On a Global Complexity Bound of the Levenberg-Marquardt Method*

Kenji Ueda and Nobuo Yamashita†

Abstract

In this paper, we investigate a global complexity bound of the Levenberg-Marquardt method (LMM) for the nonlinear least squares problem. The global complexity bound for an iterative method solving unconstrained minimization of ϕ is an upper bound on the number of iterations required to get an approximate solution such that $\|\nabla\phi(x)\| \leq \epsilon$. We show that the global complexity bound of the LMM is $O(\epsilon^{-2})$.

Keywords Levenberg-Marquardt methods, Global complexity bound, Scale parameter

Mathematics Subject Classification (2000) 90C30, 65K05, 49M15

1 Introduction

In this paper, we consider the nonlinear least squares problem of finding a local minimizer of

$$\phi(x) := \frac{1}{2}\|F(x)\|^2, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable mapping.

The Levenberg-Marquardt method (LMM) is one of the solution methods for (1.1) [1, 5, 6, 10]. For a current point x^k, the LMM adopts a search direction $d^k(\mu_k)$ given by

$$d^k(\mu_k) := -(J(x^k)^T J(x^k) + \mu_k I)^{-1} J(x^k)^T F(x^k), \tag{1.2}$$

where $J(x^k)$ is the Jacobian of F at x^k and μ_k is a positive parameter. Taking $\mu_k \to \infty$, we have $\|d^k(\mu_k)\| \to 0$ and $d^k(\mu_k)/\|d^k(\mu_k)\| \to J(x^k)^T F(x^k)/\|J(x^k)^T F(x^k)\|$. Therefore, $\phi(x^k + d^k(\mu_k)) < \phi(x^k)$ for μ_k sufficiently large, and hence the LMM converges globally if μ_k is appropriately updated. In this paper, we call μ_k the scale parameter of the search direction.

In order to guarantee a global convergence of the LMM, many updating rules of the scale parameter μ_k have been proposed [5, 6, 10]. Moré [5] proposed the updating rule based on the idea of the trust-region method [4]. A search direction $d^k(\mu_k)$ of his proposal is given as a solution of a subproblem

$$\begin{aligned} &\underset{d\in\mathbb{R}^n}{\text{minimize}} \ \ \|F(x^k) + J(x^k)d\|, \\ &\text{subject to } \|d\| \leq \Delta_k, \end{aligned}$$

and the scale parameter μ_k corresponds to the Lagrange multiplier of the Karush-Kuhn-Tucker conditions of the subproblem. Then, instead of directly updating μ_k, Δ_k is controlled for global convergence. Thus, the LMM with his updating rule requires to solve the subproblem at each iteration. On the other hand, Osborne [6] proposed a simple and direct updating rule of μ_k, and showed the global convergence of the LMM with the rule. Since μ_k is directly given, $d^k(\mu_k)$ is a solution of the linear equations which

*Technical Report, 2009-020, December 14, 2009

†Corresponding author. Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, E-mail: nobuo@i.kyoto-u.ac.jp, Tel: 81-75-7534759, Fax: 81-75-7534759

is much easier to solve than the above subproblem. The details of his proposal is presented in Section 2.

Recently, global complexity bounds have been vigorously discussed for solution methods of unconstrained minimization problems [2, 3, 7, 8, 9, 11, 13]. The global complexity bound of an iterative method for unconstrained minimization of f is an upper bound on the number of iterations required to get an approximate solution such that $\|\nabla f(x)\| \le \epsilon$, where ϵ is a given positive constant. The bound is useful when we solve large-scale problems and we want to estimate the worst computational time for a given accuracy of a solution in advance. Until now, some bounds for a steepest descent method and Newton-type methods have been presented [2, 3, 7, 8, 9, 11, 13]. Thus, if we apply the steepest descent method or the Newton-type methods to the least squares problem (1.1), then we can estimate the worst computational time in advance. However, since these method are not specialized to nonlinear least squares problems, they are not efficient. In fact, the Newton-type methods [2, 3, 7, 8, 9, 13] require the second derivative of F, and have to solve nonconvex subproblems at each iteration. Moreover, although the steepest descent method requires only the Jacobian of F, its convergence is slow in general. Thus, it is worth investigating a global complexity bound for methods specified for (1.1). Recently, Nesterov [12] proposed a modified Gauss-Newton method for solving a system of nonlinear equations and gave interesting results. However, the modified Gauss-Newton method also has to solve computationally expensive subproblems. To the authors' knowledge, the global complexity bound for the LMM remains unknown.

