Local convergence analysis of the Levenberg-Marquardt framework for nonzero-residue nonlinear least-squares problems under an error bound condition

Roger Behling[1], Douglas S. Gonçalves[2], and Sandra A. Santos[3]

[1]Federal University of Santa Catarina , Blumenau, SC, Brazil, roger.behling@ufsc.br
[2]Federal University of Santa Catarina , Florianópolis, SC, Brazil, douglas@mtm.ufsc.br
[3]University of Campinas , Campinas, SP, Brazil, sandra@ime.unicamp.br

August 23, 2018

Abstract

The Levenberg-Marquardt method (LM) is widely used for solving nonlinear systems of equations, as well as nonlinear least-squares problems. In this paper, we consider local convergence issues of the LM method when applied to nonzero-residue nonlinear least-squares problems under an error bound condition, which is weaker than requiring full-rank of the Jacobian in a neighborhood of a stationary point. Differently from the zero-residue case, the choice of the LM parameter is shown to be dictated by (i) the behavior of the rank of the Jacobian, and (ii) a combined measure of nonlinearity and residue size in a neighborhood of the set of (possibly non-isolated) stationary points of the sum of squares function.

Keywords. Local convergence; Levenberg-Marquardt method; Nonlinear least squares; Nonzero residue; Error bound

1 Introduction

This study investigates local convergence issues of the Levenberg Marquardt (LM) method applied to the nonlinear least-squares (NLS) problem. Differently from previous analyses in the literature [1, 2, 3, 4], we neither assume zero residue at a solution nor full column rank of the Jacobian at such a point.

In applied contexts, such as data fitting, parameter estimation, experimental design, and imaging problems, to name a few, admitting nonzero residue is essential for achieving meaningful solutions (see e.g. [5, 6, 7, 8, 9, 10, 11, 12]).

For a general residual function, the NLS problem is a nonconvex optimization problem for which the global minimum is unknown, opposed to the zero residue case. Thus, we will limit our attention to stationary points of the NLS objective function. We are particularly interested in the case of non-isolated solutions (stationary points), with a possible change (decrease) in the rank of the Jacobian as the generated sequence approaches the set of stationary points.

Given an initial point, sufficiently close to a stationary point of the NLS objective function, we are interested in the convergence analysis of the sequence generated by the local Levenberg-Marquardt (LM) iteration. Our contribution is to establish the local analysis based on an error bound condition upon the gradient of the NLS problem, *without* requiring zero residue neither full rank of the Jacobian at the stationary points.

In [1, 13], the local convergence of LM for the NLS problem has been established assuming full rank of the Jacobian at the solution, and the nonzero-residue case was handled by imposing a condition that combines the size of the residue and the problem nonlinearity. Under the assumption that the positive sequence of LM parameters is bounded, it was proved that the iterates converge linearly to the solution; in the zero residue case, assuming that the LM parameters are of the order of the norm of the gradient of the NLS function, such a convergence is proved to be quadratic.

The seminal work of Yamashita and Fukushima [4] showed for the first time the local convergence of LM for systems of nonlinear equations, under an error bound condition upon the norm of the residue. Assuming that the LM parameter is the squared norm of the residue, they established local quadratic convergence of the distance of the iterates to the solution set. Later, this result was improved by Fan and Yuan [2], who showed that the LM parameter may be actually chosen as the norm of the gradient to a power between 1 and 2, and the quadratic convergence is still attained. Moreover, they proved that the sequence of iterates itself converges quadratically to some point in the solution set.

In [14], Li et al. considered the problem of finding stationary points of a convex and smooth nonlinear function by an inexact regularized Newton method. Assuming an error bound condition upon the norm of the gradient, and under a convenient control of the inexactness, they proved local quadratic convergence of the iterates to a stationary point.

