Convergence properties of Levenberg-Marquardt methods with generalized regularization terms

Shumpei Ariizumi[a], Yuya Yamakawa[a], Nobuo Yamashita[a]

[a] *Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan*

Abstract

Levenberg-Marquardt methods (LMMs) are the most typical algorithms for solving nonlinear equations $F(x) = 0$, where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function. They sequentially solve subproblems represented as squared residual of the Newton equations with the L_2 regularization to determine the search direction. However, since the subproblems of the LMMs are usually reduced to linear equations with n variables, it takes much time to solve them when $m \ll n$.

In this paper, we propose a new LMM which generalizes the L_2 regularization of the subproblems of the ordinary LMMs. By virtue of the generalization, we can choose a suitable regularization term for each given problem. Moreover, we show that a sequence generated by the proposed method converges globally and quadratically under some reasonable assumptions. Finally, we conduct numerical experiments to confirm that the proposed method performs better than the existing LMMs for some problems that satisfy $m \ll n$.

Keywords: Nonlinear equation; Levenberg-Marquardt method; Global convergence; Local convergence

1. Introduction

In this paper, we consider the following nonlinear equations:

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

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function and $\mathbb{R}^n$ denotes the n–dimensional Euclidean space. We write X^* for a solution set of (1).

Nonlinear equations arise from many fields such as engineering, economics, and so on [1, 2, 3]. Since solving nonlinear equations plays an important role in such fields, various types of methods for (1) have been proposed, such as Newton's methods [4, 5], Gauss–Newton methods [6, 7, 8], and Levenberg-Marquardt methods (LMMs) [4, 9, 10]. In particular, the Newton's methods are known as typical iterative algorithms and belong to the most powerful ones. As it is well known, the simplest Newton's methods sequentially solve a sequence of the Newton equations $L_k(x) = 0$, which are the first–order approximation of the nonlinear function F at each iteration point x^k, to determine the search direction and updates x^k along the search direction. However, there is no guarantee that the Newton equations are solvable unless the current point x^k is sufficiently close to the solution of (1). On the other hand, the Gauss–Newton methods sequentially minimize $\|L_k(x)\|^2/2$ to obtain the search direction instead of solving the Newton equations. Although these minimization subproblems are solvable unlike the Newton methods because it is a linear least squares (LLS) problem, the coefficient matrix of LLS is not necessarily regular, and hence it is unstable for numerical error. Moreover, it is generally known that the Gauss–Newton methods are ineffective for problems that the nonlinearity of F is strong or the norm of F at the solution of (1) is large. The LMMs have nice global and local convergence properties. The methods can be regarded as an improvement of the Gauss–Newton methods because their subproblems correspond to LLS with the L_2 regularization, that is, their solutions are unique.

Researches regarding the LMMs have a long history, and various types of them have been proposed so far. In [11, 12], a locally convergent LMM was proposed and its superlinear and quadratic convergence was also shown under some appropriate assumptions. Yamashita and Fukushima [11] proposed a globally convergent LMM equipped with Armijo's line search. As stated above, LMMs sequentially solve the subproblems, which are described as LLS with the L_2 regularization, at each iteration. Although the subproblems are generally reduced to linear equations with n variables, their scales become large if $m \ll n$, and hence it takes much time to solve them. Moreover, search directions obtained by solving the subproblems are generally dense.

In this paper, we propose a new LMM which generalizes the regularization term of the subproblem. Thus, it enables us to select a suitable function as the regularization term depending on a given problem. By virtue of the generalization, we can deal with problems equipped with various regularizations. For example, if we use the L_1 norm, the subproblem becomes an L_1–L_2 optimization [13], or if we adopt the L_2 norm cubed, it becomes the cubic regularization [14]. For the proposed LMM equipped with Armijo's line search, we provide the following two convergence properties. The former is its global convergence, and the latter is its local and quadratic convergence. Moreover, we conduct numerical experiments to confirm the usefulness of the proposed method.

This paper is organized as follows. In Section 2, we propose the generalized LMM with Armijo's line search. In Section 3, we show that the proposed method enjoys the global convergence property. Section 4 shows the local and quadratic convergence of the proposed LMM without Armijo's line search. In Section 5, we prove that the proposed method with Armijo's line search indeed has the local and quadratic convergence property under some appropriate conditions by utilizing the convergence result given in Section 4. Some numerical results are reported in Section 6. Finally, we conclude the paper and provide several future works.

We use the following notation in this paper. The identity matrix is represented by I and the zero matrix is represented by O. For a vector $x \in \mathbb{R}^n$, $\|x\|$ and $\|x\|_1$ indicate the L_2 norm and L_1 norm, respectively. For a vector $c \in \mathbb{R}^n$ and positive real number $r \in \mathbb{R}$, we define $N(c, r) := \{x \mid \|x - c\| \leq r\}$. For a matrix $A \in \mathbb{R}^{m\times n}$, $A^\top$ denotes the transpose of A and $\|A\|$ denotes the operator norm of A, which is defined by $\|A\| := \sup\{\|Ax\|/\|x\| \mid x \in \mathbb{R}^n, x \neq 0\}$. For matrices $A_1 \in \mathbb{R}^{m\times m}$ and $A_2 \in \mathbb{R}^{n\times n}$, $\mathrm{diag}\,(A_1, A_2)$ stands for the block diagonal matrix consisting of A_1 and A_2. For a function $G\colon \mathbb{R}^m \to \mathbb{R}^n$, $G'(a)$ is the Jacobian of G at a. For a vector $a \in \mathbb{R}^n$ and a set $X \subset \mathbb{R}^n$, $\mathrm{dist}(a, X)$ means the distance between a and X, which is defined by $\mathrm{dist}(a, X) := \inf\{\|x - a\| \mid x \in X\}$. For a function $G\colon \mathbb{R}^m \to \mathbb{R}^n$ and a positive scalar $t \in \mathbb{R}$, we write $G(x) = O(t)$ $(t \to 0)$ if there exist $C > 0$ and $\delta > 0$ such that $|t| < \delta$ implies $\|G(x)\| < C|t|$. We also write $G(x) = o(t)$ $(t \to 0)$ if for all $\varepsilon > 0$ there exists $\delta > 0$ such that $|t| < \delta$ implies $\|G(x)\| < \varepsilon|t|$.

2. Generalized Levenberg-Marquardt Method

In this section, we propose a new LLM which generalizes the regularization term included in the subproblem of the ordinary LMMs. Before providing the explanation of the proposed method, we recall the existing Gauss–Newton methods and LMMs as seen in the standard text books on continuous optimization. Note that the superscript k indicates the k-th iteration of the algorithms. The Gauss–Newton methods solve the following LLS to determine a search direction d^k:

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad \|F'(x^k)d + F(x^k)\|^2. \tag{2}$$

From convexity of the objective function and the first–order optimality condition, problem (2) is reduced to solving the following linear equation:

$$F'(x^k)^\top F'(x^k)d = -F'(x^k)^\top F(x^k). \tag{3}$$

When $F'(x^k)$ is full column rank, $F'(x^k)^\top F'(x^k)$ becomes nonsingular and equation (3) has a unique solution. However, since it is not always so, equation (3) generally has various solutions and its calculation might be unstable. On the other hand, the LMMs use a solution of the following linear equations:

$$(F'(x^k)^\top F'(x^k) + \mu_k I)d = -F'(x^k)^\top F(x^k), \tag{4}$$

where μ_k is a positive parameter. Since $F'(x^k)^\top F'(x^k) + \mu_k I$ is positive definite, equation (4) has a unique solution and is equivalent to the following minimization problem:

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad \|F'(x^k)d + F(x^k)\|^2 + \mu_k\|d\|^2,$$

which corresponds to LLS (2) with the L_2 regularization.

Now, we propose a generalization related to the regularization term $\mu_k\|d\|^2$. In particular, we consider the following subproblem:

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad \|F'(x^k)d + F(x^k)\|^2 + \mu_k\psi(d), \tag{5}$$

where $\mu_k\psi(d)$ is a generalized regularization term. Throughout this paper, we suppose that the function ψ satisfies the following condition: There exist constants $0 < \gamma_1 \leq \gamma_2$ and $0 < p \leq 3$ such that

$$\gamma_1\|d\|^p \leq \psi(d) \leq \gamma_2\|d\|^p \tag{6}$$

for all $d \in \mathbb{R}^n$. This generalized regularization term includes not only the L_2 norm squared [15] but also the L_1 norm [13] and the L_2 norm cubed [14].

