On the Convergence of a New Levenberg-Marquardt Method *

Jin-yan Fan Ya-xiang Yuan
State Key Laboratory of Scientific/Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
The Academy of Mathematics and Systems Sciences,
Chinese Academy of Sciences, P.O.Box 2719, Beijing, 100080, P.R.China,
E-mail: fjy@lsec.cc.ac.cn, yyx@lsec.cc.ac.cn

Abstract

We propose a new way to choose the parameter in the Levenberg-Marquardt method for solving nonlinear equations $F(x) = 0$, where $F(x) : R^n \to R^m$ is continuously differentiable and $F'(x)$ is Lipschitz continuous. The sequence generated by the new method converges to the solution quadratically, if $||F(x)||_2$ provides a local error bound for the system of nonlinear equations. Numerical results show that the method performs well for singular problems.

Keywords: nonlinear equations, Levenberg-Marquardt method, quadratic convergence.

1 Introduction

We consider the problem for solving nonlinear equations

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

where $F(x) : R^n \to R^m$ is continuously differentiable and $F'(x)$ is Lipschitz continuous. Throughout the paper, we assume that the solution set of (1.1) is nonempty and denote it by X^*. And in all cases $||\cdot||$ refers to the 2-norm.

When $n = m$, the classical Levenberg-Marquardt method (see [1, 2]) for nonlinear equations (1.1) computes the trial step by

$$d_k = -(J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T F(x_k), \tag{1.2}$$

where $J(x_k) = F'(x_k)$ is the Jocabi, and $\mu_k \geq 0$ is a parameter being updated from iteration to iteration. Levenberg-Marquardt step (1.2) is a modification of the Newton's step

$$d_k^N = -J(x_k)^{-1} F(x_k). \tag{1.3}$$

*Supported by Chinese National Science Foundation grant 19731010 and the Knowledge Innovation Program of CAS.

The parameter μ_k is used to prevent d_k from being too large when $J(x_k)$ is nearly singular. Furthermore, when $J(x_k)$ is singular, the Newton's step is undefined. A positive μ_k guarantees that (1.2) is well defined. In some sense, the Levenberg-Marquardt method can also be regarded as a trust region algorithm [3, 9].

It is well known that the Levenberg-Marquardt method has a quadratic rate of convergence, if the Jacobi at the solution x^* is nonsingular, the initial point is sufficiently close to x^*, and if the parameter is chosen suitablely at each step. However, the condition of the nonsingularity of $J(x^*)$ is too strong. Recently, paper [8] shows, under the weaker condition that $||F(x)||$ provides a local error bound near the solution, the Levenberg-Marquardt method still has a quadratic convergence if the parameter is chosen as $\mu_k = ||F(x_k)||^2$. This is a very interesting result. However, the quadratic term $\mu_k = ||F(x_k)||^2$ has some unsatisfactory properties. In fact, when the sequence is close to the solution set, $\mu_k = ||F(x_k)||^2$ may be smaller than the machine eps, and it will lose its role. On the other hand, when the sequence is far away from the solution set, $\mu_k = ||F(x_k)||^2$ may be very large, and the step d_k will be too small, so it prevents the sequence converges quickly. Because of these observations, we consider the choice $\mu_k = ||F(x_k)||$. We prove that with this parameter, the quadratic convergence of the Levenberg-Marquardt method still holds if $||F(x)||$ provides a local error bound near the solution.

Definition 1.1. *Let N be a subset of R^n such that $N \cap X^* \neq \emptyset$. We say that $||F(x)||$ provides a local error bound on N for system (1.1), if there exists a positive constant $c > 0$ such that*

$$||F(x)|| \geq c\,dist(x, X^*), \qquad \forall x \in N.$$

Note that, if $J(x^*)$ is nonsingular at a solution x^* of (1.1), then x^* is an isolated solution, hence $||F(x)||$ provides a local error bound on some neighbourhood of x^*. However, the converse is not necessarily true, see example in [8]. So a local error bound condition is weaker than that of the nonsingularity.