By addressing zero residue NLS problems, there are significant studies in the literature with local analysis of LM considering the norm of the residue vector as an error bound [15, 16, 17, 18, 19, 20]. On the other hand, rank-deficient nonlinear least-squares problems have been examined, see e.g. [21]. Nevertheless, up to our knowledge, an analysis for the nonzero-residue case combining an error bound condition upon the gradient, with the possibility of decreasing the rank of the Jacobian around the solution is a novelty.

Our study is organized as follows. The assumptions and the auxiliary results necessary to the analysis are established in Sections 2 and 3, respectively. The local convergence results are presented in Section 4, divided in two cases, according with the behavior of the rank of the Jacobian around the solution (constant and diminishing rank). Illustrative examples are described in Section 5, and our final considerations, in Section 6.

2 Assumptions

The nonlinear least-squares (NLS) problem is stated as follows

$$\min_x \quad \frac{1}{2}\|F(x)\|_2^2 := f(x), \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is twice continuously differentiable and $m \geq n$. The gradient of f is given by $\nabla f(x) = J(x)^T F(x)$, and the Hessian, by $\nabla^2 f(x) = J(x)^T J(x) + S(x)$, where $S(x) = \sum_{i=1}^m F_i(x)\nabla^2 F_i(x)$.

If x_* is a global minimizer of $f(x)$ and $F(x_*) \neq 0$, we say that problem (1) is a *nonzero-residue* NLS problem. We will limit our attention to stationary points of f, i.e., to the set

$$X^* = \{x \in \mathbb{R}^n \mid J(x)^T F(x) = 0\}, \tag{2}$$

under the assumption $X^* \neq \emptyset$.

The local LM iteration is defined by

$$(J_k^T J_k + \mu_k I)d_k = -J_k^T F_k, \tag{3}$$

$$x_{k+1} = x_k + d_k, \tag{4}$$

where $J_k := J(x_k)$, $J_k^T F_k := J(x_k)^T F(x_k)$ and $\{\mu_k\}$ is a positive scalar sequence.

Unless stated otherwise, $\|\cdot\|$ denotes the Euclidean vector norm and its induced matrix norm. $\mathcal{R}(A)$ and $\mathcal{N}(A)$ denote, respectively, the range and the null space of a matrix A. The transpose of vectors and matrices is denoted by $(\cdot)^T$, and $\lambda_{\min}(S)$ stands for the smallest eigenvalue of a symmetric matrix S. Throughout the text, x_k, x_{k+1}, d_k, μ_k, J_k, F_k are those defined in the LM iteration, i.e. by Eqs. (3) and (4). For two symmetric matrices C and D, $C \succeq D$ means that $C - D$ is positive semidefinite, and $C \succ D$ means that $C - D$ is positive definite.

Let $x_* \in X^*$ be a stationary point of f and $B(x_*, r) := \{x \in \mathbb{R}^n \mid \|x - x_*\| \leq r\}$, with $r < 1$. Our main assumptions are presented next.

Assumption 1. The Jacobian $J(x)$ is Lipschitz in a neighborhood $B(x_*, r)$ of a stationary point x_*, i.e., there exists a constant $L_0 \geq 0$ such that

$$\|J(x) - J(y)\| \leq L_0\|x - y\|, \quad \forall x, y \in B(x_*, r). \tag{5}$$

Assumption 1 (A1) implies that the error in the linearization of $F(x)$ around $y \in B(x_*, r)$ is bounded by the squared distance between y and $x \in B(x_*, r)$:

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

where $L_1 := L_0/2$. Since there are positive constants α and β such that $\|J(x)\| \leq \alpha$ and $\|F(x)\| \leq \beta$ for all $x \in B(x_*, r)$, we also have from A1 that the variation of $F(x)$ in $B(x_*, r)$ is bounded

$$\|F(x) - F(y)\| \leq L_2\|x - y\|, \quad \forall x, y \in B(x_*, r), \tag{7}$$

where $L_2 := \alpha$, and the gradient $\nabla f(x) = J(x)^T F(x)$ is Lipschitz in $B(x_*, r)$

$$\|J(x)^T F(x) - J(y)^T F(y)\| \leq L_3\|x - y\|, \quad \forall x, y \in B(x_*, r). \tag{8}$$

A useful bound is provided next, quantifying the error of an *incomplete linearization* for the gradient of the NLS function.