In this paper, we investigate the global complexity bound for the LMM with the Osborne's simple updating rule of the scale parameter. In particular, we show that the global complexity bound is $O(\epsilon^{-2})$ without assuming the nonsingularity of $J(x)^T J(x)$.

Throughout the paper, we use the following notations. For a vector $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm defined by $\|x\| := \sqrt{x^T x}$. For a symmetric matrix $M \in \mathbb{R}^{n\times n}$, we denote the maximum eigenvalue and the minimum eigenvalue of M as $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$, respectively. For a matrix $M \in \mathbb{R}^{n\times m}$, $\|M\|$ denotes the ℓ_2 norm of M defined by $\|M\| := \sqrt{\lambda_{\max}(M^T M)}$. If M is symmetric positive semidefinite matrix, then $\|M\| = \lambda_{\max}(M)$. $B(x, r)$ denotes the closed sphere with center x and radius r, i.e., $B(x, r) := \{y \in \mathbb{R}^n \mid \|y - x\| \le r\}$. For sets $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^n$, $S_1 + S_2$ denotes the sum of S_1 and S_2 defined by $S_1 + S_2 := \{x + y \in \mathbb{R}^n \mid x \in S_1, y \in S_2\}$.

2 The global complexity bound of the LMM

First, we explain the updating rule of the scale parameter μ_k proposed by Osborne [6]. Then, we give a global complexity bound of the LMM with the rule. In what follows, we denote the LMM with the Osborne's updating rule as the LMM for simplicity. Moreover, x^k denotes the k-th iterative point, and F_k and J_k denotes $F(x^k)$ and $J(x^k)$, respectively.

The LMM adopts a search direction $d^k(\mu_k)$ defined by (1.2), and controls μ_k directly as follows. Let $\psi_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a model function of ϕ at x^k defined by

$$\psi_k(d, \mu) := \frac{1}{2}\|F_k + J_k d\|^2 + \frac{1}{2}\mu\|d\|^2. \tag{2.1}$$

Note that $d^k(\mu_k)$ is a global minimizer of $\psi_k(\cdot, \mu_k)$. Let $\rho_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the ratio of the reduction of the objective function value to that of the model function value, i.e.,

$$\rho_k(d, \mu) := \frac{\phi(x^k) - \phi(x^k + d)}{\phi(x^k) - \psi_k(d, \mu)}. \tag{2.2}$$

If $\rho_k(d^k(\mu_k), \mu_k)$ is large, then we adopt $d^k(\mu_k)$ and decrease the parameter μ_k. On the other hand, if $\rho_k(d^k(\mu_k), \mu_k)$ is small, then we increase μ_k and compute $d^k(\mu_k)$ once again.

The precise description of the LMM is as follows.

The Levenberg-Marquardt Method

Step 0 : Choose parameters $\epsilon, \mu_0, \delta, \gamma_1, \gamma_2, \eta_1, \eta_2$ such that

$$0 < \epsilon < 1,\ \mu_0 > 0,\ \delta \geq 0,\ \gamma_1 < 1 < \gamma_2,\ 0 < \eta_1 \leq \eta_2 \leq 1.$$

Choose a starting point x^0. Set $k := 0$.

Step 1 : If $\|J_k^T F_k\| \leq \epsilon$, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\mu}_{l_k} = \mu_k$.