The function ψ defined by (6) enjoys the following property when it is differentiable at $d = 0$:

$$\nabla\psi(0) = 0. \tag{7}$$

A simple proof is given below. We consider the following optimization problem:

$$\underset{d\in\mathbb{R}^n}{\text{minimize}} \quad \psi(d). \tag{8}$$

Problem (8) has a unique solution $d = 0$ from (6), and hence the first–order optimality condition of (8) implies (7).

As mentioned in Section 1, the traditional LMMs have local and global convergence properties. In particular, the global convergence has been shown by integrating Armijo's line search with the LMMs. For example, see [11, 12]. We also adopt Armijo's line search so that the proposed LMM converges globally. In the line search, we use the following merit function:

$$\phi(x) = \frac{1}{2}\|F(x)\|^2. \tag{9}$$

The formal statement of the proposed generalized LMM is given below.

Algorithm 1 (Generalized Levenberg-Marquardt Method)

1: Choose parameters $\alpha, \beta \in (0, 1)$ and an initial point $x^0 \in \mathbb{R}^n$. Set $k := 0$.
2: If $\|F'(x^k)^\top F(x^k)\| = 0$, then stop.
3: Find the solution d^k of (5).
4: Let m_k be the smallest nonnegative integer m such that

$$\phi(x^k + \beta^m d^k) - \phi(x^k) \leq \alpha \beta^m \nabla \phi(x^k)^\top d^k. \tag{10}$$

Set $x^{k+1} := x^k + \beta^{m_k} d^k$.
5: Set $k \leftarrow k + 1$, and go to Step 2.

We show that Algorithm 1 is well-defined. To this end, we make the following assumptions.

Assumption 1.

(1) *ψ is differentiable at neighborhood of $d = 0$ and is continuously differentiable at $d = 0$.*

(2) *Subgradient of ψ at neighborhood of $d = 0$ is bounded, i.e., there exist constants $M \in (0, \infty)$ and $b \in (0, 1)$ such that $\|\eta\| \leq M$ for all $\eta \in \partial\psi(d)$ and $d \in N(0, b)$.*

Note that (2) of Assumption 1 holds when ψ is norm.

Assumption 2.

(1) *When ψ is continuously differentiable at $d = 0$,*

$$\mu_k = \|F(x^k)\|^{2-p+\delta}, \quad \delta \in [1, 2].$$

(2) *When ψ is not continuously differentiable at $d = 0$,*

$$\mu_k = \min\left\{\|F(x^k)\|^{2-p+\delta}, \frac{\|F'(x^k)^\top F(x^k)\|}{M}\right\}, \quad \delta \in [1,2].$$

To prove that Algorithm 1 is well-defined, we have to show the following two properties.

- search direction is not zero until the termination criterion in Step 2 is satisfied
- the line search in Step 4 terminates finitely

We prove the former property in Lemma 1. The latter property is omitted because it clearly hold.

Lemma 1. *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Suppose that either (1) or (2) of Assumption 1 holds and that Assumption 2 holds. Then, the following inequalities hold:*

$$\|F(x^k)\|^2 \geq \|F'(x^k)d^k + F(x^k)\|^2 + \mu_k\psi(d^k),$$

$$\nabla\phi(x^k)^\top d^k \leq -\frac{1}{2}\|F'(x^k)d^k\|^2 - \frac{1}{2}\mu_k\psi(d^k).$$

Moreover, if x^k is not a stationary point of ϕ, i.e., $F'(x^k)^\top F(x^k) \neq 0$, then the solution d^k of subproblem (5) is not zero and satisfies $\nabla\phi(x^k)^\top d^k < 0$.

Proof. For simplicity, let us define $\theta_k(d) := \|F'(x^k)d + F(x^k)\|^2 + \mu_k\psi(d)$. Since d^k is an optimal solution of (5), we have

$$\|F(x^k)\|^2 = \theta_k(0) \geq \theta_k(d^k) = \|F'(x^k)d^k + F(x^k)\|^2 + \mu_k\psi(d^k). \tag{11}$$

It follows from (11) and $\nabla\phi(x^k) = F'(x^k)^\top F(x^k)$ that

$$\nabla\phi(x^k)^\top d^k \leq -\frac{1}{2}\|F'(x^k)d^k\|^2 - \frac{1}{2}\mu_k\psi(d^k). \tag{12}$$

We prove the second half of assertion by contradiction. Assume that $d^k = 0$. We divide the proof into the following two cases: (i) ψ is differentiable at $d = 0$; (ii) ψ is not differentiable at $d = 0$.

Case (i): Since d^k is a solution of (5), we have $2F'(x^k)^\top F'(x^k)d^k + 2F'(x^k)^\top F(x^k) + \mu_k\nabla\psi(d^k) = 0$. It then follows from $d^k = 0$ and (7) that

$\nabla\phi(x^k) = F'(x^k)^\top F(x^k) = 0$. This contradicts the fact that x^k is not a stationary point of ϕ.

Case (ii): Since d^k is a solution of (5), there exists $\eta \in \partial\psi(d^k)$ such that $2F'(x^k)^\top F'(x^k)d^k + 2F'(x^k)^\top F(x^k) + \mu_k\eta = 0$. Substituting $d^k = 0$ into the equality yields $2F'(x^k)^\top F(x^k) + \mu_k\eta = 0$. It follows from (2) of Assumption 1 and (2) of Assumption 2 that

$$\begin{aligned} 0 &= \|2F'(x^k)^\top F(x^k) + \mu_k\eta\| \geq 2\|F'(x^k)^\top F(x^k)\| - \mu_k\|\eta\| \\ &> 2\|F'(x^k)^\top F(x^k)\| - \frac{2\|F'(x^k)^\top F(x^k)\|}{M} \cdot M = 0, \end{aligned}$$

which contradicts. Moreover, since x^k is not a stationary point and $d^k \neq 0$, from (12), we have $\nabla\phi(x^k)^\top d^k < 0$. $\square$

3. Global Convergence of Algorithm 1

In this section, we prove that a sequence generated by Algorithm 1 converges globally.

Lemma 2. *Let $\bar{x}$ be a limit point of $\{x^k\}$. Suppose that either* (1) *or* (2) *of Assumption* 1 *holds and that Assumption* 2 *holds. If $\psi(d^k) \to 0$ $(k \to \infty)$, then $\bar{x}$ satisfies the stationary condition of* $\min\{\phi(x) \mid x \in \mathbb{R}^n\}$.

Proof. From (6), we have $0 \leq \gamma_1\|d^k\|^p \leq \psi(d^k)$. Utilizing $\psi(d^k) \to 0$ $(k \to \infty)$ yields $d^k \to 0$ $(k \to \infty)$. Since d^k is a solution of (5), there exists $\eta^k \in \partial\psi(d^k)$ such that $0 = \|2F'(x^k)^\top F'(x^k)d^k + 2F'(x^k)^\top F(x^k) + \mu_k\eta^k\| \geq 2\|F'(x^k)^\top F(x^k)\| - 2\|F'(x^k)^\top F'(x^k)d^k\| - \mu_k\|\eta^k\|$. Then we have

$$\|F'(x^k)^\top F(x^k)\| \leq \|F'(x^k)^\top F'(x^k)d^k\| + \frac{1}{2}\mu_k\|\eta^k\|. \tag{13}$$

Now we consider the following two cases: (i) ψ is continuously differentiable at $d = 0$; (ii) ψ is not continuously differentiable at $d = 0$.

Case (i): Recall that $\mu_k = \|F(x^k)\|^{2-p+\delta}$ from (1) of Assumption 2. Since $\nabla\psi$ is continuous and $d^k \to 0$ $(k \to \infty)$, it follows from (7) that $\nabla\psi(d^k) \to 0$ $(k \to \infty)$. Therefore, inequality (13) and the continuity of F and F' derive $\|F'(\bar{x})^\top F(\bar{x})\| \leq 0$.

