $\left\|r^k\right\| \le \eta \left\| F'\left(x^k\right)^T F\left(x^k\right) \right\| = \eta \left\| \nabla\phi\left(x^k\right) \right\|$. Then we can deduce from (3.7) that

$$\lim_{k\to\infty,\ k\in K_3} \left\| \nabla\phi\left(x^k\right) \right\| = \lim_{k\to\infty,\ k\in K_3} \left\| r^k \right\| = 0.$$

This contradicts the assumption that the limit point of $\left\{x^k\right\}_{k\in K}$ is not a stationary point. Therefore, we have (3.6), and hence $\{d^k\}$ is gradient related to $\{x^k\}$. Consequently we conclude that any accumulation point of $\left\{r^k\right\}$ is a stationary point of the merit function ϕ.

Now, we proceed to showing the latter half of the theorem. From Theorem 2.1, it is sufficient to show that Assumption 2.3 is satisfied and that $x^{k+1} = x^k + d^k$ with d^k being determined from (3.1) for all k large enough. Since $\left\|r^k\right\|$ satisfies (3.4), we have for any convergent subsequence $\left\{x^k\right\}_{k\in K}$

$$\begin{aligned}
\frac{\left\|r^k\right\|}{\mu_k} &\le \frac{\nu_k \left\| F'\left(x^k\right)^T F\left(x^k\right) \right\|^\delta}{\left\| F\left(x^k\right) \right\|^\delta} \\
&\le \nu_k \left\| F'\left(x^k\right) \right\|^\delta = o\left(\operatorname{dist}\left(x^k, X^*\right)\right),
\end{aligned}$$

where the last equality follows from the boundedness of $\left\{\left\|F'\left(x^k\right)\right\|\right\}$ and the definition of $\{\nu_k\}$. This means that Assumption 2.3 holds. Let x^* be an accumulation point of $\left\{x^k\right\}$, and $\left\{x^k\right\}_{k\in K}$ be a subsequence such that $\lim_{k\in K} x^k = x^*$. Then there exists $\bar{k} \in K$ such that

$$\left\| x^k - x^* \right\| \le e,\ \forall k \ge \bar{k}, k \in K,$$

where e is given in Lemma 2.4. Moreover, by choosing larger $\bar{k}$ if necessary, we can show from Lemma 2.3 that a sequence $\left\{y^l\right\}$ generated by the ILMM with $y^0 = x^{\bar{k}}$ and unit step size satisfies

$$\left\| \bar{y}^{l+1} - y^{l+1} \right\| \le \frac{c_2\gamma}{L} \left\| \bar{y}^l - y^l \right\| \quad \forall l, \tag{3.8}$$

where $\bar{y}^l$ denotes one of the nearest solutions from y^l, that is, $\bar{y}^l \in X^*$ and $\left\|\bar{y}^l - y^l\right\| = \operatorname{dist}\left(y^l, X^*\right)$. (Note that there may be more than one nearest solutions to y^l, since the solution set X^* need not be convex.) It then follows that

$$\begin{aligned}
\left\| F(y^{l+1}) \right\| &= \left\| F\left(\bar{y}^{l+1}\right) - F\left(y^{l+1}\right) \right\| \\
&\le L \left\| \bar{y}^{l+1} - y^{l+1} \right\| \\
&\le c_2\gamma \left\| \bar{y}^l - y^l \right\| \\
&\le \gamma \left\| F\left(y^l\right) \right\|,
\end{aligned}$$

where the first inequality follows from (2.3), the second inequality follows from (3.8), and the last inequality follows from Assumption 2.1 (iii). Hence, (3.2) is satisfied for $k \geq \bar{k}$, and we obtain $x^{k+1} = x^k + d^k$ for $k \geq \bar{k}$, where d^k is determined from (3.1). This completes the proof. □

Remark 3.1 *As an updating rule of ν_k which satisfies the assumption in Theorem 3.1, we may employ the rule*

$$\nu_k = \left\| F\left(x^k\right) \right\|^\tau ,$$

where τ is a constant such that $\tau > 1$. In this case, it follows from (2.3) that

$$\begin{aligned} \nu_k &= \left\| F\left(x^k\right) \right\|^\tau = \left\| F\left(\bar{x}^k\right) - F\left(x^k\right) \right\|^\tau \\ &\leq L^\tau \left\| \bar{x}^k - x^k \right\|^\tau = o\left(\operatorname{dist}\left(x^k, X^*\right)\right) \end{aligned}$$

when k is sufficiently large.

In Theorem 3.1, we have established the superlinear convergence of Algorithm 3.1 by using Theorem 2.1. Using Theorem 2.2, we can also give conditions for a quadratic convergence of Algorithm 3.1.

Theorem 3.2 *Let $\left\{x^k\right\}$ be a sequence generated by Algorithm 3.1 with $\delta = 2$. If the residual vector r^k given by (1.7) satisfies*

$$\left\|r^k\right\| \leq \min\left\{ \eta \left\| F'\left(x^k\right)^T F\left(x^k\right) \right\|,\ \nu_k \left\| F'\left(x^k\right)^T F\left(x^k\right) \right\|^\delta \right\}, \tag{3.9}$$

where $\eta \in (0,1)$ and $\nu_k = O\left(\operatorname{dist}\left(x^k, X^*\right)^2\right)$, then any accumulation point of $\left\{x^k\right\}$ is a stationary point of ϕ. Moreover, if an accumulation point x^* of $\left\{x^k\right\}$ is a solution of (1.1) that satisfies Assumption 2.1, then $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$ converges to 0 quadratically.

Proof: It is easy to verify that the assumptions of Theorem 3.1 hold, so any accumulation point of $\left\{x^k\right\}$ is a stationary point of ϕ by Theorem 3.1. Moreover, since $\left\|r^k\right\|$ satisfies (3.9), we have

$$\frac{\left\|r^k\right\|}{\mu_k} \leq \frac{\nu_k \left\| F'\left(x^k\right)^T F\left(x^k\right) \right\|^\delta}{\left\| F\left(x^k\right) \right\|^\delta} \leq \nu_k \left\| F'\left(x^k\right) \right\|^\delta = O\left(\operatorname{dist}\left(x^k, X^*\right)^2\right),$$

and hence (2.18) is satisfied. Then, in a way similar to Theorem 3.1, we can show the quadratic convergence of $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$. □

4 Numerical Results

In this section, we discuss implementation issues of Algorithm 3.1 proposed in Section 3 and report some numerical results.

In implementing Algorithm 3.1, the most expensive task is to compute the search direction d^k by solving (3.1). Since the coefficient matrix $F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I$ of the linear equation (3.1) is always positive definite, we could find the exact solution of (3.1) by means of Cholesky factorization.

However, our algorithm does not require the exact solution of (3.1), that is, the search direction d^k has only to satisfy the approximate condition (3.4). In our numerical experiments, we employ the conjugate gradient method (CGM) [1, 11] to find d^k satisfying the approximate condition (3.4) from the following two reasons. First, the CGM can find an approximate solution of (3.1) with any accuracy. So, when the approximate condition (3.4) is mild, the CGM may find d^k in a small number of iterations. Second, the CGM is suitable for large-scale problems. At each iteration of the CGM for (3.1), we need to calculate

$$-\left(F'\left(x^k\right)^T F'\left(x^k\right)+\mu_k I\right)\bar{d}^{j-1}-F'\left(x^k\right)^T F\left(x^k\right), \tag{4.1}$$

where $\bar{d}^{j-1}$ is the search direction used in the previous iteration. This is done by calculating $\bar{v}^j = F'\left(x^k\right)\bar{d}^{j-1}$ first, and then $F'\left(x^k\right)^T \bar{v}^j + \mu_k \bar{d}^{j-1}$. Thus, the calculation of (4.1) is inexpensive if $F'\left(x^k\right)$ is sparse.