Step 2.1 : Compute

$$d^k(\bar{\mu}_{l_k}) = -(J_k^T J_k + \bar{\mu}_{l_k} I)^{-1} J_k^T F_k.$$

Step 2.2 : Compute

$$\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) = \frac{\phi(x^k) - \phi(x^k + d^k(\bar{\mu}_{l_k}))}{\phi(x^k) - \psi_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k})}.$$

If $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) < \eta_1$, then update $\bar{\mu}_{l_k+1} := \gamma_2 \bar{\mu}_{l_k}$, set $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $\eta_2 > \rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq \eta_1$, then update $\mu_{k+1} := \bar{\mu}_{l_k}$.
If $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq \eta_2$, then update $\mu_{k+1} := \gamma_1 \bar{\mu}_{l_k}$.
Update $x^{k+1} = x^k + d^k(\bar{\mu}_{l_k})$. Set $k := k + 1$, and go to Step 1.

Osborne [6] showed that the LMM has a global convergence property under appropriate conditions.

Next, we discuss the global complexity bound of the LMM. In what follows, for simplicity, we denote l_k and $\bar{\mu}_{l_k}$ of the last iteration in the inner loops of Steps 2.0–2.2 at each k as l_k^* and μ_k^*, respectively.

Let K_{outer} be the total number of outer iterations when the algorithm terminates. If there does not exist such K_{outer}, we define $K_{\text{outer}} := \infty$. Moreover, let K_{total} be the total number of inner iterations such that $k < K_{\text{outer}}$, i.e.,

$$K_{\text{total}} := \sum_{k=0}^{K_{\text{outer}}-1} l_k^*.$$

Note that we cannot estimate the total computational time from K_{outer}. In contrast, K_{total} means the total number of solving linear equations. Therefore, the main task of the paper is to investigate K_{total}. To this end, we firstly make the following assumption.

Assumption 1. *The level set of ϕ at the initial point x^0 is compact, i.e., $\Omega := \{x \in \mathbb{R}^n \mid \phi(x) \leq \phi(x^0)\}$ is compact.*

Since $\{\phi(x^k)\}$ is monotonically decreasing, the sequence $\{x^k\}$ is included in the compact set Ω. Moreover, since F is continuously differentiable, there exist positive constants U_F and U_J such that

$$\|F(x)\| \leq U_F, \quad \max(\|J(x)\|, \|J(x)^T\|) \leq U_J, \quad \forall x \in \Omega. \tag{2.3}$$

Now, we give bounds of eigenvalues of $(J_k^T J_k + \mu I)^{-1}$.

Lemma 2.1. *Suppose that Assumption 1 holds. Then, for any $\mu \in (0, \infty)$,*

$$\lambda_{\max}((J_k^T J_k + \mu I)^{-1}) \leq \frac{1}{\mu},$$
$$\lambda_{\min}((J_k^T J_k + \mu I)^{-1}) \geq \frac{1}{U_J^2 + \mu}.$$

Proof. Since $J_k^T J_k$ is positive semidefinite, we have

$$\lambda_{\max}((J_k^T J_k + \mu I)^{-1}) = \frac{1}{\lambda_{\min}(J_k^T J_k + \mu I)} \leq \frac{1}{\mu}.$$

On the other hand, we have from (2.3) that

$$\lambda_{\min}((J_k^T J_k + \mu I)^{-1}) = \frac{1}{\lambda_{\max}(J_k^T J_k + \mu I)} = \frac{1}{\|J_k\|^2 + \mu} \geq \frac{1}{U_J^2 + \mu}.$$

This completes the proof. $\square$

The next lemma indicates that $\|d^k(\mu)\|$ is bounded above when $\mu \in [\mu_0, \infty)$.

Lemma 2.2. *Suppose that Assumption 1 holds. Then, for any* $\mu \in [\mu_0, \infty)$,

$$\|d^k(\mu)\| \leq U_d,$$

where $U_d := \frac{U_F U_J}{\mu_0}$.