Case (ii): Since $d^k \to 0$ $(k \to \infty)$ and (2) of Assumption 1 holds, we have $\|\eta^k\| \leq M$ for all sufficiently large k. It then follows from (2) of Assumption 2 and (13) that $\|F'(x^k)^\top F(x^k)\| \leq \|F'(x^k)^\top F'(x^k)d^k\| + \|F'(x^k)^\top F(x^k)\|/2$. Therefore, letting $k \to \infty$ implies $\|F'(\bar{x})^\top F(\bar{x})\| \leq \|F'(\bar{x})^\top F(\bar{x})\|/2$, which leads to $\|F'(\bar{x})^\top F(\bar{x})\| = 0$. $\square$

Theorem 1. *Let $\{x^k\}$ be a bounded sequence generated by Algorithm* 1*. Suppose that either* (1) *or* (2) *of Assumption* 1 *holds and that Assumption* 2 *holds. Then, any accumulation point of $\{x^k\}$ satisfies the stationary condition of* $\min\{\phi(x) \mid x \in \mathbb{R}^n\}$.

Proof. From Lemma 1, we have

$$\nabla\phi(x^k)^\top d^k \leq -\frac{1}{2}\mu_k\psi(d^k). \tag{14}$$

Combining (10) and (14) yields $\phi(x^{k+1}) - \phi(x^k) \leq \alpha\beta^{m_k}\nabla\phi(x^k)^\top d^k \leq -\alpha\beta^{m_k}\mu_k\psi(d^k)/2$, which implies

$$\phi(x^{k+1}) - \phi(x^0) = \sum_{i=0}^{k}\left(\phi(x^{i+1}) - \phi(x^i)\right) \leq -\frac{\alpha}{2}\sum_{i=0}^{k}\beta^{m_i}\mu_i\psi(d^i). \tag{15}$$

Now let $\bar{x}$ be an accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^k\}_K$ that converges to $\bar{x}$, where $K \subset \mathbb{N}\cup\{0\}$. If $K \ni k \to \infty$ in (15), then $\phi(x^{k+1}) - \phi(x^0)$ converges to $\phi(\bar{x}) - \phi(0)$, which implies that $\beta^{m_k}\mu_k\psi(d^k) \to 0$ $(k \to \infty)$. Hence, there are the two cases: (a) $\mu_k \to 0$ $(k \to \infty)$; (b) $\mu_k \not\to 0$ $(k \to \infty)$.

Case (a): From Assumption 2, it is clear that $\|F(\bar{x})\|^{2-p+\delta} = 0$ or $\|F'(\bar{x})^\top F(\bar{x})\|/M = 0$, that is, $\nabla\phi(\bar{x}) = F'(\bar{x})F(\bar{x}) = 0$.

Case (b): There exist $\bar{\mu} > 0$ and $J \subset K$ such that $\mu_k \geq \bar{\mu}$ for all $k \in J$. It follows from Lemma 1 and (6) that $\|F(x^k)\|^2 \geq \mu_k\psi(d^k) \geq \bar{\mu}\gamma_1\|d^k\|^p$, namely $\|d^k\|^p \leq \|F(x^k)\|^2/(\bar{\mu}\gamma_1)$. Since $\{x^k\}$ is bounded, so is $\{d^k\}$. Therefore without loss of generality, we assume that $\{d^k\}_K$ converges to $\bar{d}$. Now we consider the following two cases: (b1) $\beta^{m_k} \to 0$ $(k \to \infty)$; (b2) $\beta^{m_k} \not\to 0$ $(k \to \infty)$.

Case (b1): Since m_k is the smallest nonnegative integer that satisfies the Armijo's condition (10), $m_k - 1$ does not satisfy the condition, i.e., we have $\phi(x^k + \beta^{m_k-1}d^k) - \phi(x^k) > \alpha\beta^{m_k-1}\nabla\phi(x^k)^\top d^k$. From Mean Value Theorem, there exists a constant $\rho_k \in [0,1]$ such that $\phi(x^k + \beta^{m_k-1}d^k) - \phi(x^k) = \beta^{m_k-1}\nabla\phi(x^k + \rho_k\beta^{m_k-1}d^k)^\top d^k$. Then we have $\nabla\phi(x^k + \rho_k\beta^{m_k-1}d^k)^\top d^k > \alpha\nabla\phi(x^k)^\top d^k$. The continuity of $\nabla\phi$ ensures $\nabla\phi(\bar{x})^\top\bar{d} \geq \alpha\nabla\phi(\bar{x})^\top\bar{d}$. Since $\alpha \in (0,1)$, we obtain $\nabla\phi(\bar{x})^\top\bar{d} \geq 0$. Therefore, from (14), we have $\mu_k\psi(d^k) \to 0$ $(k \to \infty)$. Because of $\mu_k \not\to 0$ $(k \to \infty)$, it is clear that $\psi(d^k) \to 0$ $(k \to \infty)$. Therefore, from Lemma 2, we obtain that $\bar{x}$ satisfies the stationary condition.

Case (b2): Since $\psi(d^k) \to 0$ $(k \to \infty)$, Lemma 2 implies that $\bar{x}$ satisfies the stationary condition. □

4. Local Convergence without Line Search

This section shows the local convergence properties of Algorithm 2, which is equal to Algorithm 1 without Armijo's line search and provided below.

Algorithm 2 (Generalized Levenberg-Marquardt Method without Line Search)

1: Choose an initial point $x^0 \in \mathbb{R}^n$. Set $k := 0$.
2: If $\|F'(x^k)^\top F(x^k)\| = 0$, then stop.
3: Find the solution d^k of (5).
4: Set $x^{k+1} := x^k + d^k$.
5: Set $k \leftarrow k + 1$, and go to Step 2.

In particular, we first show superlinear convergence under appropriate assumptions. Moreover, we prove quadratic convergence by some additional assumptions. In the following, we discuss the local convergence around $x^* \in X^*$.

4.1. Superlinear Convergence

We make two assumptions for the local convergence.

Assumption 3.

(1) *The Jacobian of F is Lipschitz continuous on $N(x^*, b)$, i.e., there exists a positive constant L_1 such that*

$$\|F'(y) - F'(x)\| \leq L_1 \|y - x\|$$

for all $x, y \in N(x^*, b)$.

(2) *The norm of $F(x)$ provides a local error bound on $N(x^*, b)$ for the system* (1)*, i.e., there exists a positive constant c_1 such that*

$$c_1 \mathrm{dist}(x, X^*) \leq \|F(x)\|$$

for all $x \in N(x^*, b)$.

Assumption 4. *There exist constants $0 < \xi_1 \leq \xi_2$ such that*

$$\xi_1 \|F(x^k)\|^{2-p+\delta} \leq \mu_k \leq \xi_2 \|F(x^k)\|^{2-p+\delta}, \quad \delta \in [1, 2].$$

Remark 1. *The local error bound condition, that is,* (2) *of Assumption* 3*, holds if the Jacobian of F is nonsingular at x^*. Moreover, it is known that the condition is satisified if F is a piecewise linear function* [11].

To begin with, we provide the well-known results which can be derived from (1) of Assumption 3.

Lemma 3 ([5, Lemma 4.1.12]). *Suppose that* (1) *of Assumption* 3 *holds. Then*

$$\|F'(y)(x-y) - (F(x) - F(y))\| \leq \frac{L_1}{2}\|x-y\|^2$$

for all $x, y \in N(x^*, b)$. *Moreover, there exists a positive constant* L_2 *such that*

$$\|F(x) - F(y)\| \leq L_2\|x-y\|$$

for all $x, y \in N(x^*, b)$.

Now, we prove superlinear convergence of Algorithm 2. As stated in Section 1, Fan and Yuan [12] showed superlinear convergence of the ordinary LMM. Although the difference between the existing and proposed methods is the regularization term in (5), the same idea as [12] can be applied to the convergence analysis of the proposed method by using inequality (6) and Assumption 4 accordingly. Therefore, we omit proofs of the following Lemmas and Theorem.

Lemma 4 ([12, Lemma 2.1]). *Suppose that Assumptions* 3 *and* 4 *hold. Then there exists* $c_2 > 0$ *such that* $\|d^k\| \leq c_2 \mathrm{dist}(x^k, X^*)$ *for all* $x^k \in N(x^*, b/2)$.

Lemma 5 ([12, Lemma 2.2]). *Suppose that Assumptions* 3 *and* 4 *hold. Then there exists* $c_3 > 0$ *such that* $\mathrm{dist}(x^k + d^k, X^*) \leq c_3 \mathrm{dist}(x^k, X^*)^{(2+\delta)/2}$ *for all* $x^{k+1}, x^k \in N(x^*, b/2)$.

