Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Notes on the Convergence of a Levenberg-Marquardt Method with Singular Scaling Matrices for Non-Zero Residue Problems

Rafaela Filippozzi[1] Everton Boos[2] Douglas S. Gonçalves[3] Fermín S. V. Bazán[4]
UFSC, Florianópolis, SC

Abstract. This work addresses the local and global convergence analysis of a variant of the Levenberg-Marquardt method (LMM), designed for non-linear least-squares problems with non-zero residue. Such variant, called LMM with singular scaling (LMMSS), allows the so-called LMM scaling matrix to be singular, which can be useful in certain applications. In order to handle the non-zero residue while preserving local convergence, a judicious choice of the LMM parameter is made based on the gradient linearization error which is dictated by non-linearity and residue size. The main contributions of this work are related to local and global convergence of LMMSS in this setting. More specifically, we demonstrate that the sequence of directions d_k generated by LMMSS is gradient-related and then prove that limit points of a line-search version of LMMSS are stationary for the least-squares function. Moreover, the local analysis is further demonstrated under an error bound condition on the gradient and for different hypotheses on the linearization error.

Keywords. Levenberg-Marquardt, Singular Scaling Matrix, Convergence analysis, Non-zero residue.

1 Introduction

In this paper, we delve into the investigation of the global and local convergence properties of the Levenberg–Marquardt method applied to a specific non-linear least-squares (NLS) problem:

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|F(x)\|^2 := \phi(x), \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is twice continuously differentiable. In particular, we focus on the **overdetermined** problem, where $m \geq n$.

Unlike previous research, which predominantly focused on zero-residue cases or assumed full rank of the Jacobian [5, 6, 9, 10, 12], we do not make the assumption of a zero residue at a solution or full rank of the Jacobian at such a point.

We define problem (1) as a non-zero-residue NLS problem if x^* is a global minimum of ϕ and $F(x^*) \neq 0$. Consequently, our interest lies in identifying the stationary points of ϕ, i.e., the set $X^* = \{x \in \mathbb{R}^n \mid J(x)^T F(x) = 0\}$, where $J(x) \in \mathbb{R}^{m\times n}$ is the Jacobian of F at x, under the assumption that $X^* \neq \emptyset$.

While works such as [1, 2] explore problems with non-zero residue, presenting local convergence of the LM method, the scaling matrix remains the identity matrix. In this work, besides considering non-zero residue, we consider the following iteration:

[1]rafaela.filippozzi@posgrad.ufsc.br
[2]everton.boos@ufsc.br
[3]douglas.goncalves@ufsc.br
[4]fermin.bazan@ufsc.br

$$(J_k^T J_k + \lambda_k L^T L) d_k = -J_k^T F_k \tag{2}$$

$$x_{k+1} = x_k + \alpha_k d_k, \quad \forall k \geq 0, \tag{3}$$

where $F_k := F(x_k)$, $J_k := J(x_k)$ is the Jacobian of F at x_k, $\{\lambda_k\}$ is a positive scalar, $\{\alpha_k\}$ is the step size, and $L^T L$ is called **scaling matrix** and it is allowed to be singular.

We refer to the iteration (2)–(3) as the Levenberg-Marquardt method with Singular Scaling (LMMSS). The choice of the scaling matrix as $L^T L$, allowed to be singular, differs from the classic LMM approach. This strategy is motivated by applications in inverse problems, and it departs from the results explored for problems with zero residue in [4].

Although the classical theory of the Levenberg-Marquardt method (LMM) is well-established, the consideration of a singular scaling matrix introduces unique challenges in convergence analysis, necessitating new theoretical tools and additional assumptions. The study aims to demonstrate that the convergence of LMM with singular scaling matrices is achievable under reasonable conditions, emphasizing the impact of problem-oriented singular scaling matrices on the quality of obtained approximate solutions. This approach, combined with the consideration of non-zero residue, diverges from existing literature, providing a new perspective on local and global convergence for LMM in the presence of singular scaling matrices.