Proof. We have from (1.2) that

$$\begin{aligned}
\|d^k(\mu)\| &= \|(J_k^T J_k + \mu I)^{-1} J_k^T F_k\| \\
&\leq \|(J_k^T J_k + \mu I)^{-1}\| \cdot \|J_k^T\| \cdot \|F_k\| \\
&\leq U_J U_F \lambda_{\max}\Big((J_k^T J_k + \mu I)^{-1}\Big) \\
&\leq \frac{U_F U_J}{\mu_0},
\end{aligned}$$

where the second inequality follows from (2.3), and the last inequality follows from Lemma 2.1 and $\mu \geq \mu_0$. $\square$

In what follows, we further assume that J is Lipschitz continuous on the compact set $\Omega + B(0, U_d)$, where U_d is the constant given in Lemma 2.2.

Assumption 2. *There exists* $L > 0$ *such that*

$$\|J(x) - J(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega + B(0, U_d).$$

Note that if F is twice continuously differentiable, then Assumption 2 holds.

Next, we show that the scale parameter μ_k is bounded above.

Lemma 2.3. *Suppose that Assumptions 1 and 2 hold. Then,*

$$\mu_k^* \leq U_\mu,$$

where $U_\mu := \gamma_2 \max(\mu_0, 2(U_F + U_J U_d)L + L^2 U_d^2)$.

Proof. Since F is continuously differentiable, we have

$$F(x^k + d^k(\mu)) = F_k + J_k d^k(\mu) + \int_0^1 (J(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau. \tag{2.4}$$

Since $x + \tau d^k(\mu) \in \Omega + B(0, U_d)$ for all $\mu \in [\mu_0, \infty)$ and $\tau \in [0, 1]$, it then follows from Assumption 2 that for $\mu \geq \mu_0$,

$$\|J(x^k + \tau d^k(\mu)) - J(x^k)\| \leq L\tau\|d^k(\mu)\| \leq L\|d^k(\mu)\|. \tag{2.5}$$

Moreover, we have from (2.3) and Lemma 2.2 that for $\mu \geq \mu_0$,

$$\|F_k + J_k d^k(\mu)\| \leq \|F_k\| + \|J_k\| \cdot \|d^k(\mu)\| \leq U_F + U_J U_d. \tag{2.6}$$

Now we suppose that $\mu \geq \mu_0$. It then follows from (1.1), (2.1) and (2.4) that

$$\begin{aligned}
\phi(x^k + d^k(\mu)) - \psi_k(d^k(\mu), \mu) &= \frac{1}{2}\|F(x^k + d^k(\mu))\|^2 - \frac{1}{2}\|F_k + J_k d^k(\mu)\|^2 - \frac{1}{2}\mu\|d^k(\mu)\|^2 \\
&= \frac{1}{2}\left\|F_k + J_k d^k(\mu) + \int_0^1 (J(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau\right\|^2 \\
&\quad - \frac{1}{2}\|F_k + J_k d^k(\mu)\|^2 - \frac{1}{2}\mu\|d^k(\mu)\|^2 \\
&= (F_k + J_k d^k(\mu))^T \int_0^1 (J(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau \\
&\quad + \frac{1}{2}\left\|\int_0^1 (J(x^k + \tau d^k(\mu)) - J_k) d^k(\mu) \mathrm{d}\tau\right\|^2 - \frac{1}{2}\mu\|d^k(\mu)\|^2 \\
&\leq \|F_k + J_k d^k(\mu)\| \int_0^1 \|J(x^k + \tau d^k(\mu)) - J_k\| \cdot \|d^k(\mu)\| \mathrm{d}\tau \\
&\quad + \frac{1}{2}\int_0^1 \|J(x^k + \tau d^k(\mu)) - J_k\|^2 \cdot \|d^k(\mu)\|^2 \mathrm{d}\tau - \frac{1}{2}\mu\|d^k(\mu)\|^2 \\
&\leq (U_F + U_J U_d) L \|d^k(\mu)\|^2 + \frac{1}{2} L^2 U_d^2 \|d^k(\mu)\|^2 - \frac{1}{2}\mu\|d^k(\mu)\|^2 \\
&= \frac{2(U_F + U_J U_d) L + L^2 U_d^2 - \mu}{2}\|d^k(\mu)\|^2,
\end{aligned} \tag{2.7}$$

where the second inequality follows from (2.5), (2.6) and Lemma 2.2.