Let a positive constant r_1 be

$$r_1 := \min\left\{\frac{b}{2(1+11c_2)}, \frac{c_3^{-\frac{2}{\delta}}}{2}\right\}.$$

Lemma 6 ([12, Theorem 2.1]). *Suppose that Assumptions* 3 *and* 4 *hold. If* $x^0 \in N(x^*, r_1)$, *then* $x^k \in N(x^*, b/2)$ *for all nonnegative integer* k.

Theorem 2 ([12, Theorem 2.1]). *Let* $\{x^k\}$ *be a sequence generated by Algorithm* 2. *Suppose that Assumptions* 3 *and* 4 *hold. If* $x^0 \in N(x^*, r_1)$, *then the sequence* $\{x^k\}$ *converges to some solution* $\bar{x} \in X^*$ *superlinearly.*

4.2. Quadratic Convergence

We prove that Algorithm 2 converges quadratically. To this end, we make the following assumption.

Assumption 5. *There exists a positive constant* γ *such that*

$$\mu_k \|\eta^k\| \leq \gamma \|x^k - x^*\|^2$$

for all $\eta^k \in \partial\psi(d^k)$.

Remark 2. *Suppose that Assumptions* 3 *and* 4 *hold, and* $x^k \in N(x^*, b/2)$. *In the case of the common LMM, specifically,* $\psi(d) = \|d\|^2$, *we have* $\partial\psi(d^k) = \{2d^k\}$. *Moreover, from* (6), *we have* $p = 2$. *It then follows from Assumption* 4, *Lemmas* 3, *and* 4 *that* $\mu_k\|\eta^k\| \leq 2\xi_2\|F(x^k)\|\|d^k\| \leq 2\xi_2 L_2 c_2 \|x^k - x^*\|^2$, *and hence Assmption* 5 *holds. When the* L_1 *norm is adopted as* ψ, *we obtain* $\|\eta^k\| \leq \sqrt{n}$ *and* $p = 1$. *Therefore, from Assumption* 4 *and Lemma* 3, *we have* $\mu_k\|\eta^k\| \leq \sqrt{n}\xi_2\|F(x^k)\|^2 \leq \sqrt{n}\xi_2 L_2\|x^k - x^*\|^2$.

In the subsequent convergence analysis, we also utilize the way of Fan and Yuan [12] to show quadratic convergence of the proposed method. They proved the quadratic convergence of the normal LMM by using the fact that the search direction can be written explicitly. However, the search direction of the proposed method can not be expressed explicitly due to the generalization of the regularization term. Therefore, we provide some new lemmas which evaluate the difference between the search direction of the exisiting LMM and of the proposed method so that we can use the same approach as [12]. For this purpose, we define the search direction of the ordinary LMM as follows. Let d^k_{LM} be a solution of the following traditional LMM subproblem:

$$\underset{d \in \mathbb{R}^n}{\text{minimize}} \quad \|F'(x^k)d + F(x^k)\|^2 + \nu_k\|d\|^2, \tag{16}$$

where ν_k is a positive parameter. In the following, as with [12], ν_k is defined as below:

$$\nu_k = \|F(x^k)\|^\delta, \quad \delta \in [1, 2]. \tag{17}$$

First, we provide some lemmas related to the search direction of the ordinary LMM.

Lemma 7 ([12, Lemma 2.1])**.** *Suppose that Assumption* 3 *holds. Then there exists* $c_4 > 0$ *such that* $\|d^k_{\mathrm{LM}}\| \leq c_4 \mathrm{dist}(x^k, X^*)$ *for all* $x^k \in N(x^*, b/2)$.

Lemma 8. *Suppose that Assumptions* 3 *and* 5 *hold. Then there exists* $c_5 > 0$ *such that* $\|F'(x^k)^\top F'(x^k)(d^k - d^k_{\mathrm{LM}})\| \leq c_5\|x^k - x^*\|^2$ *for all* $x^k \in N(x^*, b/2)$.

Proof. From Lemma 3 and (17), we have

$$\nu_k = \|F(x^k)\|^\delta \leq L_2^\delta \|x^k - x^*\|^\delta \leq L_2^\delta \|x^k - x^*\|. \tag{18}$$

Since d^k and d^k_{LM} are respectively solutions of (5) and (16), we obtain $2F'(x^k)^\top F'(x^k)d^k_{\mathrm{LM}}+2F'(x^k)^\top F(x^k)+2\nu_k d^k_{\mathrm{LM}}=0$ and $2F'(x^k)^\top F'(x^k)d^k+2F'(x^k)^\top F(x^k)+\mu_k\eta^k=0$, where $\eta^k\in\partial\psi(d^k)$. By subtracting these two equations, we have $2F'(x^k)^\top F'(x^k)(d^k-d^k_{\mathrm{LM}})+\mu_k\eta^k-2\nu_k d^k_{\mathrm{LM}}=0$. It then follows from Assumption 5, Lemma 7, and (18) that

$$\begin{aligned}\|F'(x^k)^\top F'(x^k)(d^k-d^k_{\mathrm{LM}})\| &\le \frac{1}{2}\left(\mu_k\|\eta^k\|+2\nu_k\|d^k_{\mathrm{LM}}\|\right)\\ &\le \frac{\gamma+2c_4L_2^\delta}{2}\|x^k-x^*\|^2.\end{aligned}$$

Therefore, the desired inequality holds. □

As seen in the convergence analysis of [12], we will utilize the singular value decomposition (SVD) of $F'(x^*)$ as follows.

$$F'(x^*)=U^*\Sigma^*{V^*}^\top=U^*\mathrm{diag}\left(\Sigma_1^*,O\right){V^*}^\top=U_1^*\Sigma_1^*{V_1^*}^\top,$$

where $\Sigma_1^*:=\mathrm{diag}\left(\sigma_1^*,\dots,\sigma_r^*\right)$, $\sigma_1^*\ge\cdots\ge\sigma_r^*>0$ and $\mathrm{rank}\left(\Sigma_1^*\right)=r$. Let $\{x^k\}$ be a sequence converging to x^*, and $\sigma_1^{(k)}\ge\cdots\ge\sigma_{\min\{m,n\}}^{(k)}\ge 0$ be singular values of $F'(x^k)$. Note that $F'(x^k)\to F'(x^*)$ as $k\to\infty$. Then, since $\sigma_i^{(k)}\to\sigma_i^*$ $(k\to\infty)$ for all $i\in\{1,\dots,r\}$ and $\sigma_i^{(k)}\to 0$ $(k\to\infty)$ for all $i\in\{r+1,\dots,\min\{m,n\}\}$ from [16, Theorem 2.6.4], the number of positive singular values of $F'(x^k)$ is r or more for all sufficiently large k, i.e., the SVD of $F'(x^k)$ can be represented as follows:

$$\begin{aligned}F'(x^k)&=U_k\Sigma_k{V_k}^\top\\ &=[U_{k,1}\ U_{k,2}\ U_{k,3}]\,\mathrm{diag}(\Sigma_{k,1},\Sigma_{k,2},O)\begin{bmatrix}V_{k,1}^\top\\ V_{k,2}^\top\\ V_{k,3}^\top\end{bmatrix}\\ &=U_{k,1}\Sigma_{k,1}V_{k,1}^\top+U_{k,2}\Sigma_{k,2}V_{k,2}^\top,\end{aligned}\tag{19}$$

where $\Sigma_{k,1}:=\mathrm{diag}(\sigma_1^{(k)},\dots,\sigma_r^{(k)})>0$, $\Sigma_{k,2}:=\mathrm{diag}(\sigma_{r+1}^{(k)},\dots,\sigma_{r+q}^{(k)})\ge 0$, $\mathrm{rank}\left(\Sigma_{k,1}\right)=r$, and $\mathrm{rank}\left(\Sigma_{k,2}\right)=q\ge 0$. Let $\bar r:=\min\{b/2,r_2,r_3\}$, where r_2 is a radius such that SVD (19) exists for $x^k\in N(x^*,r_2)$, and r_3 is a positive constant satisfying $\sigma_r^*-L_1r_3>0$.

By utilizing SVD (19), some additional lemmas can be shown.