Now, we state some practical modifications of Algorithm 3.1. If an iterative point x^k is far from the solution set, the value of parameter μ_k determined by the rule (2.4) may become exceedingly large. In this case, a search direction d^k obtained from (3.1) is close to the steepest descent direction for the function ϕ, and hence it is likely that the algorithm converges slowly. To prevent this difficulty, we modify the updating rule for μ_k as

$$\mu_k = \min\left\{\left\|F\left(x^k\right)\right\|^{\delta}, \zeta\right\}, \tag{4.2}$$

where $\zeta > 0$ is an appropriate constant. We can expect that, even if $\left\|F(x^k)\right\|$ is large, the rule (4.2) enables us to find a good approximation to the Gauss-Newton direction. Next we consider the criterion for approximate solution of the linear equation (3.1). If an iterative point is far from the solution set, then the approximate condition (3.4) for r^k need not be tight, and hence a computed search direction may also approximate the steepest descent direction, because we use the CGM to solve (3.1). In view of this fact, we modify the approximate condition (3.4) as

$$\left\|r^k\right\| \le \min\left\{\eta\left\|F'\left(x^k\right)^T F\left(x^k\right)\right\|,\ \nu_k\left\|F'\left(x^k\right)^T F\left(x^k\right)\right\|^{\delta}, \kappa\sqrt{n}\right\}, \tag{4.3}$$

where $\kappa > 0$ is an appropriate constant. This condition ensures that, even if an iterative point x^k is far from the solution set, $\left\|r^k\right\|$ is smaller than $\kappa\sqrt{n}$, and hence, especially in the early stage of the iterations, we can expect that a search direction becomes a good approximation to the Gauss-Newton direction.

We have experimented on the following problems with $n = 100, 1000$ and 10000.

$$\begin{array}{llll}
\text{Problem 1}: & F:\Re^n \to \Re^n, & F_i(x) = \sqrt{i}\,(x_i - i), & i = 1,\dots,n \\
\text{Problem 2}: & F:\Re^n \to \Re^{\frac{n}{2}}, & F_i(x) = \sqrt{i}\,(x_i + x_{\frac{n}{2}+i} - i), & i = 1,\dots,\frac{n}{2} \\
\text{Problem 3}: & F:\Re^n \to \Re^n, & F_i(x) = x_i^2 - i, & i = 1,\dots,n \\
\text{Problem 4}: & F:\Re^n \to \Re^{\frac{n}{2}}, & F_i(x) = (x_i + x_{\frac{n}{2}+i})^2 - i, & i = 1,\dots,\frac{n}{2}
\end{array}$$

Clearly these problems have a solution. Moreover, $\|F(x)\|$ provides a local error bound in a neighborhood of the solution set. Note that $F'\left(x^k\right)^T F'\left(x^k\right)$ is singular for Problems 2 and 4.

Table 1: Results for Problems 1, 2, 3 and 4

Problem	n	i.p.	# iter.	# iter. (CGM)	time(sec.)	$\left\|F\left(x^k\right)\right\|$
Problem 1	100	$x^{0,1}$	3	154	3e-02	9.0e-02, 4.1e-04, 3.3e-08
		$x^{0,2}$	4	238	4e-02	2.0e-03, 4.1e-07, 7.6e-14
		$x^{0,3}$	4	238	4e-02	2.0e-03, 3.8e-07, 4.3e-14
		$x^{0,4}$	4	239	5e-02	1.9e-03, 3.9e-07, 4.7e-14
	1000	$x^{0,1}$	4	780	1.8e+00	4.3e-03, 1.1e-06, 7.8e-13
		$x^{0,2}$	4	784	1.8e+00	4.3e-03, 1.4e-06, 1.6e-12
		$x^{0,3}$	4	780	1.7e+00	4.4e-03, 1.1e-06, 4.4e-13
		$x^{0,4}$	4	763	1.7e+00	4.1e-03, 1.3e-06, 1.2e-12
	10000	$x^{0,1}$	4	2389	8.1e+01	9.2e-03, 5.4e-06, 2.7e-11
		$x^{0,2}$	4	2376	8.1e+01	1.3e-02, 1.0e-05, 1.0e-10
		$x^{0,3}$	4	2346	8.0e+01	9.5e-03, 5.5e-06, 2.8e-11
		$x^{0,4}$	4	2350	8.0e+01	1.3e-02, 1.0e-05, 1.0e-10
Problem 2	100	$x^{0,1}$	3	107	1e-02	9.5e-02, 4.8e-04, 1.8e-08
		$x^{0,2}$	4	160	2e-02	1.7e-03, 1.5e-07, 4.0e-15
		$x^{0,3}$	4	159	3e-02	1.7e-03, 1.4e-07, 3.2e-15
		$x^{0,4}$	4	160	3e-02	1.7e-03, 1.4e-07, 3.4e-15
	1000	$x^{0,1}$	3	345	7.2e-01	1.2e+00, 3.1e-03, 3.2e-07
		$x^{0,2}$	4	584	1.1e+00	3.2e-03, 3.9e-07, 5.6e-14
		$x^{0,3}$	4	580	1.1e+00	3.3e-03, 3.3e-07, 2.1e-14
		$x^{0,4}$	4	584	1.2e+00	3.2e-03, 3.7e-07, 3.8e-14
	10000	$x^{0,1}$	4	1798	5.1e+01	6.6e-03, 1.4e-06, 9.2e-13
		$x^{0,2}$	4	1794	5.0e+01	8.1e-03, 2.6e-06, 3.3e-12
		$x^{0,3}$	4	1770	5.0e+01	6.7e-03, 1.5e-06, 1.0e-12
		$x^{0,4}$	4	1735	5.0e+01	8.1e-03, 2.6e-06, 3.3e-12
Problem 3	100	$x^{0,1}$	9	239	5e-02	2.5e-02, 1.5e-04, 1.1e-08
		$x^{0,2}$	10	243	5e-02	2.5e-02, 1.5e-04, 1.1e-08
		$x^{0,3}$	9	235	5e-02	2.5e-02, 1.5e-04, 1.1e-08
		$x^{0,4}$	10	239	6e-02	2.5e-02, 1.5e-04, 1.1e-08
	1000	$x^{0,1}$	13	1033	3.2e+00	1.1e-03, 6.0e-07, 1.8e-12
		$x^{0,2}$	14	1038	3.3e+00	1.1e-03, 6.0e-07, 1.8e-12
		$x^{0,3}$	13	1017	3.1e+00	1.1e-03, 5.9e-07, 1.8e-12
		$x^{0,4}$	14	1010	3.3e+00	1.1e-03, 5.9e-07, 1.8e-12
	10000	$x^{0,1}$	16	2822	2.2e+02	5.9e-03, 9.8e-06, 8.3e-11
		$x^{0,2}$	17	2827	2.3e+02	5.9e-03, 9.8e-06, 8.3e-11
		$x^{0,3}$	16	2771	2.2e+02	5.8e-03, 9.6e-06, 8.2e-11
		$x^{0,4}$	17	2775	2.3e+02	5.8e-03, 9.6e-06, 8.2e-11
Problem 4	100	$x^{0,1}$	10	173	3e-02	2.5e-02, 1.5e-04, 8.0e-09
		$x^{0,2}$	11	176	4e-02	2.5e-02, 1.5e-04, 8.0e-09
		$x^{0,3}$	10	169	3e-02	2.5e-02, 1.5e-04, 8.0e-09
		$x^{0,4}$	11	172	4e-02	2.5e-02, 1.5e-04, 8.0e-09
	1000	$x^{0,1}$	14	753	2.1e+00	1.1e-03, 4.5e-07, 7.4e-13
		$x^{0,2}$	15	757	2.1e+00	1.1e-03, 4.5e-07, 7.4e-13
		$x^{0,3}$	14	744	2.0e+00	1.1e-03, 4.5e-07, 7.3e-13
		$x^{0,4}$	15	747	2.1e+00	1.1e-03, 4.5e-07, 7.3e-13
	10000	$x^{0,1}$	17	2058	1.3e+02	5.8e-03, 8.9e-06, 3.9e-11
		$x^{0,2}$	18	2062	1.4e+02	5.8e-03, 8.9e-06, 3.9e-11
		$x^{0,3}$	17	2032	1.3e+02	5.8e-03, 8.8e-06, 3.8e-11
		$x^{0,4}$	18	2036	1.4e+02	5.8e-03, 8.8e-06, 3.8e-11