We further suppose that $\mu \geq \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$. It then follows from (2.7) that

$$\phi(x^k + d^k(\mu)) \leq \psi_k(d^k(\mu), \mu),$$

and hence

$$\rho_k(d^k(\mu), \mu) = \frac{\phi(x^k) - \phi(x^k + d^k(\mu))}{\phi(x^k) - \psi_k(d^k(\mu), \mu)} \geq 1.$$

Thus, if $\bar{\mu}_{l_k} \geq \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$, then inner loops of Step 2 must terminate. Therefore, if $\bar{\mu}_1 \geq \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$ at the k-th iteration, then $\mu_k^* = \bar{\mu}_1$. On the other hand, if $\bar{\mu}_1 < \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$, then μ_k^* must satisfy $\mu_k^* \leq \gamma_2 \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$. Otherwise, $\bar{\mu}_{l_k^*-1} > \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2)$, which contradicts $\rho_k(d^k(\bar{\mu}_{l_k^*-1}), \bar{\mu}_{l_k^*-1}) < \eta_1 < 1$. Consequently, we have from the updating rule of μ that

$$\begin{aligned}
\mu_k^* &\leq \max(\bar{\mu}_1, \gamma_2\mu_0, \gamma_2(2(U_F + U_J U_d) L + L^2 U_d^2)) \\
&\leq \max(\mu_{k-1}^*, \gamma_2\mu_0, \gamma_2(2(U_F + U_J U_d) L + L^2 U_d^2)) \\
&\leq \cdots \leq \max(\mu_0, \gamma_2\mu_0, \gamma_2(2(U_F + U_J U_d) L + L^2 U_d^2)) \\
&= \gamma_2 \max(\mu_0, 2(U_F + U_J U_d) L + L^2 U_d^2).
\end{aligned}$$

This completes the proof. $\square$

By using the above lemma, we give a lower bound of the reduction of the objective function when $k < K_{\mathrm{outer}}$.

Lemma 2.4. *Suppose that Assumptions 1 and 2 hold. Then, for all k such that $k < K_{\mathrm{outer}}$,*

$$\phi(x^k) - \phi(x^{k+1}) > p\epsilon^2.$$

where $p := \frac{\eta_1}{2(U_J^2 + U_\mu)}$.

Proof. First note that $\|J_k^T F_k\| > \epsilon$ for $k < K_{\text{outer}}$. We have from Lemmas 2.1 and 2.3 that

$$\lambda_{\min}\left((J_k^T J_k + \mu_k^* I)^{-1}\right) \geq \frac{1}{U_J^2 + U_\mu}. \tag{2.8}$$

Since $\rho_k(d^k(\mu_k^*), \mu_k^*) \geq \eta_1$ from the definition of μ_k^*, we have

$$\phi(x^k) - \phi(x^{k+1}) \geq \eta_1(\phi(x^k) - \psi_k(d^k(\mu_k^*), \mu_k^*)).$$

It then follows from (1.1) and (2.1) that

$$\begin{aligned}
\phi(x^k) - \phi(x^{k+1}) &\geq \frac{\eta_1}{2}(\|F_k\|^2 - \|F_k + J_k d^k(\mu_k^*)\|^2 - \mu_k^*\|d^k(\mu_k^*)\|^2) \\
&= \eta_1(-F_k^T J_k d^k(\mu_k^*) - \frac{1}{2} d^k(\mu_k^*)^T (J_k^T J_k + \mu_k^* I) d^k(\mu_k^*)) \\
&= \frac{\eta_1}{2} F_k^T J_k (J_k^T J_k + \mu_k^* I)^{-1} J_k^T F_k \\
&\geq \frac{\eta_1}{2} \lambda_{\min}\left((J_k^T J_k + \mu_k^* I)^{-1}\right) \|J_k^T F_k\|^2 \\
&> \frac{\eta_1}{2(U_J^2 + U_\mu)} \epsilon^2,
\end{aligned}$$

where the second equality follows from (1.2), and the last inequality follows from (2.8) and $\|J_k^T F_k\| > \epsilon$. □

Now, we give the global complexity bound K_{outer}.