Our study is organized as follows. In Section 2, we start with some assumptions that will be considered throughout the manuscript and also the mathematical background and preliminary results necessary for the upcoming sections. Comments concerning the local convergence analysis and obtained results are presented in Section 3, which are categorized into two cases based on the behavior of the rank of the Jacobian around a solution (constant and diminishing rank). Finally, Section 4 proves that, for LMMSS with an Armijo line-search scheme, every limit point of the generated sequence is a stationary point of the sum-of-squares function, thereby establishing the global convergence of this method. Our final considerations are provided in Section 5.

We emphasize that this work is only a brief presentation of the complete article [7], in which the reader can find a more in depth approach of the concepts discussed here, together with proofs and illustrative examples.

2 Assumptions and auxiliary tools

Our main assumptions are presented next.

Assumption 1. *The matrix $L \in \mathbb{R}^{p \times n}$ is full rank, where $m \geq n \geq p$, and there exist $\gamma > 0$ such that, for every $x \in \mathbb{R}^n$*

$$\|J(x)v\|^2 + \|Lv\|^2 \geq \gamma \|v\|^2, \quad \forall v \in \mathbb{R}^n. \tag{4}$$

Assumption 2. *For any $x^* \in X^*$, there exists a constant $\delta \in (0,1)$ and $L_0 > 0$ such that*

$$\|J(x) - J(y)\| \leq L_0 \|x - y\|, \tag{5}$$

for all $x, y \in B(x^*, \delta)$.

Assumption 2 asks the Jacobian to be locally Lipschitz and implies:

$$\|J(y)(x-y) - (F(x) - F(y))\| \leq L_1 \|x-y\|^2, \quad \forall x, y \in B(x^*, \delta) \tag{6}$$

where $L_1 = L_0/2$, that is, the error in the linear approximation of $F(x)$ around y is $O(\|x-y\|^2)$, for x and y in a neighborhood of x^*.

Due to the compactness of the ball $B(x^*, \delta)$, there exist positive constants L_2 and β such that $\|J(x)\| \leq L_2$ and $\|F(x)\| \leq \beta$ for all $x \in B(x^*, \delta)$. Therefore, since $\|J(x)\|$ is bounded in $B(x^*, \delta)$, by applying the mean value inequality, we can infer that $\|F(x) - F(y)\| \leq L_2\|x - y\|$, for all $x, y \in B(x^*, \delta)$. Additionally, the gradient $\nabla\phi(x) = J(x)^T F(x)$ is Lipschitz in $B(x^*, \delta)$:

$$\left\|J(x)^T F(x) - J(y)^T F(y)\right\| \leq L_3\|x - y\|, \quad \forall x, y \in B(x^*, \delta), \tag{7}$$

where $L_3 = L_2^2 + \beta L_0$.

Moreover, notice that for $z \in X^* \cap B(x^*, r)$ and $x, y \in B(x^*, r)$, we have

$$\begin{aligned}\|(J(x) - J(y))^T F(y)\| &= \|(J(x) - J(z) + J(z) - J(y))^T F(y)\| \\ &\leq \|(J(x) - J(z))^T F(y)\| + \|(J(z) - J(y))^T F(y)\| \\ &\leq L_0 L_2\|x - z\|\|y - z\| + \|J(x)^T F(z)\| \\ &\quad + L_0 L_2\|y - z\|^2 + \|J(y)^T F(z)\|.\end{aligned} \tag{8}$$

Lemma 2.1. *[1, Lemma 2.1] If Assumption 2 is satisfied, then there exists a value $L_4 > 0$ such that*

$$\left\|\nabla\phi(y) - \nabla\phi(x) - J(x)^T J(x)(y - x)\right\| \leq L_4 \|x - y\|^2 + \left\|(J(x) - J(y))^T F(y)\right\|, \tag{9}$$

for all $x, y \in B(x^*, \delta)$.

Next, we present the hypothesis on the error bound assumed throughout this work.