Table 2: Results for various values of ζ and κ

Problem	1	2	3	4
$\zeta = 10^{-9}$	2	2	12	13
$\zeta = 10^{-8}$	2	2	12	13
$\zeta = 10^{-7}$	2	2	12	13
$\zeta = 10^{-6}$	2	2	12	13
$\zeta = 10^{-5}$	3	3	12	13
$\zeta = 10^{-4}$	3	3	13	14
$\zeta = 10^{-3}$	4	3	13	14
$\zeta = 10^{-2}$	4	4	13	14
$\zeta = 10^{-1}$	7	6	13	14
$\zeta = 10^{0}$	14	11	13	14
$\zeta = 10^{1}$	58	36	14	14
$\zeta = 10^{2}$	272	175	16	15
$\zeta = 10^{3}$	780	612	20	18
$\zeta = \infty$	1769	1809	71	41

Problem	1	2	3	4
$\kappa = 10^{-9}$	3	3	13	14
$\kappa = 10^{-8}$	3	3	13	14
$\kappa = 10^{-7}$	3	3	13	14
$\kappa = 10^{-6}$	3	3	13	14
$\kappa = 10^{-5}$	3	3	13	14
$\kappa = 10^{-4}$	4	3	13	14
$\kappa = 10^{-3}$	4	3	13	14
$\kappa = 10^{-2}$	4	4	13	14
$\kappa = 10^{-1}$	6	5	13	14
$\kappa = 10^{0}$	12	11	18	18
$\kappa = 10^{1}$	17	19	23	23
$\kappa = 10^{2}$	23	24	27	28
$\kappa = 10^{3}$	29	30	31	31
$\kappa = \infty$	37	39	32	32

Table 3: Results for Problems 1 and 2 with $n = 100000$

	# iter.	# iter. (CGM)	time(sec.)	$\left\|F\left(x^k\right)\right\|$
Problem 1	4	7125	5.8e+03	5.6e-02, 6.3e-05, 3.2e-09
Problem 2	4	5334	3.5e+03	2.9e-02, 1.5e-05, 9.5e-11

For each problem, we set an initial point as $x^{0,1} = \left\{\frac{n}{2}, \ldots, \frac{n}{2}\right\}^T, x^{0,2} = \{n, \ldots, n\}^T, x^{0,3} = \left\{-\frac{n}{2}, \ldots, -\frac{n}{2}\right\}^T x^{0,4} = \{-n, \ldots, -n\}^T$. Moreover, we update ν_k in the approximate condition (3.4) as described in Remark 3.1. We set the default values of parameters as $\alpha = 0.6, \beta = 0.7, \gamma = 0.8, \delta = 1.0, \eta = 0.8, \rho = 0.5, \tau = 2.0, p = 2.0, \zeta = 0.001$ and $\kappa = 0.001$, which may be altered if necessary. We use the stopping criterion $\left\|F\left(x^k\right)\right\| < 10^{-8}\sqrt{n}$ for each experiment. The algorithm was coded in C and run on a Sun Ultra 60 workstation.

Table 1 shows the computational results for each problem, with the following items: The dimension of the problem (n), the initial point (i.p.), the number of iterations of Algorithm 3.1 (# iter.), the cumulative number of iterations of the CGM (# iter. (CGM)), the CPU time in second (time(sec.)), and the values of $\left\|F\left(x^k\right)\right\|$ at last three iterations of Algorithm 3.1 ($\left\|F\left(x^k\right)\right\|$). Note that, from Assumption 2.1 (iii), we have $\left\|F\left(x^k\right)\right\| = O\left(\text{dist}\left(x^k, X^*\right)\right)$. For each problem, the algorithm always stopped successfully, and the generated sequence converged to the solution set superlinearly.

Table 2 consists of two tables; one shows the number of iterations for various values of ζ in the new updating rule (4.2) for μ_k, while the other shows the number of iterations for various values of κ in the new approximate condition (4.3). In these experiments, we set $n = 1000$ and $x^0 = x^{0,1}$. In Table 2, $\zeta = \infty$ and $\kappa = \infty$ mean that we use the updating rule (2.4) and the approximate condition (3.4), respectively. From Table 2, we see that when ζ and κ are small, the algorithm solved all problems successfully and the number of iterations is small. On the other hand, when ζ and κ are large or these parameters are not used ($\zeta = \infty/\kappa = \infty$), the number of iterations becomes large. This observation indicates that the performance of the algorithm can be improved by choosing appropriate values for ζ and κ in (4.2) and (4.3), respectively.

Finally, we show Table 3, which contains results for Problems 1 and 2 with $n = 100000$ and $x^0 = x^{0,1}$. Thus we may conclude that the algorithm can deal with large-scale problems efficiently.

5 Concluding Remarks

In this paper, we have discussed the convergence properties of the ILMM under a local error bound condition on F. Using an approach similar to [20], we have showed that the ILMM converges to the solution set superlinearly under appropriate conditions on the approximate solution of the system of linear equations solved at each iteration. For large scale problems, this property is very useful. On the other hand, it was shown in [20] that the LMM converges quadratically under the assumption that $\mu_k = \left\|F\left(x^k\right)\right\|^2$. In that case, if μ_k is very small and $F'\left(x^*\right)^T F\left(x^*\right)$ is singular, the system of linear equations tends to be unstable numerically. Since the ILMM converges superlinearly even if $\mu_k = \left\|F\left(x^k\right)\right\|^\delta, 0 < \delta \leq 2$, we can expect the numerical robustness of the ILMM when it is implemented with $0 < \delta < 2$.

References

[1] D. P. Bertsekas, *Nonlinear Programming* (Athena Scientific, Massachusetts, 1995).

[2] F. H. Clarke, *Optimization and Nonsmooth Analysis* (Wiley, New York, 1983).

[3] T. De Luca, F. Facchinei and C. Kanzow, A semismooth equation approach to the solution of nonlinear complementarity problems, *Mathematical Programming* 75 (1996) 407-439.

[4] J. E. Dennis, Jr. and J. J. Moré, A characterization of superlinear convergence and its application to Quasi-Newton methods, *Mathematics of Computation* 28 (1974) 549-560.

[5] F. Facchinei and C. Kanzow, A nonsmooth inexact Newton Method for the solution of large-scale nonlinear complementarity problems, *Mathematical Programming* 76 (1997) 493-512.

[6] M. C. Ferris and J.-S. Pang, Engineering and economic applications of complementarity problem, *SIAM Review* 39 (1997) 669-713.

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

[8] R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, New York, 1987).