Theorem 2.1. *Suppose that Assumptions 1 and 2 hold. Then,*

$$K_{\text{outer}} \leq \left\lceil \frac{\phi(x^0)}{p} \epsilon^{-2} + 1 \right\rceil.$$

Proof. Let K be $\lceil (\phi(x^0)\epsilon^{-2}/p) + 1 \rceil$. Suppose the contrary, i.e., $K_{\text{outer}} > K$. Then, we have from Lemma 2.4 that

$$\phi(x^0) \geq \phi(x^0) - \phi(x^K) = \sum_{j=0}^{K-1} (\phi(x^j) - \phi(x^{j+1})) > \sum_{j=0}^{K-1} p\epsilon^2 = p\epsilon^2 K. \tag{2.9}$$

On the other hand, we have from the definition of K that

$$p\epsilon^2 K = p\epsilon^2 \left\lceil \left(\frac{\phi(x^0)}{p\epsilon^2}\right) + 1 \right\rceil > \phi(x^0).$$

This contradicts (2.9), and hence we obtain the result of the theorem. □

By using Theorem 2.2, we show the main theorem of the paper.

Theorem 2.2. *Suppose that Assumptions 1 and 2 hold. Then,*

$$K_{\text{total}} \leq \left\lceil \log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0 \gamma_1^{K_{\text{outer}}}}\right) + 1 \right\rceil,$$

and hence $K_{\text{total}} = O(\epsilon^{-2})$.

Proof. Suppose the contrary, i.e., $K_{\text{total}} > \lceil \log_{\gamma_2}(U_\mu \gamma_2^{K_{\text{outer}}}/\mu_0 \gamma_1^{K_{\text{outer}}}) + 1 \rceil$. The number of satisfying $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) < \eta_1$ is $\sum_{k=0}^{K_{\text{outer}}-1}(l_k^* - 1)$. Moreover, the number of satisfying $\rho_k(d^k(\bar{\mu}_{l_k}), \bar{\mu}_{l_k}) \geq \eta_2$ is at most K_{outer}. Thus, we have from the updating rule of μ_k that

$$\begin{aligned}
\mu_{K_{\text{outer}}-1}^* &\geq \mu_0 \gamma_2^{\sum_{k=0}^{K_{\text{outer}}-1}(l_k^*-1)} \gamma_1^{K_{\text{outer}}} \\
&= \mu_0 \gamma_2^{K_{\text{total}}} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} \\
&> \mu_0 \gamma_2^{\log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0 \gamma_1^{K_{\text{outer}}}}\right)} \gamma_2^{-K_{\text{outer}}} \gamma_1^{K_{\text{outer}}} \\
&= U_\mu,
\end{aligned}$$

where the last inequality follows from the assumption that $K_{\text{total}} > \lceil \log_{\gamma_2}(U_\mu \gamma_2^{K_{\text{outer}}}/\mu_0\gamma_1^{K_{\text{outer}}}) + 1\rceil$. This contradicts Lemma 2.3. Therefore, we have from Theorem 2.1 that

$$\begin{aligned} K_{\text{total}} &\le \left\lceil \log_{\gamma_2}\left(\frac{U_\mu \gamma_2^{K_{\text{outer}}}}{\mu_0\gamma_1^{K_{\text{outer}}}}\right) + 1 \right\rceil \\ &= \lceil K_{\text{outer}}(1 - \log_{\gamma_2}\gamma_1) + \log_{\gamma_2} U_\mu - \log_{\gamma_2}\mu_0 + 1 \rceil \\ &\le \left\lceil \left\lceil \frac{\phi(x^0)}{p}\epsilon^{-2} + 1 \right\rceil (1 - \log_{\gamma_2}\gamma_1) + \log_{\gamma_2} U_\mu - \log_{\gamma_2}\mu_0 + 1 \right\rceil, \end{aligned}$$