In the next section, we show that under the local error bound condition, the sequence generated by the new Levenberg-Marquardt method without line search converges to the solution quadratically. In section 3, the global convergence result is given when the line search is used. Finally in section 4, we present numerical results for some singular nonlinear equations.

2 Local convergence of the Levenberg-Marquardt method

To study the convergence properties of the method, we make the following assumptions.

Assumption 2.1. *(a) $F(x)$ is continuously differentiable, and the Jacobi $J(x)$ is Lipschitz continuous on some neighbourhood of $x^* \in X^*$, i. e., there exist positive constants L_1 and $b_1 < 1$ such that*

$$||J(y) - J(x)|| \leq L_1||y - x||, \qquad \forall x, y \in N(x^*, b_1) = \{x \mid ||x - x^*|| \leq b_1\}. \tag{2.1}$$

(b) $||F(x)||$ provides a local error bound on $N(x^*, b_1)$ for the system (1.1), i.e., there exists a constant $c_1 > 0$ such that

$$||F(x)|| \geq c_1\,dist(x, X^*), \qquad \forall x \in N(x^*, b_1). \tag{2.2}$$

Note that, by Assumption 2.1(a), we have

$$||F(y) - F(x) - J(x)(y-x)|| \le L_1||y-x||^2, \qquad \forall x, y \in N(x^*, b_1), \tag{2.3}$$

and, there exists a constant $L_2 > 0$ such that

$$||F(y) - F(x)|| \le L_2||y-x||, \qquad \forall x, y \in N(x^*, b_1). \tag{2.4}$$

We discuss the local convergence of the Levenberg-Marquardt method without line search, i. e., the next iterate x_{k+1} is computed by

$$x_{k+1} = x_k + d_k,$$

where d_k is given by (1.2). For simplification, we use the notations $F_k = F(x_k)$, $J_k = J(x_k)$. And we assume

Assumption 2.2. *$\mu_k = ||F(x_k)||$ for all k.*

Paper [8] shows the quadratic convergence of the Levenberg-Marquardt method by choosing $\mu_k = ||F(x_k)||^2$, and bases its analyses on an unconstrained optimization problem. Here, we first prove the superlinear convergence of the new Levenberg-Marquardt method when choosing $\mu_k = ||F(x_k)||$, then, based on the singular value decomposition of the Jacobian matrix, we obtain the quadratic convergence. In the following, we denote $\bar{x}_k$ the vector in X^* that satisfies

$$||x_k - \bar{x}_k|| = \mathrm{dist}(x_k, X^*).$$

Lemma 2.1. *Under the conditions of Assumptions 2.1 and 2.2, if $x_k \in N(x^*, b_1)$, then there exists a constant $c_2 > 0$ such that*

$$||d_k|| \le c_2\, dist(x_k, X^*). \tag{2.5}$$

Proof. It follows from Assumption 2.1(b), Assumption 2.2, and (2.4) that the Levenberg-Marquardt parameter μ_k satisfies

$$c_1||\bar{x}_k - x_k|| \le \mu_k = ||F_k|| \le L_2||\bar{x}_k - x_k||. \tag{2.6}$$

Define

$$v_k = -J_k^T F_k - (J_k^T J_k + \mu_k I)(\bar{x}_k - x_k),$$

then from (2.3) and (2.4), we have

$$\begin{array}{rcl} ||v_k|| & \le & ||J_k|| ||F_k + J_k(\bar{x}_k - x_k)|| + \mu_k||\bar{x}_k - x_k|| \\ & \le & L_2(L_1+1)||\bar{x}_k - x_k||^2. \end{array}$$

Let $w_k = d_k - (\bar{x}_k - x_k)$, it is easy to see that

$$(J_k^T J_k + \mu_k I)w_k = v_k.$$

Thus it follows that

$$\begin{array}{rcl} ||w_k|| & \le & \mu_k^{-1}||v_k|| \\ & \le & \dfrac{1}{c_1}||\bar{x}_k - x_k||^{-1}||v_k|| \\ & \le & \dfrac{L_2(L_1+1)}{c_1}||\bar{x}_k - x_k||. \end{array}$$

The above inequality and the definition of w_k imply that

$$||d_k|| \le c_2 \text{dist}(x_k, X^*),$$

where $c_2 = L_2(L_1+1)/c_1 + 1$. $\square$

Lemma 2.2. *Under the conditions of Assumptions 2.1 and 2.2, if $x_k \in N(x^*, b_1)$, then we have*

$$dist(x_k + d_k, X^*) \le c_3 \, dist(x_k, X^*)^{\frac{3}{2}}, \tag{2.7}$$

where $c_3 = (\sqrt{L_1^2 + L_2} + L_1 c_2^2)/c_1$.