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, second ed., (The Johns Hopkins University Press, 1989).

[10] 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) 161-220.

[11] M. R. Hestenes, *Conjugate Direction Methods in Optimization* (Springer-Verlag, New York, 1980).

[12] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic, New York, 1970).

[13] J.-S. Pang, Error bounds in mathematical programming, *Mathematics Programming* 79 (1997) 299-332.

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

[15] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, *Mathematics of Operations Research* 18 (1993) 227-244.

[16] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* 14 (1976), 877-898.

[17] N. Yamashita and M. Fukushima, The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem, to appear in *SIAM Journal on Optimization*.

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

[19] N. Yamashita, H. Dan and M. Fukushima, On the identification of degenerate set of the nonlinear complementarity problem with the proximal point algorithm, to appear.

[20] N. Yamashita and M. Fukushima, On the rate of convergence of the Levenberg-Marquardt method, Technical Report 2000-008, Department of Applied Mathematics and Physics, Kyoto University (November 2000).

A Appendix: Quadratic or superlinear algorithm for monotone nonlinear complementarity problem without degeneracy conditions

In this appendix, we introduce one of the application using the method which we propose in this paper.

A superlinearly convergent algorithm for the monotone nonlinear complementarity problem without nondegeneracy conditions

Abstract

In this paper, we consider an algorithm for solving the monotone nonlinear complementarity problem (NCP). Recently, Yamashita and Fukushima proposed a method based on the proximal point algorithm (PPA) for monotone NCP. The method enjoys the favorable property that a generated sequence converges to the solution set of NCP superlinearly. However, when a generated sequence converges to a degenerate solution, the method may need much computational time to solve subproblems, and hence the method does not have genuine superlinear convergence. More recently, Yamashita, Dan and Fukushima presented a technique to identify whether a solution is degenerate or not. Using this technique, we construct a differentiable system of nonlinear equations whose solution is a solution of the original NCP. Moreover, we propose a hybrid algorithm which is based on the PPA and uses this system. We show that the proposed algorithm has a quadratic or superlinear rate of convergence even if it converges to a degenerate solution.

A.1 Introduction

The nonlinear complementarity problem (NCP) is to find a vector $\bar{x} \in \Re^n$ such that

$$\mathrm{NCP}(F): \quad \bar{x}_i \geq 0, \;\; F_i(\bar{x}) \geq 0, \;\; \bar{x}_i F_i(\bar{x}) = 0, \quad i = 1, \ldots, n,$$

where F is a mapping from $\Re^n$ to $\Re^n$. When F is affine, NCP(F) is called the linear complementarity problem (LCP). NCP can be found in various fields, e.g., operations research, engineering, finance and so on [7]. Throughout this paper, F is assumed to be continuously differentiable and monotone.

For a solution $\bar{x}$ of NCP(F), let $P(\bar{x}), N(\bar{x})$ and $C(\bar{x})$ be defined by

$$\begin{aligned} P(\bar{x}) &:= \{i \mid \bar{x}_i > 0, F_i(\bar{x}) = 0\}, \\ N(\bar{x}) &:= \{i \mid \bar{x}_i = 0, F_i(\bar{x}) > 0\}, \\ C(\bar{x}) &:= \{i \mid \bar{x}_i = 0, F_i(\bar{x}) = 0\}, \end{aligned}$$

respectively. Note that each index of a solution $\bar{x}$ of NCP(F) belongs to one of these sets. If $C(\bar{x}) \neq \emptyset$, we call $\bar{x}$ a degenerate solution, otherwise we call $\bar{x}$ a nondegenerate solution. In this paper, we will focus on these index sets for a particular solution, and develop an algorithm for solving NCP(F) by estimating the correct index sets.

Various methods for solving NCP, such as the generalized Newton method (GNM) [4], the smoothing method [2] and the regularization method [6], have been proposed and shown to have nice convergence properties. However, those methods generally require the local uniqueness of a solution for a superlinear rate of convergence. Recently, Yamashita and Fukushima [11] proposed a method, henceforth called the PPA, based on the proximal point algorithm, and showed that it has a superlinear rate of convergence without the local uniqueness of a solution. However, when a sequence $\{x^k\}$ generated by the PPA converges to a degenerate solution, subproblems may become computationally expensive. This difficulty comes from the fact that we do not know in advance whether or not $\{x^k\}$ converges to a degenerate solution. More recently, Yamashita, Dan and Fukushima [12] presented a technique that enables us to identify $P^* = P(x^*), N^* = N(x^*)$ and $C^* = C(x^*)$ when x^k enters a certain vicinity of x^*. Once we identify P^*, N^* and C^*, we can find a solution of NCP(F) by solving the system of nonlinear equations

$$G_{x^*}(x) := \begin{pmatrix} F_{P^*}(x) \\ x_{N^*} \\ F_{C^*}(x) \\ x_{C^*} \end{pmatrix} = 0. \tag{A.1.1}$$

In fact, if a solution $\hat{x}$ of (A.1.1) is sufficiently near to x^*, then we have $\hat{x}_P > 0$ and $F_N(\hat{x}) > 0$, and hence $\hat{x}$ is also a solution of NCP(F). Moreover, since the mapping $G_{x^*} : \Re^n \to \Re^{n+|C^*|}$ is differentiable, we can use any Newton like method which requires differentiability of the mapping, and we can expect that such a method has a quadratic or superlinear rate of convergence no matter whether x^* is degenerate or nondegenerate.

In this paper, we construct a differentiable system of nonlinear equations (A.1.1) by using the technique proposed in [12]. Moreover, we propose a hybrid algorithm which generates a sequence

$\{x^k\}$ by the PPA primarily and also tries to find a solution of the system of nonlinear equations (A.1.1) by using the inexact Levenberg-Marquardt method (ILMM) proposed in [3], when an iterative point x^k is judged to lie sufficiently close to a solution. We show that the proposed method has a quadratic or superlinear convergence property if $\|G_{x^*}(x)\|$ provides a local error bound for (A.1.1), i.e., there exists constants $b_G > 0$ and $c_G > 0$ such that

$$c_G \text{dist}\,(x, X_G^*) \leq \|G_{x^*}(x)\| \quad \forall x \in B\,(x^*, b_G)\,, \tag{A.1.2}$$

where X_G^* is the solution set of (A.1.1) and $B\,(x^*, b_G) := \{x \mid \|x - x^*\| \leq b_G\}$. More specifically, we show that either a sequence $\{x^k\}$ generated by the proposed algorithm converges to the solution set of NCP quadratically or $\|G_{x^*}(x^k)\|$ converges to 0 superlinearly, without assuming the nondegeneracy of a solution.

This paper is organized as follows: In Section A.2, we review some mathematical concepts and results such as error bounds for NCP, the PPA [11] and the ILMM [3]. In Section A.3, we construct a differentiable system of nonlinear equations whose solution is a solution of NCP(F). Moreover, we propose the hybrid algorithm based on (A.1.1) and show its global convergence. In Section A.4, we make some concluding remarks and discuss future research topics.

A.2 Preliminaries

In this section, we review some concepts and results which are used in the subsequent discussion.

A.2.1 Error bound for NCP

First, we recall some mathematical concepts related to a mapping F.