Assumption 3 (*Error bound*)**.** *For any $x^* \in X^*$, $\left\|J(x)^T F(x)\right\|$ provides an error bound in $B(x^*, \delta)$, i.e., there exists $\omega \in (0, \infty)$ such that*

$$\omega \operatorname{dist}(x, X^*) \leq \left\|J(x)^T F(x)\right\|, \quad \forall x \in B(x^*, \delta), \tag{10}$$

where $\operatorname{dist}(x, X^*) = \inf_{z \in X^*} \|x - z\|$. Throughout the text, given x, we shall denote by $\bar{x}$ an element of X^* such that $\|x - \bar{x}\| = dist(x, X^*)$.

From Assumptions 3 and (7), we obtain

$$\omega \operatorname{dist}(x_k, X^*) \leq \left\|J_k^T F_k\right\| \leq L_3 \operatorname{dist}(x_k, X^*). \tag{11}$$

The remaining assumptions focus, as outlined in (8), on the terms

$$\left\|J(x)^T F(z)\right\| \text{ and } \left\|J(y)^T F(z)\right\|.$$

These terms play a crucial role in controlling the error, as expressed in (9), of the "incomplete linearization" of the gradient. It is worth noting that each of these assumptions, motivated by [1], gives rise to distinct convergence rates and corresponding analyses.

Assumption 4. *For every $x \in B(x^*, \delta)$ and every $z \in X^* \cap B(x^*, \delta)$, the following inequality holds:*

$$\|(J(x) - J(z))^T F(z)\| \leq \theta\|x - z\|,$$

with $0 \leq \theta < \bar{\theta}(\omega, L_3, \lambda^*)$, where $\theta(\omega, L_3, \lambda^*)$ is a positive constant depending on ω from (10), L_3 from (7), and the smallest positive eigenvalue of $J_*^T J_*$, denoted by λ^*.

Assumption 5. *For every $x \in B(x^*, \delta)$ and every $z \in X^* \cap B(x^*, \delta)$, the following inequality holds:*

$$\|(J(x) - J(z))^T F(z)\| \leq C\|x - z\|^{1+r}, \tag{12}$$

with $r \in]0, 1[$ and $C \geq 0$.

DOI: 10.5540/03.2025.011.01.0446

Assumption 6. *For every $x \in B(x^*, \delta)$ and every $z \in X^* \cap B(x^*, \delta)$, the following inequality holds:*

$$\|(J(x) - J(z))^T F(z)\| \leq K\|x - z\|^2, \tag{13}$$

with $K \geq 0$.

In [1], a few simple non-zero-residue examples with two variables are presented, illustrating the motivation for such assumptions. Furthermore, it is evident that Assumption 6 implies Assumption 5, which in turn implies Assumption 4.

Apart from the above assumptions, a few auxiliary tools are necessary, the main one being the application of the Generalized Singular Value Decomposition (GSVD) [11]. A particularized version of the GSVD to the context of this work may be found in [8, p. 22]. This is an important tool in theoretical analysis because allows the connection, in a single decomposition, of a pair of unrelated matrices, as is the case with Jacobians of F throughout x_k and L.

The following results establish a limit on the step length based on the distance from the current iteration to the solution set, considering two distinct cases. Lemma 2.2 addresses the scenario where the rank of the Jacobian around the solution is constant, while Lemma 2.3 deals with the case of diminishing rank, depending on the definition of the LM sequence λ_k.

Lemma 2.2. *Suppose that Assumptions 1-3 are valid, and $rank(J(x)^T J(x)) = rank(J(x^*)^T J(x^*)) = q \geq 1$ for all $x \in B(x^*, \delta)$. If $x_k \in B(x^*, \delta)$ and $\lambda_k := \lambda(x_k) > 0$, then there exists $c_1 > 0$ such that*

$$\|d_k\| \leq c_1 \text{dist}(x_k, X^*). \tag{14}$$

Lemma 2.3. *Suppose that Assumptions 1–2 are valid. For $x_k \in B(x^*, \delta)$ and $rank(J_k) = \ell \geq rank(J_*) = q \geq 1$, and*

(a) Assumption 5 is satisfied with $\lambda_k = \left\|J_k^T F_k\right\|^r$, and $r \in]0, 1[$,

(b) Assumption 6 is satisfied with $\lambda_k = \left\|J_k^T F_k\right\|$,

then there exists $c_1 > 0$ such that $\|d_k\| \leq c_1 \text{dist}(x_k, X^*)$.