Proof. Define

$$\varphi_k(d) = ||F_k + J_k d||^2 + \mu_k ||d||^2.$$

From (1.2), d_k is a stationary point of $\varphi_k(d)$. The convexity of $\varphi_k(d)$ indicates that d_k is also a minimizer of $\varphi_k(d)$. Thus, using $b_1 < 1$, we have

$$\begin{array}{rcl} \varphi_k(d_k) & \le & \varphi_k(\bar{x}_k - x_k) \\ & = & ||F_k + J_k(\bar{x}_k - x_k)||^2 + \mu_k ||\bar{x}_k - x_k||^2 \\ & \le & L_1^2 ||\bar{x}_k - x_k||^4 + L_2 ||\bar{x}_k - x_k||^3 \\ & \le & (L_1^2 + L_2) ||\bar{x}_k - x_k||^3. \end{array}$$

Thus,

$$\begin{array}{rcl} ||F(x_k + d_k)|| & \le & ||F_k + J_k d_k|| + L_1 ||d_k||^2 \\ & \le & \sqrt{\varphi_k(d_k)} + L_1 ||d_k||^2 \\ & \le & \sqrt{L_1^2 + L_2} ||\bar{x}_k - x_k||^{\frac{3}{2}} + L_1 c_2^2 ||\bar{x}_k - x_k||^2 \\ & \le & (\sqrt{L_1^2 + L_2} + L_1 c_2^2) ||\bar{x}_k - x_k||^{\frac{3}{2}}. \end{array}$$

Hence

$$\text{dist}(x_k + d_k, X^*) \le \frac{1}{c_1} ||F(x_k + d_k)|| \le c_3 \text{dist}(x_k, X^*)^{\frac{3}{2}}.$$

The proof is completed. $\square$

Theorem 2.1. *Under the conditions of Assumptions 2.1 and 2.2, if x_0 is chosen sufficient close to X^*, then we have $x_{k+1} = x_k + d_k$ converges superlinearly.*

Proof. From (2.7), if

$$\text{dist}(x_0, X^*) < c_3^{-2},$$

we have

$$\text{dist}(x_{k+1}, X^*) < \text{dist}(x_k, X^*) \qquad \text{for all} \quad k \ge 0.$$

Therefore, we can assume that all x_k are in $N(x^*, b_1)$. Again, from (2.7), it follows that

$$\sum_{k=0}^{\infty} \text{dist}(x_k, X^*) < +\infty,$$

which implies, due to Lemma 2.1, that

$$\sum_{k=0}^{\infty} ||d_k|| < +\infty.$$

Thus $\{x_k\}$ converges to some point $\bar{x} \in X^*$. It is obviously that

$$\text{dist}(x_k, X^*) \leq \text{dist}(x_k + d_k, X^*) + ||d_k||.$$

The above inequality and (2.7) imply that

$$\text{dist}(x_k, X^*) \leq 2||d_k|| \tag{2.8}$$

for all large k. Thus from (2.5), (2.8) and (2.7) we see that $||d_{k+1}|| = O(||d_k||^{\frac{3}{2}})$. Therefore $\{x_k\}$ converges to the solution $\bar{x}$ superlinearly. $\square$