Lemma 9 ([12, Inequality (2.13) in Theorem 2.2]). *Suppose that Assumption 3 holds. Then there exists $c_6>0$ such that $\|F'(x^k)d^k_{\mathrm{LM}}+F(x^k)\|\le c_6\|x^k-x^*\|^2$ for all $x^k\in N(x^*,\bar r)$.*

Lemma 10. *Suppose that Assumption* 3 *holds. Then*

(i) $\|\Sigma_{k,2}\| \le L_1 \|x^k - x^*\|$ *for all* $x^k \in N(x^*, \bar{r})$*;*

(ii) *there exists* $c_7 > 0$ *such that* $c_7 \le (\Sigma_{k,1})_{ii}$ *for all* $i \in \{1, \ldots, r\}$ *and all* $x^k \in N(x^*, \bar{r})$*.*

Proof. From [17, Theorem 4.11 (Mirsky)] and (1) of Assumption 3, we have

$$\|\mathrm{diag}\,(\Sigma_{k,1} - \Sigma_1^*, \Sigma_{k,2}, O)\,\| \le \|F'(x^k) - F'(x^*)\| \le L_1 \|x^k - x^*\|,$$

which leads to $\|\Sigma_{k,1} - \Sigma_1^*\| \le L_1 \|x^k - x^*\|$ and $\|\Sigma_{k,2}\| \le L_1 \|x^k - x^*\|$. Then, we obtain

$$\sigma_r^* - \sigma_i^{(k)} \le \sigma_i^* - \sigma_i^{(k)} \le \max_{1 \le i \le r} |\sigma_i^{(k)} - \sigma_i^*| = \|\Sigma_{k,1} - \Sigma_1^*\| \le L_1 r_3$$

for all $i \in \{1, \ldots, r\}$. Therefore, we have $\sigma_i^{(k)} \ge \sigma_r^* - L_1 r_3 > 0$ for all $i \in \{1, \ldots, r\}$. □

Lemma 11. *Suppose that Assumptions* 3*,* 4*, and* 5 *hold. Then there exist* $c_8 > 0$ *and* $c_9 > 0$ *such that*

$$\|\Sigma_{k,1} V_{k,1}^\top (d^k - d_{\mathrm{LM}}^k)\| \le c_8 \|x^k - x^*\|^2, \ \|\Sigma_{k,2} V_{k,2}^\top (d^k - d_{\mathrm{LM}}^k)\| \le c_9 \|x^k - x^*\|^2$$

for all $x^k \in N(x^*, \bar{r})$*.*

Proof. From Lemma 8, we have

$$\begin{aligned} c_5 \|x^k - x^*\|^2 &\ge \|F'(x^k)^\top F'(x^k)(d^k - d_{\mathrm{LM}}^k)\| \\ &= \|(V_{k,1}\Sigma_{k,1}^2 V_{k,1}^\top + V_{k,2}\Sigma_{k,2}^2 V_{k,2}^\top)(d^k - d_{\mathrm{LM}}^k)\| \\ &\ge \|V_{k,1}\Sigma_{k,1}^2 V_{k,1}^\top (d^k - d_{\mathrm{LM}}^k)\| - \|V_{k,2}\Sigma_{k,2}^2 V_{k,2}^\top (d^k - d_{\mathrm{LM}}^k)\|. \end{aligned} \tag{20}$$

Lemmas 4, 7, and 10 imply

$$\begin{aligned} \|V_{k,2}\Sigma_{k,2}^2 V_{k,2}^\top (d^k - d_{\mathrm{LM}}^k)\| &\le \|V_{k,2}\| \|\Sigma_{k,2}\|^2 \|V_{k,2}^\top\| \|d^k - d_{\mathrm{LM}}^k\| \\ &\le \|\Sigma_{k,2}\|^2 \left(\|d^k\| + \|d_{\mathrm{LM}}^k\|\right) \\ &\le L_1^2 (c_2 + c_4) \|x^k - x^*\|^2. \end{aligned} \tag{21}$$

Combining (20) and (21) yields

$$\begin{aligned}\|V_{k,1}\Sigma_{k,1}^2 V_{k,1}^\top (d^k - d^k_{\mathrm{LM}})\| &\le c_5\|x^k - x^*\|^2 + \|V_{k,2}\Sigma_{k,2}^2 V_{k,2}^\top (d^k - d^k_{\mathrm{LM}})\| \\ &= \left\{L_1^2(c_2+c_4)+c_5\right\}\|x^k - x^*\|^2.\end{aligned}$$

It then follows from Lemma 10 that

$$\begin{aligned}\|\Sigma_{k,1}V_{k,1}^\top(d^k - d^k_{\mathrm{LM}})\| &= \|\Sigma_{k,1}^{-1}V_{k,1}^\top V_{k,1}\Sigma_{k,1}^2 V_{k,1}^\top(d^k - d^k_{\mathrm{LM}})\| \\ &\le \|\Sigma_{k,1}^{-1}\|\|V_{k,1}^\top\|\|V_{k,1}\Sigma_{k,1}^2V_{k,1}^\top(d^k - d^k_{\mathrm{LM}})\| \\ &\le \frac{L_1^2(c_2+c_4)+c_5}{c_7}\|x^k - x^*\|^2, \\ \|\Sigma_{k,2}V_{k,2}^\top(d^k - d^k_{\mathrm{LM}})\| &\le \|\Sigma_{k,2}\|\|V_{k,2}^\top\|\|d^k - d^k_{\mathrm{LM}}\| \\ &= \|\Sigma_{k,2}\|\|d^k - d^k_{\mathrm{LM}}\| \\ &\le L_1(c_2+c_4)\|x^k - x^*\|^2.\end{aligned}$$

Therefore, the desired inequalities hold. □

Lemma 12. *Suppose that Assumptions 3, 4, and 5 hold. Then there exists a positive constant c_{10} such that $\|F(x^{k+1})\| \le c_{10}\|x^k - x^*\|^2$ for all $x^{k+1} \in N(x^*, b/2)$ and $x^k \in N(x^*, \bar{r})$.*

Proof. Using Lemma 11 yields

$$\begin{aligned}\|F'(x^k)(d^k - d^k_{\mathrm{LM}})\|^2 &= \|\Sigma_{k,1}V_{k,1}^\top(d^k - d^k_{\mathrm{LM}})\|^2 + \|\Sigma_{k,2}V_{k,2}^\top(d^k - d^k_{\mathrm{LM}})\|^2 \\ &\le \left(c_8^2 + c_9^2\right)\|x^k - x^*\|^4,\end{aligned}$$

and hence

$$\|F'(x^k)(d^k - d^k_{\mathrm{LM}})\| \le \sqrt{c_8^2 + c_9^2}\|x^k - x^*\|^2. \tag{22}$$

Since $\|F(x^{k+1})\| - \|F'(x^k)d^k + F(x^k)\| \le \|F'(x^k)(x^{k+1} - x^k) - (F(x^{k+1}) - F(x^k))\| \le \frac{L_1}{2}\|d^k\|^2$, inequality (22) and Lemmas 4 and 9 imply

$$\begin{aligned}\|F(x^{k+1})\| &\le \|F'(x^k)d^k + F(x^k)\| + \frac{L_1}{2}\|d^k\|^2 \\ &\le \|F'(x^k)d^k_{\mathrm{LM}} + F(x^k)\| + \|F'(x^k)(d^k - d^k_{\mathrm{LM}})\| + \frac{L_1}{2}\|d^k\|^2 \\ &= \left(\frac{L_1c_2^2}{2} + c_6 + \sqrt{c_8^2 + c_9^2}\right)\|x^k - x^*\|^2.\end{aligned}$$

Therefore, the assertion is proven. □

Theorem 3. *Let $\{x^k\}$ be a sequence generated by Algorithm 2. Suppose that Assumptions 3, 4, and 5 hold. If $x^0 \in N(x^*, r_1)$, then the sequence $\{x^k\}$ converges to the solution x^* quadratically.*

Proof. Let a positive constant $\tilde{r}$ be

$$\tilde{r} := \min\left\{\left(\frac{1}{2c_3}\right)^{\frac{2}{\delta}}, \bar{r}\right\}.$$