Definition A.1 *The mapping $F : \Re^n \to \Re^n$ is called*

(i) monotone if

$$(x - y)^T (F(x) - F(y)) \geq 0 \qquad \forall x, y \in \Re^n,$$

(ii) strongly monotone with modulus $\mu > 0$ if

$$(x - y)^T (F(x) - F(y)) \geq \mu\|x - y\|^2 \qquad \forall x, y \in \Re^n.$$

Next, we consider a system of nonlinear equations which is equivalent to NCP(F). In order to reformulate NCP, we may use a function $\phi : \Re^2 \to \Re$ that has the following property:

$$\phi(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0. \tag{A.2.3}$$

Such a function ϕ is called an NCP-function. Using an NCP-function ϕ, let $H_F : \Re^n \to \Re^n$ be defined by

$$H_F(x) = \begin{pmatrix} \phi\,(x_1, F_1(x)) \\ \vdots \\ \phi\,(x_n, F_n(x)) \end{pmatrix}.$$

From (A.2.3), it is easily seen that NCP(F) is equivalent to the system of nonlinear equations

$$H_F(x) = 0.$$

The following two NCP-functions are well-known:

$$\begin{aligned}\phi_{NR}(a,b) &= \min\{a,b\},\\ \phi_{FB}(a,b) &= \sqrt{a^2+b^2} - a - b.\end{aligned}$$

The functions ϕ_{NR} and ϕ_{FB} are called the natural residual function and the Fischer-Burmeister function [8], respectively. The function ϕ_{FB} is equivalent to ϕ_{NR} in the sense that they satisfy the following inequalities [8]:

$$(2-\sqrt{2})\,|\phi_{NR}(a,b)| \;\le\; |\phi_{FB}(a,b)| \;\le\; (2+\sqrt{2})\,|\phi_{NR}(a,b)| \quad \forall (a,b)^T \in \Re^2.$$

Throughout this paper, we assume that the mapping H_F is given by any NCP-function ϕ satisfying

$$\nu_1\,|\phi_{NR}(a,b)| \;\le\; |\phi(a,b)| \;\le\; \nu_2\,|\phi_{NR}(a,b)| \quad \forall (a,b)^T \in \Re^2, \tag{A.2.4}$$

where ν_1 and ν_2 are positive constants.

The next theorem states some error bound properties of $\|H_F(x)\|$ for NCP.

Theorem A.1 *[8]*

(i) Suppose that F is strongly monotone with modulus $\mu > 0$ and Lipschitz continuous with constant $L > 0$. Then $\|H_F(x)\|$ provides a global error bound for NCP(F), that is,

$$\|x - \hat{x}\| \le \frac{K_1(L+1)}{\mu}\,\|H_F(x)\| \qquad \forall x \in \Re^n,$$

where $\hat{x}$ is the unique solution of NCP(F) and $K_1 > 0$ is a constant independent of F.

(ii) Suppose that F is affine and there exists a solution of NCP(F). Then $\|H_F(x)\|$ provides a local error bound for NCP(F), that is, there exist positive constants K_2 and K_3 such that

$$\|H_F(x)\| \le K_2 \;\Rightarrow\; \operatorname{dist}(x, X^*) \le K_3\,\|H_F(x)\|\,,$$

where X^* denotes the solution set of NCP(F).

A.2.2 The PPA and identification of the index sets

The PPA presented in [11] is stated as follows:

Algorithm PPA

Step 0: Choose parameters $\beta \in (0,1), c_0 \in (0,1)$ and an initial point $x^0 \in \Re^n$. Set $k := 0$.

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

Step 2: Let $F^k : \Re^n \to \Re^n$ be given by

$$F^k(x) = F(x) + c_k \left(x - x^k\right).$$

Find an approximate solution x^{k+1} of NCP(F^k) such that

$$\left\|H_{F^k}(x^{k+1})\right\| \leq \beta^k \min\left\{1, \left\|x^{k+1} - x^k\right\|\right\}. \tag{A.2.5}$$

Step 3: Choose $c_{k+1} > 0$ and set $k := k + 1$. Go to Step 2. □

This algorithm enjoys nice convergence properties. Specifically, the sequence generated by the PPA converges to a solution of NCP(F) globally, and the distance between the generated sequence and the solution set of NCP(F) converges to 0 superlinearly under mild assumptions. In what follows, $\{x^k\}$ denotes a sequence generated by Algorithm PPA, and x^* denotes the limit point of $\{x^k\}$. The following convergence theorem has been established in [11].

Theorem A.2 *Suppose that F is monotone and Lipschitzian. Suppose also that NCP(F) has a solution. Then, $\{x^k\}$ converges to a solution x^* of NCP(F) whenever $\{c_k\}$ is bounded. Moreover, if $\|H_F(x)\|$ provides a local error bound in a neighborhood of x^* and $c_k \to 0$, then $\{\mathrm{dist}(x^k, X^*)\}$ converges to 0 superlinearly, where X^* is the solution set of NCP(F).*

The most expensive task in Algorithm PPA is to solve NCP(F^k) in Step 2. In [11], it is proposed that NCP(F^k) be solved using Generalized Newton Method (GNM) [4]. Note that F^k is strongly monotone when F is monotone. Then, the GNM can find an approximate solution of subproblem NCP(F^k) rapidly when x^* is nondegenerate, as stated in the following theorem [11].

Theorem A.3 *Suppose that x^* is a nondegenerate solution of NCP(F) and $\|H_F(x)\|$ provides a local error bound in a neighborhood of x^*. Then, a single iteration of the GNM for NCP(F^k) yields a point x^{k+1} that satisfies (A.2.5), provided k is sufficiently large.*

When x^* is a degenerate solution, Theorem A.3 no longer guarantees that the GNM can find a solution of subproblem NCP(F^k) in a few iterations, and hence it may take much time to solve subproblems. To overcome this difficulty, we will use the technique proposed in [12] to identify the index sets $P(x^*), N(x^*)$ and $C(x^*)$, which is based on the following idea: Suppose that $\|H_F(x)\|$ provides a local error bound for NCP(F) and we let $c_k = \alpha^k$ for all k in Algorithm PPA, where α is a constant such that $\beta < \alpha < 1$. Let a function sequence $\{\rho^k\}$ be defined by

$$\rho^k(x) = \sqrt{\frac{\|H_F(x)\|}{\alpha^k}},$$

where the index sets P^k, N^k and C^k are defined by

$$\begin{aligned} P^k &:= \left\{i \mid x_i^k > \rho^k(x^k), F_i(x^k) \leq \rho^k(x^k)\right\}, \\ N^k &:= \left\{i \mid x_i^k \leq \rho^k(x^k), F_i(x^k) > \rho^k(x^k)\right\}, \\ C^k &:= \left\{i \mid x_i^k \leq \rho^k(x^k), F_i(x^k) \leq \rho^k(x^k)\right\}, \end{aligned} \tag{A.2.6}$$

respectively. Then, we have

$$P^k = P(x^*), \quad N^k = N(x^*), \quad C^k = C(x^*).$$

for sufficiently large k [12, Theorem 3.4].

A.2.3 Inexact Levenberg-Marquardt method

The Levenberg-Marquardt method (LMM) [1, 9] is a method for solving the system of nonlinear equations

$$G(y) = 0, \tag{A.2.7}$$

where $G : \Re^n \to \Re^m$ is a continuously differentiable mapping. The LMM generates a sequence $\{y^l\}$ by $y^{l+1} := y^l + \hat{d}^l$, where $\hat{d}^l$ is a solution of the system of linear equations

$$\left(\nabla G(y^l)\nabla G(y^l)^T + \mu_l I\right) d = -\nabla G(y^l)G(y^l). \tag{A.2.8}$$

Here $\nabla G(y) \in \Re^{n\times m}$ is the transposed Jacobian of G, μ_l is a positive parameter and I is the identity matrix. Since $\nabla G(y^l)\nabla G(y^l)^T + \mu_l I$ is positive definite, (A.2.8) has a unique solution. However, it is expensive to find an exact solution of (A.2.8) when n is large. In that case, the inexact Levenberg-Marquardt method (ILMM) [3, 5] is useful. The ILMM uses an approximate solution d^l of (A.2.8) as a search direction, and generates a sequence $\{y^l\}$ by $y^{l+1} := y^l + d^l$.