Without loss of generality, we assume that $\{x_k\}$ converges to $x^* \in X^*$, and the singular value decomposition (SVD) of $J(x^*)$ is

$$\begin{aligned} J(x^*) &= U^* \Sigma^* V^{*T} \\ &= U^* \begin{pmatrix} \sigma_1^* & & & & & \\ & \ddots & & & & \\ & & \sigma_r^* & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} V^{*T} \\ &= U_1^* \Sigma_1^* V_1^{*T}, \end{aligned}$$

where $\sigma_1^* \geq \sigma_2^* \geq \cdots \geq \sigma_r^* > 0$ and $\text{rank}(\Sigma_1^*) = r$. Suppose the SVD of $J(x_k)$ and its decomposite form are as follows

$$\begin{aligned} J(x_k) &= U_k \Sigma_k V_k^T \\ &= (U_{k,1}, U_{k,2}, U_{k,3}) \begin{pmatrix} \sigma_1^{(k)} & & & & & & & & \\ & \ddots & & & & & & & \\ & & \sigma_r^{(k)} & & & & & & \\ & & & \sigma_{r+1}^{(k)} & & & & & \\ & & & & \ddots & & & & \\ & & & & & \sigma_{r+q}^{(k)} & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix} \begin{pmatrix} V_{k,1}^T \\ V_{k,2}^T \\ V_{k,3}^T \end{pmatrix} \\ &= U_{k,1} \Sigma_{k,1} V_{k,1}^T + U_{k,2} \Sigma_{k,2} V_{k,2}^T, \end{aligned} \tag{2.9}$$

where $\Sigma_{k,1}, \Sigma_{k,2} > 0$, $\text{rank}(\Sigma_{k,1}) = r$ and $\text{rank}(\Sigma_{k,2}) = q \geq 0$. In the following, if the context is clear, we submit the subscription k in $\Sigma_{k,i}$ and $U_{k,i}, V_{k,i} (i = 1, 2, 3)$. Consequently, (2.9) can be written as

$$J_k = U_1 \Sigma_1 V_1^T + U_2 \Sigma_2 V_2^T.$$

To prove the sequence converges quadratically, we first give the following lemma.

Lemma 2.3. *Under the conditions of Assumption 2.1, if $x_k \in N(x^*, b_1)$, then we have*
(a) $||U_1U_1^TF_k|| \le L_2||x_k - \bar{x}_k||$;
(b) $||U_2U_2^TF_k|| \le 2L_1||x_k - x^*||^2$;
(c) $||U_3U_3^TF_k|| \le L_1||x_k - \bar{x}_k||^2$.

Proof. Result (a) follows immediately from (2.4). By the theory of matrix perturbation [7] and Assumption 2.1(a), we have

$$||\mathrm{diag}(\Sigma_1 - \Sigma_1^*, \Sigma_2, 0)|| \le ||J_k - J^*|| \le L_1||x_k - x^*||.$$

The above relation gives

$$||\Sigma_1 - \Sigma_1^*|| \le L_1||x_k - x^*|| \qquad \text{and} \qquad ||\Sigma_2|| \le L_1||x_k - x^*||. \tag{2.10}$$

Let $s_k = -J_k^+F_k$, where J_k^+ is the pseudo-inverse of J_k. It is easy to see that s_k is the least squares solution of $\min||F_k + J_ks||$, so we obtain from (2.3) that

$$||U_3U_3^TF_k|| = ||F_k + J_ks_k|| \le ||F_k + J_k(\bar{x}_k - x_k)|| \le L_1||x_k - \bar{x}_k||^2.$$

Let $\tilde{J}_k = U_1\Sigma_1V_1^T$ and $\tilde{s}_k = -\tilde{J}_k^+F_k$. Since $\tilde{s}_k$ is the least squares solution of $\min||F_k + \tilde{J}_ks||$, it follows from (2.3) and (2.10) that

$$\begin{aligned} ||(U_2U_2^T + U_3U_3^T)F_k|| &= ||F_k + \tilde{J}_k\tilde{s}_k|| \\ &\le ||F_k + \tilde{J}_k(\bar{x}_k - x_k)|| \\ &\le ||F_k + J_k(\bar{x}_k - x_k)|| + ||(\tilde{J}_k - J_k)(\bar{x}_k - x_k)|| \\ &\le L_1||\bar{x}_k - x_k)||^2 + ||U_2\Sigma_2V_2^T(\bar{x}_k - x_k)|| \\ &\le L_1||\bar{x}_k - x_k||^2 + L_1||x^* - x_k||||\bar{x}_k - x_k|| \\ &\le 2L_1||x_k - x^*||^2. \end{aligned}$$