From Theorem 2 and $x^0 \in N(x^*, r_1)$, we have $x^k \to x^*$ $(k \to \infty)$. Then we obtain $\|d^k\| \leq \|x^{k+1} - x^*\| + \|x^k - x^*\| \to 0$ $(k \to \infty)$. Therefore, there exists a constant $n_1 \in \mathbb{N}$ such that $x^k \in N(x^*, \tilde{r})$ and

$$\|d^k\| \leq \min\left\{\left(\frac{1}{3c}\right)^{\frac{2}{\delta}}, \frac{1}{3c'}\right\} \tag{23}$$

for all $k \geq n_1$, where $c = 2^{\frac{2+\delta}{2}} c_2 c_3$ and $c' = 9c_2 c_{10}/(4c_1)$. Let $\ell \geq n_1$ be an arbitrary. Since $x^\ell \in N(x^*, b/2)$ and $x^\ell + d^\ell = x^{\ell+1} \in N(x^*, b/2)$, Lemma 5 ensures $\mathrm{dist}(x^\ell, X^*) \leq \mathrm{dist}(x^\ell + d^\ell, X^*) + \|d^\ell\| \leq c_3 \mathrm{dist}(x^\ell, X^*)^{(2+\delta)/2} + \|d^\ell\|$, which implies that $\left\{1 - c_3 \mathrm{dist}(x^\ell, X^*)^{\delta/2}\right\} \mathrm{dist}(x^\ell, X^*) \leq \|d^\ell\|$. Since we have $c_3 \mathrm{dist}(x^\ell, X^*)^{\delta/2} \leq c_3 \|x^\ell - x^*\|^{\delta/2} \leq 1/2$, we obtain $\mathrm{dist}(x^\ell, X^*) \leq 2\|d^\ell\|$. Therefore, from Lemmas 4 and 5, we have

$$\|d^{\ell+1}\| \leq c_2 \mathrm{dist}(x^{\ell+1}, X^*) \leq c_2 c_3 \mathrm{dist}(x^\ell, X^*)^{\frac{2+\delta}{2}} \leq 2^{\frac{2+\delta}{2}} c_2 c_3 \|d^\ell\|^{\frac{2+\delta}{2}}. \tag{24}$$

Suppose that $m > \ell$. Then from (23) and (24), we have

$$\begin{aligned}
\|d^j\| &\leq c\|d^{j-1}\|^{\frac{2+\delta}{2}} \leq \frac{1}{3}\|d^{j-1}\| \leq \frac{c}{3}\|d^{j-2}\|^{\frac{2+\delta}{2}} \\
&\leq \left(\frac{1}{3}\right)^2 \|d^{j-2}\| \leq \cdots \leq \left(\frac{1}{3}\right)^{j-\ell} \|d^\ell\|
\end{aligned}$$

for all $j \in \{\ell + 1, \ldots, m - 1\}$. Hence we obtain

$$\begin{aligned}
\|x^m - x^\ell\| &= \left\|\sum_{j=\ell}^{m-1} d^j\right\| \leq \|d^\ell\| + \sum_{j=\ell+1}^{m-1} \|d^j\| \\
&\leq \left\{1 + \sum_{l=1}^{m-\ell-1} \left(\frac{1}{3}\right)^l\right\} \|d^\ell\| = \frac{3 - 3^{-m+\ell+1}}{2}\|d^\ell\|,
\end{aligned}$$

which implies from $m \to \infty$ that $\|x^\ell - x^*\| \le (3/2)\|d^\ell\|$. Therefore, it follows from (2) of Assumption 3, Lemmas 4, and 12 that

$$\begin{aligned}\|d^{\ell+1}\| &\le c_2 \mathrm{dist}(x^{\ell+1}, X^*) \le \frac{c_2}{c_1}\|F(x^{\ell+1})\| \\ &\le \frac{c_2 c_{10}}{c_1}\|x^\ell - x^*\|^2 \le \frac{9 c_2 c_{10}}{4 c_1}\|d^\ell\|^2. \qquad (25)\end{aligned}$$

Then, from (23) and (25), we have

$$\begin{aligned}\|d^j\| &\le c'\|d^{j-1}\|^2 \le \frac{1}{3}\|d^{j-1}\| \le \frac{c'}{3}\|d^{j-2}\|^2 \\ &\le \left(\frac{1}{3}\right)^2 \|d^{j-2}\| \le \cdots \le \left(\frac{1}{3}\right)^{j-\ell} \|d^\ell\|\end{aligned}$$

for all $j \in \{\ell+1, \ldots, m-1\}$. Therefore, we obtain

$$\begin{aligned}\|x^m - x^\ell\| &= \left\|\sum_{j=\ell}^{m-1} d^j\right\| \ge \|d^\ell\| - \sum_{j=\ell+1}^{m-1}\|d^j\| \\ &\ge \left\{1 - \sum_{l=1}^{m-\ell-1}\left(\frac{1}{3}\right)^l\right\}\|d^\ell\| = \frac{1 + 3^{-m+\ell+1}}{2}\|d^\ell\|. \qquad (26)\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\|d^j\| &\le c'\|d^{j-1}\|^2 \le \frac{1}{3}\|d^{j-1}\| \le \frac{c'}{3}\|d^{j-2}\|^2 \\ &\le \left(\frac{1}{3}\right)^2 \|d^{j-2}\| \le \cdots \le \left(\frac{1}{3}\right)^{j-\ell-1} \|d^{\ell+1}\|\end{aligned}$$

for all $j \in \{\ell+2, \ldots, m-1\}$. It then follows that

$$\begin{aligned}\|x^m - x^{\ell+1}\| &= \left\|\sum_{j=\ell+1}^{m-1} d^j\right\| \le \|d^{\ell+1}\| + \sum_{j=\ell+2}^{m-1}\|d^j\| \\ &\le \left\{1 + \sum_{l=1}^{m-\ell-2}\left(\frac{1}{3}\right)^l\right\}\|d^{\ell+1}\| = \frac{3(1 - 3^{-m+\ell+1})}{2}\|d^{\ell+1}\|. \qquad (27)\end{aligned}$$

Hence, inequalities (25), (26), and (27) yield

$$\|x^{\ell+1} - x^m\| \leq \frac{3c'(1 - 3^{-m+\ell+1})}{2}\|d^\ell\|^2 \leq \frac{6c'(1 - 3^{-m+\ell+1})}{(1 + 3^{-m+\ell+1})^2}\|x^\ell - x^m\|^2.$$

Let $\bar{c} := 6c' > 0$ and $m \to \infty$ leads to $\|x^{\ell+1} - x^*\| \leq \bar{c}\|x^\ell - x^*\|^2$, which implies that the sequence $\{x^k\}$ converges to the solution x^* quadratically. $\square$

5. Local Convergence of Algorithm 1

In this section, we prove local and quadratic convergence of Algorithm 1 which converges globally under Assumptions 1 and 2. Although we consider applying Theorem 3 to the current convergence analysis, there are some differences between Algorithms 1 and 2 as follows:

- Algorithm 1 set the step size as $\beta^{m_k} \in (0, 1]$ whereas Algorithm 2 always adopt 1.
- Algorithm 1 supposes that μ_k satisfies Assumption 2. Meanwhile, Algorithm 2 needs to satisfy Assumption 4.

Therefore, we provide sufficient conditions under which the step size becomes 1 and μ_k given in Assumption 2 satisfies Assumption 4.

Throughout this section, we use the following notation:

$$u := \sup\left\{\sigma_1\left(F'(x)\right) \mid x \in N\left(x^*, b\right)\right\},$$

where $\sigma_i(M) \colon \mathbb{R}^{m\times n} \to \mathbb{R}$ is a function that returns the i-th largest singular value of M. Then, the continuity of σ_i and F' implies $u < \infty$, and the definition of u ensures

$$(\Sigma_k)_{ii} = \sigma_i\left(F'(x^k)\right) \leq \sigma_1\left(F'(x^k)\right) \leq u \tag{28}$$

for all $x^k \in N(x^*, b)$ and $i \in \{1, \ldots, \min\{m, n\}\}$. By using the positive constant u, we also define the following positive constant ω:

$$\omega = \frac{2c_2^2u^2}{c_1^2} + \frac{b\bar{\gamma}c_2}{2}.$$

Now we make the following three assumptions.

Assumption 6. *When ψ is not continuously differentiable at $d = 0$, the following two conditions hold.*

(1) $2 - p + \delta \geq 1$.

(2) *For the merit function ϕ, Polyak-Lojasiewicz inequality holds at a neighborhood of x^*, i.e., there exists a positive constant μ such that*

$$\frac{1}{2}\|\nabla \phi(x)\|^2 \geq \mu\left(\phi(x) - \phi(x^*)\right)$$

for all $x \in N(x^*, b/2)$.

Assumption 7. *There exists a positive constant $\bar{\gamma}$ such that*

$$\mu_k \|\eta^k\| \leq \bar{\gamma}\|F(x^k)\|^2$$

for all $\eta^k \in \partial\psi(d^k)$.