Recently, Dan, Yamashita and Fukushima [3] showed the following theorem which says that the ILMM has a quadratic rate of convergence under a local error bound condition, which is milder than the nonsingularity condition at a solution.

Theorem A.4 *Let $\{y^l\}$ be a sequence generated by the ILMM and y^* be a solution of (A.2.7). Suppose that the following two conditions hold.*

(i) There exist constants $b_1 \in (0, 1)$ and $\kappa_1 \in (0,\infty)$ such that

$$\|G(y) - G(x) - \nabla G(x)(y - x)\| \le \kappa_1 \|y - x\|^2 \quad \forall x, y \in B(y^*, b_1),$$

where $B(y^*, b_1) := \{y \mid \|y - y^*\| \le b_1\}$.

(ii) $\|G(y)\|$ provides an error bound for (A.2.7) on $B(y^*, b_1)$, i.e., there exists a constant $\kappa_2 > 0$ such that

$$\kappa_2 \operatorname{dist}(y, Y^*) \le \|G(y)\| \quad \forall y \in B(y^*, b_1),$$

where Y^* is the solution set of (A.2.7).

Suppose that the parameters μ_l satisfy

$$\mu_l = \left\|G(y^l)\right\|^2$$

and

$$\frac{\|r^l\|}{\mu_l} = O\left(\operatorname{dist}\left(y^l, Y^*\right)^2\right),$$

where r^l are residual vectors defined by

$$r^l := \left(\nabla G(y^l)\nabla G(y^l)^T + \mu_l I\right) d^l + \nabla G(y^l)G(y^l). \tag{A.2.9}$$

Then there exists a positive constant c such that $y^l \in B\left(y^*, c\left\|y^0 - y^*\right\|\right)$ for all l, provided an initial point y^0 is sufficiently close to y^*. Moreover, $\{\operatorname{dist}(y^l, Y^*)\}$ converges to 0 quadratically.

The assumption (i) of Theorem A.4 holds when ∇G is locally Lipschitzian [10, Theorem 3.2.12].

A.3 Proposed algorithm and its convergence properties

In this section, we describe the proposed algorithm and show that it has at least a superlinear rate of convergence for NCP, without assuming the local uniqueness of a solution.

A.3.1 Differentiable system of nonlinear equations

First, we make the following assumption to guarantee that the PPA has a superlinear convergence and the technique to identify the index sets works well.

Assumption A.1

(i) ∇F *is locally Lipschitzian.*

(ii) $\|H_F(x)\|$ *provides a local error bound for NCP(*F*), i.e., there exist constants* $b_H > 0$ *and* $c_H > 0$ *such that*

$$b_H \text{dist}\,(x, X^*) \leq \|H_F(x)\| \quad \forall x \in \{x \mid \|H_F(x)\| \leq c_H\},$$

where X^* *is the solution set of NCP(*F*).*

Sufficient conditions under which $\|H_F(x)\|$ provides a local error bound for NCP(F) are given in Theorem A.1.

We consider the mapping $G_{x^*} : \Re^n \to \Re^{n+|C^*|}$ defined by (A.1.1). Since $x_{P^*} > 0$ and $F_{N^*}(x) > 0$ in a sufficiently small neighborhood of x^*, a solution $\hat{x}$ of (A.1.1) also solves NCP(F) if $\hat{x}$ is sufficiently close to x^*. We note that $G_{x^*}(x)$ is differentiable and the Jacobian $\nabla G_{x^*}(x)$ is locally Lipschitzian, and hence the assumption (i) of Theorem A.4 holds for G_{x^*}. So, the ILMM applied to the equation (A.1.1) has a quadratic rate of convergence if $\|G_{x^*}(x)\|$ provides a local error bound. Accordingly, we make the following assumption.

Assumption A.2 $\|G_{x^*}(x)\|$ *provides a local error bound for (A.1.1) in a neighborhood of* x^**, i.e., there exist positive constants* b_G *and* c_G *which satisfy (A.1.2).*

Assumption A.2 holds when F_{P^*} and F_{C^*} are affine. Though this assumption does not necessarily hold when F is nonlinear, it does not seem very restrictive.

Assumptions A.1 and A.2 are closely related. In fact, as shown in the next lemma, Assumption A.2 is implied by Assumption A.1 under some normal circumstances.

Lemma A.1 *Suppose that Assumption A.1 holds. If there exists a constant* $r_1 > 0$ *such that* $X^* \cap B\,(x^*, r_1) = X_G^* \cap B\,(x^*, r_1)$, *where* X_G^* *is the solution set of (A.1.1), then* $\|G_{x^*}(x)\|$ *provides a local error bound for (A.1.1) in a neighborhood of* x^**, i.e., there exist constants* b_G *and* c_G *which satisfy (A.1.2). In particular, if* x^* *is a locally unique solution of NCP(*F*), then* $\|G_{x^*}(x)\|$ *provides a local error bound for (A.1.1) in a neighborhood of* x^**.*

Proof: Choosing $r_2 > 0$ sufficiently small yields that, for any $x \in B\,(x^*, r_2)$,

$$x_i \geq F_i(x) \quad \forall i \in P^*,$$
$$x_i \leq F_i(x) \quad \forall i \in N^*.$$

It then follows that, for any $x \in B(x^*, r_2)$,

$$\begin{aligned} \|H_F(x)\| &= \sqrt{\sum_{i\in P^*\cup N^*\cup C^*} \phi^2(x_i, F_i(x))} \\ &\le \nu_2 \sqrt{\sum_{i\in P^*\cup N^*\cup C^*} |\min\{x_i, F_i(x)\}|^2} \\ &\le \nu_2 \sqrt{\sum_{i\in P^*} F_i^2(x) + \sum_{i\in N^*} x_i^2 + \sum_{i\in C^*} (x_i^2 + F_i^2(x))} \\ &= \nu_2 \|G_{x^*}(x)\|, \end{aligned} \tag{A.3.10}$$

where ν_2 is the positive constant given in (A.2.4). Moreover, by choosing $r_3 > 0$ sufficiently small, we have

$$\|G_{x^*}(x)\| \le c_H/\nu_2 \quad \forall x \in B(x^*, r_3). \tag{A.3.11}$$

Let $r := \min\{r_1, r_2, r_3\}$. It then follows from (A.3.10) and (A.3.11) that

$$\|H_F(x)\| \le c_H \quad \forall x \in B(x^*, r).$$

Therefore, from Assumption A.1 (ii) and (A.3.10), we get

$$b_H \mathrm{dist}(x, X^*) \le \|H_F(x)\| \le \nu_2 \|G_{x^*}(x)\| \quad \forall x \in B(x^*, r). \tag{A.3.12}$$

Let $x \in B\left(x^*, \frac{r}{2}\right)$, and let $\bar{x}_G$ and $\bar{x}$ be one of the nearest points from x in X_G^* and X^*, respectively. Since $x^* \in X_G^*$ and $x^* \in X^*$, we have, for any $x \in B\left(x^*, \frac{r}{2}\right)$,

$$\begin{aligned} \|\bar{x}_G - x\| &\le \|x^* - x\| \le \frac{r}{2}, \\ \|\bar{x} - x\| &\le \|x^* - x\| \le \frac{r}{2}, \end{aligned}$$

and hence

$$\begin{aligned} \|\bar{x}_G - x^*\| &\le \|\bar{x}_G - x\| + \|x^* - x\| \le r, \\ \|\bar{x} - x^*\| &\le \|\bar{x} - x\| + \|x^* - x\| \le r. \end{aligned}$$