Due to the orthogonality of U_2 and U_3, we get result (b). □

Theorem 2.2. *Under the conditions of Assumptions 2.1 and 2.2, if the sequence $\{x_k\}$ be generated by the new Levenberg-Marquardt method without line search with x_0 sufficiently close to x^*. Then $\{x_k\}$ converges to the solution quadratically.*

Proof. By the SVD of J_k, we know the step at the current iterate is

$$d_k = -V_1(\Sigma_1^2 + \mu_kI)^{-1}\Sigma_1U_1^TF_k - V_2(\Sigma_2^2 + \mu_kI)^{-1}\Sigma_2U_2^TF_k. \tag{2.11}$$

So we have

$$\begin{aligned} F_k + J_kd_k &= F_k - U_1\Sigma_1(\Sigma_1^2 + \mu_kI)^{-1}\Sigma_1U_1^TF_k - U_2\Sigma_2(\Sigma_2^2 + \mu_kI)^{-1}\Sigma_2U_2^TF_k \\ &= \mu_kU_1(\Sigma_1^2 + \mu_kI)^{-1}U_1^TF_k + \mu_kU_2(\Sigma_2^2 + \mu_kI)^{-1}U_2^TF_k + U_3U_3^TF_k. \end{aligned} \tag{2.12}$$

Since $\{x_k\}$ converges to x^* superlinearly, without loss of generality, we assume that $L_1||x_k - x^*|| < \sigma_r^*/2$ holds for all sufficient large k. Thus we abtain from (2.9) that

$$||(\Sigma_1^2 + \mu_kI)^{-1}|| \le ||\Sigma_1^{-2}|| \le \frac{1}{(\sigma_r^* - L_1||x_k - x^*||)^2} < \frac{4}{\sigma_r^{*2}},$$

and

$$||(\Sigma_2^2 + \mu_k I)^{-1}|| \le \mu_k^{-1}.$$

The above two inequalities, together with (2.6) and Lemma 2.3 imply that

$$\begin{aligned} ||F_k + J_k d_k|| &\le \frac{4L_2^2}{\sigma_r^{*2}}||x_k - x^*||^2 + 3L_1||x_k - x^*||^2 \\ &= (\frac{4L_2^2}{\sigma_r^{*2}} + 3L_1)||x_k - x^*||^2. \end{aligned} \tag{2.13}$$

Let $c_4 = 4L_2^2/\sigma_r^{*2} + 3L_1$, then we get

$$\begin{aligned} c_1 \mathrm{dist}(x_{k+1}, X^*) &\le ||F(x_k + d_k)|| \\ &\le ||F_k + J_k d_k|| + L_1||d_k||^2 \\ &\le (c_4 + c_2^2 L_1)||x_k - x^*||^2. \end{aligned}$$

It now follows from (2.8) and Theorem 2.1 that

$$||d_{k+1}|| = O(||d_k||^2).$$

The above relation implies that $\{x_k\}$ converges quadratically, namely,

$$||x_{k+1} - x^*|| = O(||x_k - x^*||^2).$$

This completes our proof. $\square$

3 Global convergence of the Levenberg-Marquardt method

In this section, we discuss the global convergence properties of the Levenberg-Marquardt method. For any initial point x_0, to guarantee the sequence $\{x_k\}$ converges to a solution of the system of nonlinear equations (1.1) or to a stationary point of the merit function