Assumption 8. *The parameter α in Algorithm 1 is chosen to satisfy the following inequality:*

$$1 - \alpha\omega > 0.$$

Remark 3. *The Polyak-Lojasiewicz inequality, that is, (2) of Assumption 6 is satisfied if ϕ is strongly convex. Moreover, it is known that there exist several sufficient conditions, which are milder than strong convexity, for the Polyak-Lojasiewicz inequality [18].*

Remark 4. *As with Remark 2, we show that Assumption 7 is satisfied in the case of $\psi(d) = \|d\|^2$ and of $\psi(d) = \|d\|_1$. Suppose that Assumptions 3 and 4 hold, and $x^k \in N(x^*, b/2)$. When $\psi(d) = \|d\|^2$, from Assumptions 3, 4, and Lemma 4, we have $\mu_k\|\eta^k\| \leq 2\xi_2\|F(x^k)\|\|d^k\| \leq 2\xi_2 c_2\|F(x^k)\| \cdot \mathrm{dist}(x^k, X^*) \leq (2\xi_2 c_2/c_1)\|F(x^k)\|^2$. On the other hand, when $\psi(d) = \|d\|_1$, we obtain $\mu_k\|\eta^k\| \leq \sqrt{n}\xi_2\|F(x^k)\|^2$.*

Remark 5. *Suppose that Assumption 3 holds. If Assumption 7 holds and $x^k \in N(x^*, b)$, then from Lemma 3, we have $\mu_k\|\eta^k\| \leq \bar{\gamma}\|F(x^k)\|^2 \leq \bar{\gamma}{L_2}^2\|x^k - x^*\|^2$, which implies that Assumption 5 holds.*

The next lemma shows that μ_k given in Assumption 2 satisfies Assumption 4 at a neighborhood of x^*.

Lemma 13. *Suppose that Assumptions 2 and 6 hold. If $x^k \in N(x^*, b/2)$ and $\|F(x^k)\| \leq 1$, then the following inequalities hold:*

$$\min\left\{1, \frac{\sqrt{\mu}}{M}\right\}\|F(x^k)\|^{2-p+\delta} \leq \mu_k \leq \|F(x^k)\|^{2-p+\delta}.$$

Proof. When ψ is continuously differentiable at $d = 0$, (1) of Assumption 2 implies that the desired inequality holds. In the following, we consider the case where ψ is not continuously differentiable at $d = 0$. From (2) of Assumption 6 and (9), we have $\sqrt{\mu}\|F(x)\| \leq \|F'(x)^\top F(x)\|$ for all $x \in N(x^*, b)$. It then follows from (1) of Assumption 6 that $(\sqrt{\mu}/M)\|F(x^k)\|^{2-p+\delta} \leq (\sqrt{\mu}/M)\|F(x^k)\| \leq \|F'(x^k)^\top F(x^k)\|/M$. Therefore, we obtain

$$\begin{aligned}\min\left\{1, \frac{\sqrt{\mu}}{M}\right\}\|F(x^k)\|^{2-p+\delta} &\leq \min\left\{\|F(x^k)\|^{2-p+\delta}, \frac{\|F'(x^k)^\top F(x^k)\|}{M}\right\} \\ &= \mu_k \leq \|F(x^k)\|^{2-p+\delta}.\end{aligned}$$

This completes the proof. $\square$

The next lemma shows that 1 is chosen as the step size.

Lemma 14. *Suppose that Assumptions* 2, 3, 6, 7, *and* 8 *hold. If* $x^k + d^k, x^k \in N(x^*, b/2)$ *and*

$$\|F(x^k)\| \leq \min\left\{\left(\frac{c_1^{\frac{2+\delta}{2}}\sqrt{1-\alpha\omega}}{L_2 c_3}\right)^{\frac{2}{\delta}}, 1\right\},$$

then the step size β^{m_k} *becomes* 1.

Proof. Note that from Lemma 13, Assumption 4 holds. Since d^k is a solution of (5), there exists $\eta^k \in \partial\psi(d^k)$ such that $2F'(x^k)^\top F'(x^k)d^k + 2F'(x^k)^\top F(x^k) + \mu_k\eta^k = 0$. Therefore, we have $\nabla\phi(x^k) = F'(x^k)^\top F(x^k) = -F'(x^k)^\top F'(x^k)d^k - \mu_k\eta^k/2$, which implies that

$$\nabla\phi(x^k)^\top d^k = -\|F'(x^k)d^k\|^2 - \frac{1}{2}\mu_k\left(\eta^k\right)^\top d^k. \tag{29}$$

From (2) of Assumption 3, Lemma 4, and (28), we obtain

$$\begin{aligned}-\|F'(x^k)d^k\|^2 &= -\|U_k\Sigma_k V_k^\top d^k\|^2 \geq -\|U_k\|^2\|\Sigma_k\|^2\|V_k^\top\|^2\|d^k\|^2 \\ &\geq -u^2 \cdot c_2^2 \mathrm{dist}(x^k, X^*)^2 \geq -\frac{c_2^2 u^2}{c_1^2}\|F(x^k)\|^2.\end{aligned} \tag{30}$$

On the other hand, from (2) of Assumption 3, Assumption 7, and Lemma 4, we have

$$
\begin{aligned}
-\frac{1}{2}\mu_k\left(\eta^k\right)^\top d^k &\geq -\frac{1}{2}\mu_k\|\eta^k\|\|d^k\| \\
&\geq -\frac{\bar{\gamma}c_2}{2}\|F(x^k)\|^2 \cdot \operatorname{dist}(x^k, X^*) \\
&\geq -\frac{b\bar{\gamma}c_2}{4}\|F(x^k)\|^2. \qquad (31)
\end{aligned}
$$

Hence, using (29), (30), and (31) yields

$$
\begin{aligned}
\nabla\phi(x^k)^\top d^k &= -\|F'(x^k)d^k\|^2 - \frac{1}{2}\mu_k\left(\eta^k\right)^\top d^k \\
&\geq -\frac{c_2^2u^2}{c_1^2}\|F(x^k)\|^2 - \frac{b\bar{\gamma}c_2}{4}\|F(x^k)\|^2 \\
&= -\frac{\omega}{2}\|F(x^k)\|^2.
\end{aligned}
$$

Let $x_p^k \in X^*$ denotes the vector such that $\|x_p^k - x^k\| = \operatorname{dist}(x^k, X^*)$. Since $\|x_p^k - x^*\| \leq \|x_p^k - x^k\| + \|x^k - x^*\| \leq b$, (2) of Assumption 3 and Lemmas 3 and 5 derive

$$
\begin{aligned}
\|F(x^k + d^k)\| &\leq L_2\|x^k + d^k - x_p^{k+1}\| \\
&\leq L_2 c_3 \operatorname{dist}(x^k, X^*)^{\frac{2+\delta}{2}} \\
&\leq L_2 c_1^{-\frac{2+\delta}{2}} c_3\|F(x^k)\|^{\frac{2+\delta}{2}} \\
&= L_2 c_1^{-\frac{2+\delta}{2}} c_3\|F(x^k)\|^{\frac{\delta}{2}} \cdot \|F(x^k)\| \\
&\leq \sqrt{1-\alpha\omega}\|F(x^k)\|.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
\phi(x^k + d^k) - \phi(x^k) &= \frac{1}{2}\left(\|F(x^{k+1})\|^2 - \|F(x^k)\|^2\right) \\
&\leq \frac{1}{2}\left\{(1-\alpha\omega)\|F(x^k)\|^2 - \|F(x^k)\|^2\right\} \\
&= -\frac{\alpha\omega}{2}\|F(x^k)\|^2 \\
&\leq \alpha\nabla\phi(x^k)^\top d^k,
\end{aligned}
$$

which leads to $\beta^{m_k} = 1$. □

Theorem 4. *Let $\{x^k\}$ be a bounded sequence generated by Algorithm 1.*

Suppose that either (1) *or* (2) *of the Assumption* 1 *holds and that Assumption* 2 *holds. Then, any accumulation point of* $\{x^k\}$ *satisfies the stationary condition of* $\min\{\phi(x) \mid x \in \mathbb{R}^n\}$*. Moreover, if the accumulation point* x^* *solves* (1) *and Assumptions* 3, 6, 7, *and* 8 *hold at* x^**, then the sequence* $\{x^k\}$ *converges to* x^* *quadratically.*

Proof. From Theorem 1, we have already proved the former part of this theorem. In the following, we show the latter part. Note that Assumption 5 holds from Remark 5. By Theorem 3, it suffices to show that there exists a positive integer $\tilde{k}$ such that μ_k satisfies Assumption 4 and $\beta^{m_k} = 1$ for all $k \geq \tilde{k}$.