Remark 2.1. *For the scenario where the rank of the Jacobian matrix diminishes, under Assumption 4, the outcome regarding an upper bound on the step length in terms of the distance from the current iterate to the solution set remains open.*

3 Local convergence

Lemmas 2.2 and 2.3 showed that, under specific assumptions, $\|d_k\| \leq c_1 \|x_k - \bar{x}_k\|$. Such inequality is key for the local convergence analysis under the error bound condition (Assumption 3). In this section, we consider the "pure" LMMSS iteration, i.e., (2)–(3) with $\alpha_k = 1$ for every k.

The approach to obtain results concerning the local convergence of the proposed method is separated into two cases according to whether the Jacobian rank near the solution set is constant or not and according to the assumption under consideration (Assumption 4, 5 or 6). Due to the complexity on the explanation and to shorten the exposition, we mention that local quadratic convergence can be established under Assumption 6, with the Jacobian rank being either constant or diminishing. Additionally, local superlinear convergence can be obtained under Assumption 5, again considering constant rank on the Jacobian or not. Under Assumption 4, however, local linear convergence is proved only for the case in which the rank of the Jacobian matrix remains constant. Assuming that Assumption 4 is satisfied when the rank is decreasing, i.e., $rank(J_k) = \ell \geq rank(J_*) = q \geq 1$, the proof of local convergence is still open.

Again, for a complete explanation and proofs of this results, the reader is directed to [7].

4 Global convergence

For the analysis of global convergence, we consider Algorithm 1, a version of LMMSS with line-search for non-zero residue nonlinear least-squares. It is worth to point out that Algorithm 1 differs from the algorithm proposed in [4] in the choice of the LM parameter λ_k as well as the full step ($\alpha_k = 1$) acceptance criterion.

Initially, we will demonstrate that the sequence of directions d_k generated by Algorithm 1 is gradient-related (see [3, Proposition 1.2.1]). Subsequently, we will establish that any limit point of the sequence produced by this algorithm is a stationary point for (1), regardless of the initial point. This serves to prove the Global Convergence Theorem. Furthermore, in this section, we assume that Assumption 6 is satisfied to simplify notation and proofs.

We recall the definition of gradient-related directions from [3, Eq. (1.13)]:

Definition 4.1 (Gradient-related)**.** *Let $\{x_k\}$ and $\{d_k\}$ be sequences in $\mathbb{R}^n$. The sequence $\{d_k\}$ is said to be gradient-related to $\{x_k\}$ if, for each subsequence $\{x_k\}_{k\in\mathcal{K}}$ (with $\mathcal{K} \subseteq \mathbb{N}$) converging to a non-stationary point of ϕ, the corresponding subsequence $\{d_k\}_{k\in\mathcal{K}}$ is bounded and satisfies*

$$\limsup_{k\to\infty,\ k\in\mathcal{K}} \nabla\phi(x_k)^T d_k < 0. \tag{15}$$

Proposition 4.1. *Suppose that Assumptions 1-3, 6, and $\lambda_k = \left\|J_k^T F_k\right\|$ are satisfied. Let $\{d_k\}$, $\{x_k\}$ be sequences generated by Algorithm 1. Then, $\{d_k\}$ is* gradient-related.

This proposition ensures that Algorithm 1 generates a sequence of directions $\{d_k\}$ that are gradient-related. Hence, using [3, Proposition 1.2.1], global convergence can be established, as follows.