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

we use the line search techniques. That is, we compute the next iterate by

$$x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is computed by a certain rule. One commonly used inexact line search is the Wolfe line search which finds an $\alpha_k > 0$ that satisfies

$$||F(x_k + \alpha_k d_k)||^2 \le ||F(x_k)||^2 + \alpha_k \beta_1 F_k^T J_k d_k \tag{3.1}$$

and

$$F(x_k + \alpha_k d_k)^T J(x_k + \alpha_k d_k) d_k \ge \beta_2 F_k^T J_k d_k, \tag{3.2}$$

where $\beta_1 \le \beta_2$ are two constants in $(0,1)$. Another famous inexact line search is the Armijo line search which sets $\alpha_k = \delta^t \bar{\alpha}$, where $\bar{\alpha}$ is a positive constant and t is the smallest non-negative integer satisfying

$$||F(x_k + \delta^t \bar{\alpha} d_k)||^2 \le ||F(x_k)||^2 + \beta_1 \delta^t \bar{\alpha} F_k^T J_k d_k. \tag{3.3}$$

Both inexact line searches imply that

$$||F(x_{k+1})||^2 \leq ||F(x_k)||^2 - \beta_1\beta_3 \frac{(F_k^T J_k d_k)^2}{||d_k||^2}, \tag{3.4}$$

where β_3 is some positive constant. For more details, see [10].

Algorithm 3.1. *(New Levenberg-Marquardt method with line search)*

Step 1 *Given* $x_0 \in R^n, \beta_1, \beta_2, \eta \in (0,1), \varepsilon > 0, k := 0.$

Step 2 *If* $||J_k^T F_k|| \leq \varepsilon$ *then stop; Set* $\mu_k := ||F(x_k)||$; *Compute* d_k *by (1.2).*

Step 3 *If* d_k *satisfies*

$$||F(x_k + d_k)|| \leq \eta ||F(x_k)||, \tag{3.5}$$

then $x_{k+1} = x_k + d_k$; *Compute the step size* α_k *by the Wolfe or Armijo line search,* $x_{k+1} = x_k + \alpha_k d_k$.

Step 4 $k := k+1$; *go to Step 2.*

Theorem 3.1. *Suppose Assumptions 2.1 and 2.2 hold. Let the sequence* $\{x_k\}$ *be generated by the Levenberg-Marquardt method with inexact line search. Then* $\{x_k\}$ *converges to a stationary point of* ϕ. *Moreover, if the stationary point* x^* *is a solution of nonlinear equations (1.1), then* $\{x_k\}$ *converges to the solution quadratically.*

Proof. It follows from (3.4) and (3.5) that $\{||F(x_k)||\}$ is monotonically decreasing, hence $\{x_k\}$ converges to a stationary point x^* of ϕ.

We now proceed to prove the second part of the theorem. Since $\{x_k\}$ converges to the solution of nonlinear equations (1.1), due to Theorem 2.2, it is sufficient to prove that the step size $\alpha_k = 1$ for all large k in Algorithm 3.1. In view of (2.6), we see

$$\frac{||F_{k+1}||}{||F_k||} \leq \frac{L_2 ||x_{k+1} - \bar{x}_{k+1}||}{c_1 ||x_k - \bar{x}_k||} \to 0.$$

Thus, we have $\{x_k\}$ converges to the solution quadratically. □

4 Numerical results

We tested our new Levenberg-Marquardt method with unit step size on some singular problems, and compared it to the choice of $\mu_k = ||F_k||^2$.

First we test the Powell singular function [5], where $n = 4$ and $\text{rank}(J(x^*)) = 2$. The results are given in Table 1.

The other test problems were created by modifying the nonsingular problems given by Moré, Garbow and Hillstrom [4], and have the same form as in [6]

$$\hat{F}(x) = F(x) - J(x^*)A(A^TA)^{-1}A^T(x - x^*), \tag{4.1}$$

where $F(x)$ is the standard nonsingular test function, x^* is its root, and $A \in R^{n \times k}$ has full column rank with $1 \leq k \leq n$. Obviously, $\hat{F}(x^*) = 0$ and

$$\hat{J}(x^*) = J(x^*)(I - A(A^TA)^{-1}A^T)$$

has rank $n-k$. A disadvantage of these problems is that $\hat{F}(x)$ may have roots that are not roots of $F(x)$. We created two sets of singular problems, with $\hat{J}(x^*)$ having rank $n-1$ and $n-2$, by using

$$A \in R^{n\times 1}, \qquad A^T = (1, 1, \cdots, 1)$$

and

$$A \in R^{n\times 2}, \qquad A^T = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & -1 & 1 & -1 & \cdots & \pm 1 \end{pmatrix},$$

respectively. Meanwhile, we made a slight alteration on the variable dimension problem, which has $n+2$ equations in n unknowns; we eliminate the $n-1$th and nth equations. (The first n equations in the standard problem are linear.)