Since $x^k \to x^*$ $(k \to \infty)$ and $F(x^k) \to F(x^*) = 0$ $(k \to \infty)$, there exists a constant $\tilde{k}$ such that $x^l \in N(x^*, r_1)$ and

$$\|F(x^l)\| \leq \min\left\{ \left(\frac{c_1^{\frac{2+\delta}{2}} \sqrt{1-\alpha\omega}}{L_2 c_3} \right)^{\frac{2}{\delta}}, 1 \right\}$$

for all $l \geq \tilde{k}$. Suppose that $k \geq \tilde{k}$. From Lemma 13, Assumption 4 holds for μ_k. It then follows from Lemma 4 that $\|x^k + d^k - x^*\| \leq \|x^k - x^*\| + c_2 \mathrm{dist}(x^k, X^*) \leq (1 + c_2) r_1 \leq b/2$, which implies $x^k + d^k \in N(x^*, b/2)$. Therefore, Lemma 14 guarantees $\beta^{m_k} = 1$. Hence, Theorem 3 implies that $\{x^k\}$ converges to x^* quadratically. $\square$

6. Numerical Experiments

In this section, we conduct numerical experiments to compare Algorithm 1 with the existing LMM [12, Algorithm 3.1]. All experiments were performed using MATLAB R2021b on a machine with Intel Core i5-5350U 1.80GHz of CPU and 8GB of RAM. In both Algorithm 1 and the ordinary LMM, we chose $\alpha = 0.0001$ and $\beta = 0.5$ for line search, and used $\|F'(x^k)^\top F(x^k)\| < \varepsilon$ for termination criteria, where $\varepsilon = n \cdot 10^{-6}$. The regularization parameter was set to $\min\{\mu_k, 0.1\}$ for preventing it getting too large. Algorithm 1 adopted the L_1 norm as the regularization term ψ of (5) and solved the subproblem by using dual methods [19].

We consider a continuously differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}^m$ as follows:

$$F(x) \coloneqq (F_1(x),\ \ldots,\ F_m(x))^\top, \quad F_i(x) \coloneqq \frac{1}{2} x^\top A_i x + b_i^\top x + c_i,$$

where $F_i \colon \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function and $A_i \in \mathbb{R}^{n\times n}$ is a symmetric matrix. We generated constants of the test problems as below.

(1) We generate $B_i \in \mathbb{R}^{n\times n}, b_i \in \mathbb{R}^n$ and $x^* \in \mathbb{R}^n$ whose elements are generated randomly from the interval $[0,1]$, and
(2) set $A_i := (B_i + B_i{}^\top)/2$ and $c_i := -(x^*)^\top A_i x^*/2 + b_i^\top x^*$

We fix $m = 10$, and change n to $n = 10$, 100, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000. For each n, we solved 10 test problems by using the different initial points $x^0 \in \mathbb{R}^n$ whose elements are generated randomly from the interval $[0,1]$.

Table 1: Comparison of average CPU time between the LMM and Algorithm 1 [sec.]

n	LMM (CD)	LMM (SMW)	proposed method
10	0.0236	0.0143	0.1109
100	0.0255	0.0037	0.0373
500	0.0430	0.0238	0.0302
1000	0.1360	0.0527	0.0337
1500	0.3706	0.1118	0.0533
2000	0.8421	0.1915	0.0647
2500	1.1657	0.2570	0.0745
3000	1.5159	0.3365	0.0800
3500	2.2266	0.4512	0.1035
4000	3.0280	0.5722	0.1193
4500	4.6363	0.7494	0.1682
5000	5.9642	0.9137	0.1920

Table 1 and Figure 1 indicate the average CPU time of the existing LMM and Algorithm 1. In particular, LMM (CD) and LMM (SMW) mean that they respectively solve (4) by using the Cholesky decomposition (CD) and the Sherman-Morrison-Woodbury (SMW) formula. The table and figure imply that the proposed method was superior to the other methods when n is getting larger while m is fixed.

7. Conclusions

In this paper, we proposed a new LMM with generalized regularization terms. We showed not only the global but also the local quadratic convergence of the proposed method. Moreover, we conducted numerical experiments to verify the efficiency of the proposed method.

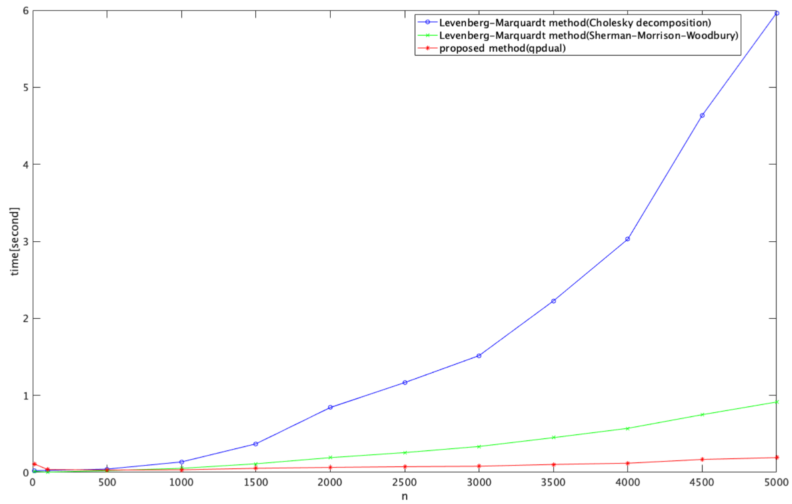

Figure 1: Comparison of average CPU time between the LMM and Algorithm 1

Future works will be to consider better regularization terms that satisfy condition (6), to povide inexact criteria which do not affect the convergence property for the case where it is expensive to solve the subproblems exactly, and to develop a regularized Newton method with generalized regularization terms for convex optimization.

Declarations

Conflict of interest The authors declare no competing interests.

References

[1] J. M. Ortega, W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, SIAM, 2000.

[2] M. Yano, J. D. Penn, G. Konidaris, A. T. Patera, Math, Numerics, & Programming (for Mechanical Engineers) (2012).

[3] K. L. Judd, Numerical Methods in Economics, MIT Press, 1998.

[4] J. Nocedal, S. J. Wright, Numerical Optimization, Springer, 1999.

[5] J. E. Dennis, R. B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Society for Industrial and Applied Mathematics, 1996. doi:10.1137/1.9781611971200.

[6] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, 1996. doi:10.1137/1.9781611971484.

[7] W. Sun, Y.-X. Yuan, Optimization Theory and Methods: Nonlinear Programming, Vol. 1, Springer, 2006.

[8] R. Fletcher, Practical Methods of Optimization, John Wiley & Sons, 2013.

[9] K. Levenberg, A Method for the Solution of Certain Non-Linear Problems in Least Squares, Quarterly of Applied Mathematics 2 (2) (1944) 164–168.

[10] D. W. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics 11 (2) (1963) 431–441.

[11] N. Yamashita, M. Fukushima, On the Rate of Convergence of the Levenberg-Marquardt Method, in: G. Alefeld, X. Chen (Eds.), Topics in Numerical Analysis, Springer, Vienna, 2001, pp. 239–249.

[12] J.-Y. Fan, Y.-X. Yuan, On the Quadratic Convergence of the Levenberg-Marquardt Method without Nonsingularity Assumption, Computing 74 (2005) 23–39. doi:10.1007/s00607-004-0083-1.

[13] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288.

[14] Y. Nesterov, B. T. Polyak, Cubic regularization of Newton method and its global performance, Mathematical Programming 108 (1) (2006) 177–205.

[15] A. E. Hoerl, R. W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics 12 (1) (1970) 55–67.

[16] R. A. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press, 2012.

[17] G. W. Stewart, J.-G. Sun, Matrix Perturbation Theory, Academic Press, 1990.

[18] H. Karimi, J. Nutini, M. Schmidt, Linear Convergence of Gradient and Proximal-Gradient Methods Under the Polyak-Łojasiewicz Condition, in: P. Frasconi, N. Landwehr, G. Manco, J. Vreeken (Eds.), Machine Learning and Knowledge Discovery in Databases, Springer, Cham, 2016, pp. 795–811.

[19] D. Goldfarb, A. Idnani, A Numerically Stable Dual Method for Solving Strictly Convex Quadratic Programs, Mathematical programming 27 (1) (1983) 1–33.