Theorem 4.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 1. Then, every limit point $\hat{x}$ of $\{x_k\}$ is such that $\nabla\phi(\hat{x}) = 0$.*

Proof. Let $K_1 = \{k \in \mathbb{N} \mid \|J(x_k+d_k)^T F(x_k+d_k)\| \le \vartheta\|J(x_k)^T F(x_k)\|\}$. If K_1 is infinite, it follows that $\|J(x_k)^T F(x_k)\| \to 0$, and therefore any limit point $\hat{x}$ of $\{x_k\}$ is such that $J(\hat{x})^T F(\hat{x}) = 0$, hence $\nabla\phi(\hat{x}) = 0$. Otherwise, if K_1 is finite, let us assume, without loss of generality, that $\|J(x_k + d_k)^T F(x_k + d_k)\| > \vartheta\|J(x_k)^T F(x_k)\|$, for each k, such that the step size is chosen to satisfy the Armijo condition. Since, by Proposition 4.1, the directions of Algorithm 1 are gradient-related, it follows from [3, Proposition 1.2.1] that any limit point $\hat{x}$ of $\{x_k\}$ is a stationary point of ϕ. ☐

Now, to connect local and global convergence, the next result establishes that under certain conditions $\alpha_k = 1$ for all k sufficiently large and then we can apply the local convergence theory to show that $\{\text{dist}(x_k, X^*)\}$ converges to zero quadratically.

Theorem 4.2. *Suppose that Assumption 1-3, 6 are satisfied and $\lambda_k = \left\|J_k^T F_k\right\|$. Moreover, assume that the level set $C_n := \{x \in \mathbb{R}^n : \phi(x) \le \phi(x_0)\}$ is compact for some x_0. Let $\{x_k\}$ be generated by Algorithm 1, using x_0 as the initial point. Then, for every k sufficiently large $\alpha_k = 1$ and the sequence $\{\text{dist}(x_k, X^*)\}$ converges to 0 quadratically.*

Remark 4.1. *We observe that in the case of zero residue, a previous result in [12, Theorem 3.1] requires the assumption that the limit point x^* is such that $F(x^*) = 0$. Also, in the case of unconstrained optimization, in order to prove a similar result the required assumption is that the Hessian at x^* is positive definite; however, such assumption would imply that x^* is an isolated stationary point [3, Proposition 1.2.1]. As we want to address the case of non-zero residue and possibly non-isolated stationary points, we considered the boundness of the level set instead of these other two.*

Algorithm 1: LMMSS for problems with non-zero residue with line-search

Input: $\nu, \zeta, \vartheta \in (0,1)$, F, J, L, and $x_0 \in \mathbb{R}^n$
Set $\lambda_0 = \|\nabla\phi(x_0)\|$
for $k = 0, 1, 2, \ldots$ **do**
  if $\|\nabla\phi(x_k)\| = 0$ **then**
    Stop with $\bar{x} = x_k$
  end
  Calculate $d_k = -(J_k^T J_k + \lambda_k L^T L)^{-1} J_k^T F_k$
  if $\|\nabla\phi(x_k + d_k)\| \le \vartheta \|\nabla\phi(x_k)\|$ **then**
    $\alpha_k = 1$
  else
    Choose m as the smallest non-negative integer such that

$$\phi(x_k + \zeta^m d_k) - \phi(x_k) \le \nu\zeta^m \nabla\phi(x_k)^T d_k \tag{16}$$

    Set $\alpha_k = \zeta^m$
  end
  Update $x_{k+1} = x_k + \alpha_k d_k$
  $\lambda_{k+1} = \|\nabla\phi(x_{k+1})\|$
end

Remark 4.2. *In the scenario where Assumption 5 is satisfied with a diminishing rank, is necessary to adjust the parameter λ_k of Algorithm 1 to $\lambda_k = \left\|J_k^T F_k\right\|^r$. Furthermore, when this hypothesis holds, the above results follow with minor changes in some constants, and we can show there exists k_0 such that $\alpha_k = 1, \forall k \ge k_0$ and thus $\operatorname{dist}(x_k, X^*)$ goes to zero superlinearly. When only Assumption 4 is satisfied, although the proofs of the Proposition 4.1 and Theorem 4.1 are analogous (when we consider constant rank), the proof that the full step size $\alpha_k = 1$ is always accepted for sufficiently large k remains open.*

5 Conclusion and future works

Motivated by insights from the article [4], which employs a version of the Levenberg-Marquardt method with a singular scaling matrix (named LMMSS) for non-linear least-squares problems with zero residue, we propose a new LMMSS method for problems with non-zero residue. Our study thoroughly investigated the local and global convergence of the proposed method.