Table 1
Results on Powell singular problem.

			$\mu_k = \|\|F(x_k)\|\|$			$\mu_k = \|\|F(x_k)\|\|^2$		
Function	n	x_0	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?
Powell singular	4	1	12	.20-7	Y	14	.19-8	Y
		10	33	.86-8	Y	–	–	–
		100	198	.84-8	Y	–	–	–

Table 2
Results on first singular test set with $\text{rank}(F'(x^*)) = n-1$.

			$\mu_k = \|\|F(x_k)\|\|$			$\mu_k = \|\|F(x_k)\|\|^2$		
Function	n	x_0	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?
Brown alm.linear	10	1	8	.80-8	Y	9	.15-8	Y
		10	22	.11-7	Y	34	.10-7	Y
		100	46	.13-8	Y	–	–	–
Broyden banded	30	1	15	.40-12	Y	16	.97-13	Y
		10	22	.72-13	Y	–	–	–
		100	27	.38-12	Y	–	–	–
Broyden tridiag.	30	1	10	.27-10	Y	8	.33-12	Y
		10	19	.10-10	Y	–	–	–
		100	26	.31-12	Y	–	–	–
Discrete boundary	10	1	3	.14-7	N	2	.27-8	N
		10	–	–	–	22	.12-10	N
		100	99	.11-9	N	–	–	–
Discrete integral	30	1	16	.85-8	Y	4	.76-13	Y
		10	68	.14-7	N	35	.27-10	N
		100	30	.60-13	N	–	–	–
Helical valley	3	1	17	.45-12	N	15	.26-12	N
		10	21	.32-13	N	67	.11-14	N
		100	31	.18-13	N	–	–	–
Rosenbrock	2	1	42	.15-12	Y	23	.97-14	Y
		10	62	.13-12	Y	124	.21-13	Y
Trigonometric	30	1	5	.62-12	N	6	.11-13	N
		10	11	.11-10	N	12	.28-12	N
Variable dimension (altered)	10	1	13	.25-9	Y	14	.79-9	Y
		10	15	.12-9	Y	111	.25-9	Y
		100	35	.10-7	Y	–	–	–
Wood	4	1	24	.16-12	Y	27	.56-13	Y
		10	30	.58-13	Y	–	–	–
		100	61	.16-12	Y	–	–	–

The algorithm is terminated when the norm of $J_k^T F_k$, e.g.,the derivative of ϕ, is less than 10^{-5}, or when the number of the iterations exceeds 150. We compare the Levenberg-Marquardt method by using $\mu_k = ||F_k||$ with the choice of $\mu_k = ||F_k||^2$. The results for the first set problems of rank $n-1$ are listed in Table 2, and the second set of rank $n-2$ in Table 3. The third column of the table indicates that the starting point is $x_0, 10x_0$, and $100x_0$, where x_0 is suggested by Moré, Garbow and Hillstrom in [4]; "iter" represents the number of iterations; $\frac{1}{2}||F(x^*)||^2$ represents the half squares norm of the function at the final iterate, using abbreviated notation: e.g. .11−7 means 0.11×10^{-7}; and "n.s.x^*?" gives a Y(yes) if the method converges to the same solution as the corresponding nonsingular problem, a N(no) otherwise. If the method failed to find the solution in 150 iterations, we denoted it by the sign "–".

Table 3
Results on second singular test set with $\text{rank}(F'(x^*)) = n - 2$.

