Convergence Properties of the Inexact Levenberg-Marquardt Method under Local Error Bound Conditions[1]

Hiroshige Dan, Nobuo Yamashita and Masao Fukushima

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

January 16, 2001

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.

Keywords: the Levenberg-Marquardt method; Local error bound; Inexact method; Superlinear convergence

[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

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 [5] 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, 3, 4] 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 [9] 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 for (1.1) if there exist positive constants b and c such that

$$b\,\mathrm{dist}(x, X^*) \leq \|F(x)\| \qquad \forall x \in \{x \in \Re^n \mid \|F(x)\| \leq c\},$$

where X^* is the solution set of (1.1). Concrete examples of local error bounds can be found in [8]. 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^* [2]. Moreover, $\|F(x)\|$ may provide a local error bound even if $F'(x)$ is singular at a solution [9]. 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 [9], 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

$$r^k := \left(F'\left(x^k\right)^T F'\left(x^k\right) + \mu_k I\right) d^k + F'\left(x^k\right)^T F\left(x^k\right). \tag{1.6}$$

The vector r^k is a residual vector associated with an approximate solution d^k. Facchinei and Kanzow [3] 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 [9], 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 [9] 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 [7, 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

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

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*

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

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*

$$\begin{aligned}
\left\|d^k\right\| &\leq c_3 \operatorname{dist}\left(x^k, X^*\right) + \frac{\left\|r^k\right\|}{\mu_k}, && (2.6)\\
\left\|F'\left(x^k\right) d^k + F\left(x^k\right)\right\| &\leq c_4 \operatorname{dist}\left(x^k, X^*\right)^{1+\frac{\delta}{2}} + \left\|F'\left(x^k\right)\right\| \frac{\left\|r^k\right\|}{\mu_k}, && (2.7)
\end{aligned}$$

where r^k *is the residual vector given by (1.6),* $c_3 = \sqrt{\frac{c_1^2}{c_2^\delta}\left(\frac{b_1}{2}\right)^{2-\delta} + 1}$ *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

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

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

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

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

$$\begin{aligned}
\mu_k &= \left\|F\left(x^k\right)\right\|^\delta \geq c_2^\delta \left\|\bar{x}^k - x^k\right\|^\delta, && (2.9)\\
\mu_k &= \left\|F\left(x^k\right)\right\|^\delta = \left\|F(\bar{x}^k) - F\left(x^k\right)\right\|^\delta \leq L^\delta \left\|\bar{x}^k - x^k\right\|^\delta. && (2.10)
\end{aligned}$$

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

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

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 (1.6), 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

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

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

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

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

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

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

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

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 (1.6) 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*

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

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 $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$ 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

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

Since

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

we have from Lemma 2.2

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

Next we consider the case $k \geq 1$. 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 $\operatorname{dist}\left(x^l, X^*\right) \leq \left\|x^* - x^l\right\| \leq \bar{b} \leq b_3$ and Lemma 2.3 that

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

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

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

Consequently, we obtain

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

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, $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$ 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 $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$ converges to 0.

Note that we have

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

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

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

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 $\{x^k\}$ 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 [9]. 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*

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

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,* $\{\operatorname{dist}(x^k, X^*)\}$ *is convergent to* 0 *quadratically. Moreover,* $\{x^k\}$ *converges to a solution* $\hat{x} \in B\left(x^*, \bar{b}\right)$.

Proof: From (2.18), we immediately obtain

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

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

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

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 [3].

Theorem 3.1 *Let* $\{x^k\}$ *be a sequence generated by Algorithm 3.1. If the residual vector* r^k *given by (1.6) 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* $\nu_k = o\left(\operatorname{dist}\left(x^k, X^*\right)\right)$, *then any accumulation point of* $\{x^k\}$ *is a stationary point of* ϕ. *Moreover, if an accumulation point* x^* *of* $\{x^k\}$ *is a solution of (1.1) that satisfies Assumption 2.1, then* $\left\{\operatorname{dist}\left(x^k, X^*\right)\right\}$ *converges to* 0 *superlinearly.*

Proof: First we show that any accumulation point of $\{x^k\}$ is a stationary point of ϕ. Let $K_1 := \left\{k \;\middle|\; \left\|F\left(x^k + d^k\right)\right\| \le \gamma\left\|F\left(x^k\right)\right\|\right\}$. If K_1 is an infinite set, then we have $\left\|F\left(x^k\right)\right\| \to 0$ as $k \to \infty$ because $\left\{\left\|F\left(x^k\right)\right\|\right\}$ 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 $\left\|F\left(x^k + d^k\right)\right\| > \gamma\left\|F\left(x^k\right)\right\|$ 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

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

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

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 $K_2 := \left\{k \in K \;\middle|\; d^k = -\nabla\phi\left(x^k\right)\right\}$. 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 (1.6) 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 $\mu_k \geq \xi > 0$ 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 (1.6) and (3.4) that

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

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 (1.6)

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

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

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

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

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

This contradicts the assumption that the limit point of $\{x^k\}_{k\in K}$ 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 $\{r^k\}$ 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 $\|r^k\|$ satisfies (3.4), we have for any convergent subsequence $\{x^k\}_{k\in K}$

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

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

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

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 $\{y^l\}$ generated by the ILMM with $y^0 = x^{\bar{k}}$ and unit step size satisfies

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

where $\bar{y}^l$ denotes one of the nearest solutions from y^l, that is, $\bar{y}^l \in X^*$ and $\|\bar{y}^l - y^l\| = \operatorname{dist}\left(y^l, X^*\right)$. (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

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

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 $\{x^k\}$ be a sequence generated by Algorithm 3.1 with $\delta = 2$. If the residual vector r^k given by (1.6) satisfies*

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

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

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

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

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

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, 6] 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.

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

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.

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 $\eta = 0.001$, 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 $\left\|F\left(x^k\right)\right\| = O\left(\operatorname{dist}\left(x^k, X^*\right)\right)$. 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 $x^0 = x^{0,1}$.

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

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 [9], 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 [9] 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] 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.

[3] 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.

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

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

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

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

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

[9] 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).