Lemma 1. *If A1 holds, then there exists $L_4 > 0$ such that, for all $x, y \in B(x_*, r)$,*

$$\|\nabla f(y) - \nabla f(x) - J(x)^T J(x)(y - x)\| \leq L_4\|x - y\|^2 + \|(J(x) - J(y))^T F(y)\|. \tag{9}$$

Proof From (6), and the consistency property of matrix norms, for all $x, y \in B(x_*, r)$, we obtain

$$\|J(x)^T F(y) - J(x)^T F(x) - J(x)^T J(x)(y - x)\| \leq L_1\|J(x)\|\|y - x\|^2,$$

and since $\|J(x)\| \leq \alpha$ for any $x \in B(x_*, r)$, we may write

$$\begin{aligned}&\big\|\big(J(x)^T F(y) - J(y)^T F(y)\big) + \\ &\big(J(y)^T F(y) - J(x)^T F(x) - J(x)^T J(x)(y - x)\big)\big\| \leq L_1\alpha\|x - y\|^2.\end{aligned}$$

From the reverse triangle inequality

$$\begin{aligned}\|J(y)^T F(y) - J(x)^T F(x) - J(x)^T J(x)(y - x)\| \quad &\leq \quad L_4\|x - y\|^2 \\ &\quad +\|(J(x) - J(y))^T F(y)\|,\end{aligned}$$

where $L_4 := L_1\alpha$. □

We have used the term 'incomplete linearization' because both the Gauss-Newton and the Levenberg-Marquardt methods consider $J(x)^T J(x)$ instead of $\nabla^2 f(x)$ in the first-order approximation of $\nabla f(y) = J(y)^T F(y)$ around $x \in B(x_*, r)$.

Also notice that, for $\bar{x} \in X^* \cap B(x_*, r)$ and $x, y \in B(x_*, r)$

$$\begin{aligned}\|(J(x) - J(y))^T F(y)\| &= \|(J(x) - J(\bar{x}) + J(\bar{x}) - J(y))^T F(y)\| \\ &\leq \|(J(x) - J(\bar{x}))^T F(y)\| + \|(J(\bar{x}) - J(y))^T F(y)\| \\ &= \|(J(x) - J(\bar{x}))^T (F(y) - F(\bar{x}) + F(\bar{x}))\| + \\ &\quad \|(J(y) - J(\bar{x}))^T (F(y) - F(\bar{x}) + F(\bar{x}))\| \\ &\leq L_0 L_2\|x - \bar{x}\|\|y - \bar{x}\| + \|J(x)^T F(\bar{x})\| + \\ &\quad L_0 L_2\|y - \bar{x}\|^2 + \|J(y)^T F(\bar{x})\|.\end{aligned} \tag{10}$$

Next we state the error bound hypothesis assumed along this work.

Assumption 2. $\|J(x)^T F(x)\|$ provides a local error bound (w.r.t. X^*) on $B(x_*, r)$, i.e., there exists a positive constant ω such that

$$\omega \operatorname{dist}(x, X^*) \leq \|J(x)^T F(x)\|, \quad \forall x \in B(x_*, r), \tag{11}$$

where $\operatorname{dist}(x, X^*) = \inf_{z \in X^*} \|x - z\|$.

Throughout the text, let $\bar{x} \in X^*$ be such that $\|x - \bar{x}\| = \operatorname{dist}(x, X^*)$.