Therefore we have $\bar{x}_G \in X_G^* \cap B(x^*, r)$ and $\bar{x} \in X^* \cap B(x^*, r)$. It then follows from the assumption $X^* \cap B(x^*, r) = X_G^* \cap B(x^*, r)$ that $\mathrm{dist}(x, X^*) = \mathrm{dist}(x, X_G^*)$ for any $x \in B\left(x^*, \frac{r}{2}\right)$. Consequently, by (A.3.12), we have

$$\nu_2^{-1} b_H \, \mathrm{dist}(x, X_G^*) \le \|G_{x^*}(x)\| \quad \forall x \in B\left(x^*, \frac{r}{2}\right),$$

i.e., $\|G_{x^*}(x)\|$ provides a local error bound for (A.1.1) in a neighborhood of x^*. □

A.3.2 The algorithm

As stated in Section A.2.2, $P^k = P(x^*), N^k = N(x^*)$ and $C^k = C(x^*)$ hold when k is sufficiently large, and hence, the system of nonlinear equations

$$G^k(x) := \begin{pmatrix} F_{P^k}(x) \\ x_{N^k} \\ F_{C^k}(x) \\ x_{C^k} \end{pmatrix} = 0 \tag{A.3.13}$$

coincides with the system of nonlinear equations (A.1.1). In this case, a solution $\hat{x}$ of (A.3.13) is a solution of NCP(F) if $\hat{x}$ is sufficiently close to x^*. Then we naturally come up with the following method: We generate a sequence $\{x^k\}$ by the PPA primarily, and if an iterative point is judged to be close to the solution x^*, then we solve (A.3.13) by the ILMM. Based on this idea, we propose the following hybrid algorithm.

Algorithm Hybrid

Step 0: Choose parameters $M_1 > 0, M_2 > 0, 0 < \beta < \alpha < 1, 0 < \gamma < 1$, and an initial point x^0. Let the initial index sets be $P^0 = N^0 = C^0 = \emptyset$. Set $k := 1$.

Step 1: Let $c_k = \alpha^k$ and obtain x^k by applying a single iteration of Algorithm PPA. Determine the index sets P^k, N^k, C^k by (A.2.6).

Step 2: If $\|H_F(x^k)\| \leq M_1, P^k = P^{k-1}, N^k = N^{k-1}, C^k = C^{k-1}$ and $P^k \cup N^k \cup C^k = \{1, 2, \ldots, n\}$, then go to Step 3. Otherwise, set $k := k + 1$ and go to Step 1.

Step 3: (ILMM for (A.3.13))

Step 3.0: Set $y^{k,0} := x^k$, $\mu_{k,0} = \|G^k(y^{k,0})\|$, and $l := 0$.

Step 3.1: Find an approximate solution $d^{k,l}$ of the system of linear equations

$$\left(\nabla G^k(y^{k,l})\nabla G^k(y^{k,l})^T + \mu_{k,l} I\right) d = -\nabla G^k(y^{k,l})^T G^k(y^{k,l}). \tag{A.3.14}$$

Step 3.2: If $y^{k,l}$ satisfies a stopping criterion, then exit. Otherwise, if $y^{k,l}$ does not satisfy

$$\left(y^{k,l} + d^{k,l}\right)_{P^k} > 0, \tag{A.3.15}$$

$$F_{N^k}(y^{k,l} + d^{k,l}) > 0, \tag{A.3.16}$$

$$\|G^k(y^{k,l} + d^{k,l})\| \leq \gamma^l \|G^k(y^{k,l})\| \tag{A.3.17}$$

$$\|x^k - \left(y^{k,l} + d^{k,l}\right)\| \leq M_2, \tag{A.3.18}$$

then set $k := k + 1$ and go to Step 1.

Step 3.3: Set $y^{k,l+1} := y^{k,l} + d^{k,l}, \mu_{k,l+1} = \|G^k(y^{k,l+1})\|$ and $l := l + 1$. Go to Step 3.1. □

In Algorithm Hybrid, we try to identify the three index sets $P(x^*), N(x^*)$ and $C(x^*)$ in Step 2. The conditions in Step 2 will be satisfied when k is sufficiently large. Note, however, that those conditions do not guarantee that the index sets P^*, N^* and C^* are identified correctly. Therefore, we check conditions (A.3.15) – (A.3.18) in Step 3.2 to ensure that the sequence $\{y^{k,l}\}$ is converging to a solution of NCP(F). The conditions (A.3.15) and (A.3.16) check whether or not P^k, N^k and C^k are identified correctly. The condition (A.3.17) guarantees that $\|G^k(y^{k,l})\|$ is converging to 0 superlinearly, and the condition (A.3.18) guarantees that $\{y^{k,l}\}_{l=0,1,2,\ldots}$ is not diverging.

A.3.3 Convergence theorem

Let the residual vector $r^{k,l}$ associated with an approximate solution $d^{k,l}$ of the system of linear equations (A.3.14) be defined by

$$r^{k,l} := \left(\nabla G^k(y^{k,l})\nabla G^k(y^{k,l})^T + \mu_{k,l}I\right) d^{k,l} + \nabla G^k(y^{k,l})G^k(y^{k,l}).$$

Now we show the following convergence theorem for Algorithm Hybrid.

Theorem A.5 *Suppose that Assumptions A.1 and A.2 hold. Suppose also that* $\{y^{k,l}\}_{l=0,1,2,\ldots}$ *satisfies*

$$\frac{\|r^{k,l}\|}{\mu_{k,l}} = O\left(\operatorname{dist}\left(y^{k,l}, X^*_{G^k}\right)^2\right), \tag{A.3.19}$$

where $X^*_{G^k}$ *is the solution set of (A.3.13). Then, there exists a positive integer* k *for which either of the following statements is true:*

(a): *Any accumulation point of* $\{y^{k,l}\}_{l=0,1,2,\ldots}$ *is a solution of NCP(F), and* $\{\|G^k(y^{k,l})\|\}_{l=0,1,2,\ldots}$ *converges to* 0 *superlinearly.*

(b): $\{y^{k,l}\}_{l=0,1,2,\ldots}$ *converges to a solution of NCP(F), and* $\{\operatorname{dist}(y^{k,l}, X^*_G)\}_{l=0,1,2,\ldots}$ *converges to* 0 *quadratically.*

Proof: First, we consider the case where the inner loop in Step 3 cycles infinitely for some k, that is, an infinite sequence $\{y^{k,l}\}_{l=0,1,2,\ldots}$ satisfies (A.3.15) – (A.3.18), even if the index sets $P(x^*), N(x^*)$ and $C(x^*)$ have yet to be identified correctly. In this case, it follows from (A.3.18) that $\{y^{k,l}\}_{l=0,1,2,\ldots}$ has accumulation points, and from (A.3.17), $\{\|G^k(y^{k,l})\|\}$ converges to 0 superlinearly. Moreover, from (A.3.15) and (A.3.16), any accumulation point of $\{y^{k,l}\}_{l=0,1,2,\ldots}$ is a solution of NCP(F). Therefore the statement (a) holds.