For local convergence, we have shown that local quadratic convergence of $\operatorname{dist}(x_k, X^*)$ to zero is possible, even in the presence of a non-zero residue, with the Jacobian rank being either constant or diminishing. When the Jacobian rank is constant for $x \in B(x^*, \delta)$, even considering weaker assumptions, we manage to establish local linear and superlinear convergence (see Assumption 4 and 5, respectively). However, if only Assumption 4 holds in the case of diminishing rank, local convergence remains an open question.

Furthermore, we demonstrated the stationary nature of the limit points in the sequence generated by a line-search version of LMMSS (Algorithm 1). Global convergence was established under Assumption 6. Unfortunately, when only Assumption 4 is satisfied, in the case of diminishing rank, the global convergence of the algorithm remains inconclusive.

As future works, apart from tackling these open questions, we aim to conduct computational experiments to validate our theoretical results as well as to test the algorithm in certain inverse problems to assess its practical behaviour.

Acknowledgments

This work was supported by Brazilian agencies FAPESC (Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina), and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). DG acknowledges the support of CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil [grant number 305213/2021-0].

References

[1] R. Behling, D. S. Gonçalves, and S. A. Santos. "Local convergence analysis of the Levenberg–Marquardt framework for nonzero-residue nonlinear least-squares problems under an error bound condition". In: **Journal of Optimization Theory and Applications** 183.3 (2019), pp. 1099–1122.

[2] E. H. Bergou, Y. Diouane, and V. Kungurtsev. "Convergence and Complexity Analysis of a Levenberg–Marquardt Algorithm for Inverse Problems". In: **Journal of Optimization Theory and Applications** 185 (2020), pp. 927–944. DOI: https://doi.org/10.1007/s10957-020-01666-1.

[3] D. P. Bertsekas. **Nonlinear Programming**. Second. Athena Scientific, 1999.

[4] E. Boos, D.S. Gonçalves, and F.S.V. Bazán. "Levenberg-Marquardt method with singular scaling and applications". In: **Applied Mathematics and Computation** 474 (2024), p. 128688. DOI: https://doi.org/10.1016/j.amc.2024.128688.

[5] J. E. Dennis Jr. "Nonlinear least squares and equations". In: **The State of the Art in Numerical Analysis**. Ed. by D. Jacobs. London: Academic Press, 1977, pp. 269–312.

[6] J. E. Dennis Jr and R. B. Schnabel. **Numerical Methods for Unconstrained Optimization and Nonlinear Equations**. Philadelphia: SIAM, 1996.

[7] R. Filippozzi, E. Boos, D.S. Gonçalves, and F.S.V. Bazán. "Convergence analysis of Levenberg-Marquardt method with Singular Scaling for nonzero residue nonlinear least-squares problems". In: **arxiv** (2024). URL: https://arxiv.org/abs/2408.10370.

[8] P. C. Hansen. **Rank-Deficient and Discrete Ill-Posed Problems**. Philadelphia: SIAM, 1998.

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

[10] D. Marquardt. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". In: **SIAM Journal on Applied Mathematics** 11.2 (1963), pp. 431–441.

[11] C. F. Van Loan. "Generalizing the Singular Value Decomposition". In: **SIAM Journal on Numerical Analysis** 13.1 (1976), pp. 76–83.

[12] N. Yamashita and M. Fukushima. "On the rate of convergence of the Levenberg-Marquardt method". In: **Topics in Numerical Analysis**. Springer, 2001, pp. 239–249.