			$\mu_k = \|\|F(x_k)\|\|$			$\mu_k = \|\|F(x_k)\|\|^2$		
Function	n	x_0	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?	iters	$\frac{1}{2}\|\|F(x^*)\|\|^2$	n.s.x^*?
Brown alm.linear	10	1	8	.80-8	Y	9	.15-8	Y
		10	22	.11-7	Y	34	.10-7	Y
		100	55	.35-16	N	–	–	–
Broyden banded	30	1	14	.70-12	Y	17	.22-12	Y
		10	21	.86-13	Y	–	–	–
		100	26	.68-12	Y	–	–	–
Broyden tridiag.	30	1	13	.69-8	Y	–	–	–
		10	22	.65-8	Y	–	–	–
		100	29	.53-8	Y	–	–	–
Discrete boundary	10	1	3	.19-7	N	–	–	–
		100	112	.13-8	N	–	–	–
Discrete integral	30	1	8	.48-6	N	5	.19-9	N
		10	121	.51-6	N	42	.21-11	N
Helical valley	3	1	28	.53-12	Y	20	.28-12	Y
		10	34	.41-12	Y	73	.62-13	Y
		100	82	.43-12	Y	–	–	–
Rosenbrock	2	1	10	.28-8	Y	11	.83-8	Y
		10	13	.11-8	N	57	.87-9	N
		100	16	.27-8	N	–	–	–
Variable dimension (altered)	10	1	13	.25-9	Y	14	.79-9	Y
		10	15	.12-9	N	108	.10-9	N
		100	35	.20-9	N	–	–	–
Wood	4	1	13	.32-10	N	22	.12-10	N
		10	16	.43-10	N	–	–	–
		100	20	.10-10	N	–	–	–

The $100x_0$ case is excluded for three problems, Rosenbrock and trigonometric when $\text{rank}(F'(x^*)) = n - 1$, and Discrete integral when $\text{rank}(F'(x^*)) = n - 2$, because both methods failed to find the solution in 150 iterations. And the choice $\mu_k = ||F_k||^2$ is undesired for trigonometric problem when $\text{rank}(F'(x^*)) = n - 2$; the method failed from either of the three starting points.

From the results, we can see that when the starting point is near the root, the results favor the choice $\mu_k = ||F_k||^2$ slightly if $\text{rank}(J(x^*)) = n - 1$. However for $\text{rank}(J(x^*)) = n - 2$, the choice of $\mu_k = ||F_k||$ maybe preferable.

When the starting point is far away from the root, our new method almost always performs much better than the method adopting $\mu_k = ||F_k||^2$, whatever the rank of $J(x^*)$ is. The second method usually can not solve the problem in 150 iterations.

References

[1] K. Levenberg, A method for the solution of certain nonlinear problems in least squares, *Quart. Appl. Math.,* 2(1944), 164-166.

[2] D. W. Marquardt, An algorithm for least-squares estimation of nonlinear inequalities, *SIAM J. Appl. Math.,* 11(1963), 431-441.

[3] J. J. Moré, Recent developments in algorithms and software for trust region methods, In: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical Programming: The State of Art* (Springer, Berlin, 1983) 258-287.

[4] J. J. Moré, B. S. Garbow and K. H. Hillstrom, Testing unconstrained optimization software, *ACM Trans. Math. Software,* 7(1981), 17-41.

[5] M. J. D. Powell, An iterative method for finding stationary values of a function of several variables, *Comput. J.* , 5(1962), 147-151.

[6] R. B. Schnabel and P. D. Frank, Tensor methods for nonlinear equations, *SIAM J. Numer. Anal.,* 21(1984), 815-843.

[7] G. W. Stewart and J. G. Sun, Matrix Perturbation Theory, Academic Press, San Diego, CA, 1990.

[8] N. Yamashita and M. Fukushima, On the rate of convergence of the Levenberg-Marquardt method, *Report, presented in the first Sino-Japan optimization meeting,* October, 2000.

[9] Y. X. Yuan, Trust region algorithms for nonlinear equations, *Information,* 1(1998), 7-20.

[10] Y. X. Yuan, Problems on convergence of unconstrained optimization algorithms, In: Y.X.Yuan, ed.,*Numerical Linear Algebra and Optimization*, Science Press, Beijing, 1999, 95-107.