Next, we show that, even if (a) does not hold, the statement (b) holds eventually. In fact, for sufficiently large k, $\|H_F(x^k)\|$ becomes sufficiently small by Theorem A.2, and $P^k = P(x^*), N^k = N(x^*), C^k = C(x^*)$ hold as shown in [12]. Hence, the system (A.3.13) coincides with the system (A.1.1) for sufficiently large k. Note that, by Assumptions A.1 and A.2, $G_{x^*}(x)$ and x^* satisfy the assumptions (i) and (ii) in Theorem A.4, where G and y^* are regarded as G_{x^*} and x^*, respectively. In what follows, we consider a sequence $\{z^{k,l}\}_{l=0,1,2,\ldots}$ generated by the ILMM for the equation $G_{x^*}(x) = 0$ with an initial point $z^{k,0} := x^k$ for sufficiently large k, and show that $z^{k,l}$ satisfies the conditions (A.3.15)–(A.3.18) for all l. It follows from Lemma 2.3 of [3] and (A.3.19) that, if k is sufficiently large and $z^{k,0} = x^k$ is sufficiently close to x^*, then we have $\|z^{k,l+1} - \bar{z}^{k,l+1}\| \leq$

$c\,\|z^{k,l}-\bar{z}^{k,l}\|^2$ for all l, where c is a positive constant and $\bar{z}^{k,l}$ is one of the nearest points of X_G^* from $z^{k,l}$. Then, from Assumption A.2, we have

$$\frac{\left\|G_{x^*}(z^{k,l+1})\right\|}{\left\|G_{x^*}(z^{k,l})\right\|^2} = \frac{\left\|G_{x^*}(z^{k,l+1})-G_{x^*}(\bar{z}^{k,l+1})\right\|}{\left\|G_{x^*}(z^{k,l})\right\|^2} \le \frac{L_{x_*}\left\|z^{k,l+1}-\bar{z}^{k,l+1}\right\|}{c_G\left\|z^{k,l}-\bar{z}^{k,l}\right\|^2} \le \frac{cL_{x^*}}{c_G} \quad l=0,1,2,\ldots,$$

where L_{x^*} is a Lipschitz constant of $G_{x^*}(x)$. When k is sufficiently large, we have $\|G_{x^*}(z^{k,0})\| \le \frac{\gamma c_G}{cL_{x^*}}$, and it follows that for each $l=0,1,2,\ldots$

$$\begin{aligned}
\left\|G_{x^*}(z^{k,l+1})\right\| &\le \frac{cL_{x^*}}{c_G}\left\|G_{x^*}(z^{k,l})\right\|^2\\
&\le \left(\frac{cL_{x^*}}{c_G}\left\|G_{x^*}(z^{k,l-1})\right\|\right)^2\left\|G_{x^*}(z^{k,l})\right\|\\
&\vdots\\
&\le \left(\frac{cL_{x^*}}{c_G}\left\|G_{x^*}(z^{k,0})\right\|\right)^{2^l}\left\|G_{x^*}(z^{k,l})\right\|\\
&\le \gamma^{2^l}\left\|G_{x^*}(z^{k,l})\right\|\\
&\le \gamma^{l}\left\|G_{x^*}(z^{k,l})\right\|.
\end{aligned}$$

Therefore, (A.3.17) holds when k is sufficiently large. Let $\hat{r}$ be the distance between $y^{k,0}$ and x^*. When $\hat{r}$ is sufficiently small, Theorem A.4 says that a generated sequence $\{z^{k,l}\}$ satisfies $z^{k,l}\in B(x^*,c_3\hat{r})$ for all l, where $c_3>0$ is a constant. Moreover, choosing sufficiently large k if necessary, $\hat{r}$ can be made arbitrarily small. Then, it follows from $x_P^*>0$ and $F_N(x^*)>0$ that (A.3.15), (A.3.16) and (A.3.18) hold for sufficiently large k. Consequently, (A.3.15) – (A.3.18) are satisfied for sufficiently large k. Then, from Theorem A.4, $\{\mathrm{dist}(z^{k,l},X_G^*)\}_{l=0,1,2,\ldots}$ converges to 0 quadratically, i.e., the case (b) holds. This completes the proof. □

Some remarks about Algorithm Hybrid are in order.

- The system of nonlinear equations (A.3.13) has only $|P^k|$ variables actually, since $x_{N^k}=0$ and $x_{C^k}=0$. Hence, (A.3.13) becomes smaller than the original problem.
- The condition $\|H_F(x^k)\|\le M_1$ in Step 2 is not needed to establish Theorem A.5. However, using this condition, we can skip useless calculation which may occur when

  $$P(x^*)\ne P^k=P^{k-1}, N(x^*)\ne N^k=N^{k-1}, C(x^*)\ne C^k=C^{k-1}, P^k\cup N^k\cup C^k=\{1,2,\ldots,n\}.$$

- We may replace the condition in Step 2 by

  $$P^k=\cdots=P^{k-j}, N^k=\cdots=N^{k-j}, C^k=\cdots=C^{k-j}, P^k\cup N^k\cup C^k=\{1,2,\ldots,n\},$$

  where j is any positive integer. This modification may also enable us to omit unnecessary calculation when the index sets are not identified correctly.
- From Theorem A.3, the PPA has a genuine superlinear rate of convergence if x^* is nondegenerate. Consequently, when we judge x^* to be nondegenerate in Step 2, we may immediately return to Step 1 instead of proceeding to Step 3.

A.4 Concluding Remarks

In this paper, we have constructed the system of nonlinear equations which is useful in dealing with degenerate NCP, and proposed an algorithm based on it. We have shown that the proposed algorithm has a quadratic or superlinear rate of convergence even if NCP has a degenerate solution.

Finally, we mention some future research topics.

(i) In order to guarantee that Algorithm Hybrid has a quadratic or superlinear rate of convergence, we need Assumption A.2. In Lemma A.1, we give a sufficient condition for Assumption A.2 to hold, but this condition is imposed on the solution set itself. It is an interesting and important subject to find a milder and/or simpler sufficient condition for Assumption A.2.

(ii) We have used the technique proposed in [12] to identify P^*, N^* and C^*. However, it is difficult to confirm that $P^k = P^*, N^k = N^*$ and $C^k = C^*$ are attained. If we have a criterion to make sure that the correct identification of the index sets has been accomplished, we may design an algorithm that avoids the useless calculation in Step 3 of Algorithm Hybrid.

References

[1] D. P. Bertsekas, *Nonlinear Programming* (Athena Scientific, Massachusetts, 1995).

[2] X. Chen, Smoothing Methods for Complementarity Problems and their Applications: A Survey, *Journal of the Operations Research Society of Japan* 43 (2000) 32-47.

[3] H. Dan, N. Yamashita and M. Fukushima, Convergence Properties of the Inexact Levenberg-Marquardt Method under Local Error Bound Conditions, Technical Report 2001-001, Department of Applied Mathematics and Physics, Kyoto University (January 2001).

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

[5] F. Facchinei and C. Kanzow, A nonsmooth inexact Newton Method for the solution of large-scale nonlinear complementarity problems, *Mathematical Programming* 76 (1997) 493-512.

[6] F. Facchinei and C. Kanzow, Beyond monotonicity in regularization methods for complementarity problems, *SIAM Journal on Control and Optimization* 37 (1999) 1150-1161.

[7] M. C. Ferris and J.-S. Pang, Engineering and economic applications of complementarity problem, *SIAM Review* 39 (1997) 669-713.

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

[9] R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, New York, 1987).

[10] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic, New York, 1970).

[11] N. Yamashita and M. Fukushima, The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem, to appear in *SIAM Journal on Optimization.*

[12] N. Yamashita, H. Dan and M. Fukushima, On the identification of degenerate set of the nonlinear complementarity problem with the proximal point algorithm, Technical Report 2001-003, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University (February 2001).