and hence $K_{\text{total}} = O(\epsilon^{-2})$. □

Note that the theorem does not assume the nonsigurality of $J(x)^T J(x)$ or $J(x)J(x)^T$. Note also that since $J(x)^T F(x) = 0$ does not imply $F(x) = 0$, the theorem does not provide a global complexity bound of $\|F_k\| \le \hat{\epsilon}$ for some positive constant $\hat{\epsilon}$. Now, by assuming the nonsingularity of $J(x)J(x)^T$, we give such a bound as a direct application of Theorem 2.2.

Corollary 2.1. *Suppose that there exists a positive constant σ such that $\lambda_{\min}(J(x)J(x)^T) \ge \sigma$ for all $x \in \Omega$. Suppose also that Assumptions 1 and 2 hold. Let $\hat{\epsilon}$ be a positive constant, and let the termination criterion in Step 1 be replaced with $\|F_k\| \le \hat{\epsilon}$. Then, $\hat{K}_{\text{total}} = O(\hat{\epsilon}^{-2})$, where $\hat{K}_{\text{total}}$ is the total number of inner iterations of the modified LMM.*

Proof. Now, we consider the original LMM with $\epsilon = \sqrt{\sigma}\hat{\epsilon}$. Since $\lambda_{\min}(J_k J_k^T) \ge \sigma$ by the assumption, we have $\|J_k^T F_k\| \ge \sqrt{\sigma}\|F_k\|$, and hence $\|J_k^T F_k\| \le \sqrt{\sigma}\hat{\epsilon}$ implies $\|F_k\| \le \hat{\epsilon}$. Thus, $\hat{K}_{\text{total}}$ is less than or equal to the total number of inner iterations of the original LMM. From Theorem 2.1 we have $\hat{K}_{\text{total}} \le K_{\text{total}} = O((\sqrt{\sigma}\hat{\epsilon})^{-2}) = O(\hat{\epsilon}^{-2})$. □

References

[1] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, New York, 1995.

[2] C. Cartis, N. I. M. Gould, and P. L. Toint, *Adaptive cubic overestimation methods for unconstrained optimization*, Technical Report 07/05, Department of Mathematics, FUNDP - University of Namur, 2007.

[3] ——, *On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization problems*, Technical Report 09/14, Department of Mathematics, FUNDP - University of Namur, 2009.

[4] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, USA, 2000.

[5] J. J. Moré, *The Levenberg-Marquardt algorithm: Implementation and theory*, Numerical Analysis, 630 (1978), pp. 105–116.

[6] M. R. Osborne, *Nonlinear least squares — the Levenberg algorithm revisited*, Journal of the Australian Mathematical Society, 19 (1976), pp. 343–357.

[7] R. A. Polyak, *Regularized Newton method for unconstrained convex optimization*, Mathematical Programming, Series B, 120 (2009), pp. 125–145.

[8] K. Ueda and N. Yamashita, *Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization*, Applied Mathematics and Optimization, (2009). DOI: 10.1007/s00245-009-9094-9.

[9] ——, *A regularized Newton method without line search for unconstrained optimization*, Technical Report 2009-007, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, 2009. http://www.amp.i.kyoto-u.ac.jp/tecrep/ps_file/2009/2009-007.pdf.

[10] S. J. Wright and J. N. Holt, *An inexact Levenberg-Marquardt method for large sparse nonlinear least squares*, Journal of the Australian Mathematical Society, 26 (1985), pp. 387–403.

[11] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[12] ———, *Modified Gauss-Newton scheme with worst-case guarantees for global performance*, Optimization Methods and Software, 22 (2007), pp. 469–483.

[13] Yu. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Mathematical Programming, 108 (2006), pp. 177–205.