Remark 1. *When J_* has full rank, and $F(x_*) = 0$, it is possible to show that $\omega = \lambda_{\min}(J(x_*)^T J(x_*))/2$ in the error bound condition by using the result in [22, Thm. 4.9] applied to $G(x) = J(x)^T F(x)$ and noting that $G'(x_*) = J(x_*)^T J(x_*)$ because $S(x_*) = 0$.*

From Assumption 2 (A2), and (8) we obtain

$$\omega \operatorname{dist}(x_k, X^*) \leq \|J_k^T F_k\| \leq L_3 \operatorname{dist}(x_k, X^*). \tag{12}$$

The remaining assumptions concern the terms $\|J(x)^T F(\bar{x})\|$ and $\|J(y)^T F(\bar{x})\|$, present in (10), that are used to bound the error (9) in the incomplete linearization of the gradient. Each of these assumptions leads to different convergence rates (and analyses).

Assumption 3. There exists a constant $\sigma \geq 0$ such that, for all $x \in B(x_*, r)$, and for all $z \in X^* \cap B(x_*, r)$, it holds

$$\|(J(x) - J(z))^T F(z)\| \leq \sigma \|x - z\|. \tag{13}$$

Remark 2. *In [1], the authors analyzed the local convergence of Gauss-Newton and Levenberg-Marquardt methods under nonsingularity of $J(x_*)^T J(x_*)$, and the following assumption*

$$\|(J(x) - J(x_*))^T F(x_*)\| \leq \sigma \|x - x_*\|, \quad \forall x \in B(x_*, r), \tag{14}$$

where $0 \leq \sigma < \lambda_{\min}(J(x_*)^T J(x_*))$. They also comment that

$$(J(x) - J(x_*))^T F(x_*) \approx S(x_*)(x - x_*),$$

and interpret σ in (14) as a combined absolute measure of nonlinearity and residue size.

Remark 3. *Due to A1, the consistency of the matrix norm, and the continuity of $\|F(x)\|$, we obtain*

$$\|(J(x) - J(z))^T F(z)\| \leq \|J(x) - J(z)\| \|F(z)\| \leq \beta L_0 \|x - z\|,$$

for all $x, z \in B(x_*, r)$. Thus, under the previous hypotheses, A3 is satisfied with $\bar{\sigma} := \beta L_0$. However, the above inequality provides a loose bound, and the corresponding $\bar{\sigma}$ may not be small enough to ensure convergence, as we will see ahead.

Assumption 4. For all $x \in B(x_*, r)$ and for all $z \in X^* \cap B(x_*, r)$, it holds

$$\|(J(x) - J(z))^T F(z)\| \leq C\|x - z\|^{1+\delta}, \tag{15}$$

with $\delta \in (0, 1)$ and $C \geq 0$.

Remark 4. *This assumption is somehow related to Assumption 2 in [23] which imposes a condition on the "quality of approximation" for $\nabla f(x)$.*

Assumption 5. For all $x \in B(x_*, r)$ and for all $z \in X^* \cap B(x_*, r)$, it holds

$$\|(J(x) - J(z))^T F(z)\| \leq K \|x - z\|^2, \tag{16}$$

with $K \geq 0$.

Remark 5. *It is easy to see that if $F(x)$ is linear or $F(z) = 0$ for $z \in X^*$, then A3, A4 and A5 are satisfied with $\sigma = C = K = 0$.*

3 Auxiliary results

This section gathers preliminary results that will show useful for our subsequent analysis. We first recall a classical bound, followed by a conveniently stated consequence. Let A^+ denote the Moore-Penrose pseudo-inverse.

Lemma 2. *[24, p. 43] Assume that* $\text{rank}(A) = p \geq 1$ *and rank*$(A + E) \leq \text{rank}(A)$ *and* $\|A^+\|\|E\| < 1$. *Then,* $\text{rank}(A + E) = p$ *and*

$$\|(A + E)^+\| \leq \frac{\|A^+\|}{1 - \|A^+\|\|E\|}.$$

Corollary 1. *Given $\kappa > 1$, if* $\text{rank}(J(x)^T J(x)) = \text{rank}(J_*^T J_*) = p \geq 1$*, and*

$$\|J(x)^T J(x) - J_*^T J_*\| \leq \left(1 - \frac{1}{\kappa}\right) \frac{1}{\|(J_*^T J_*)^+\|},$$

then $\|(J(x)^T J(x))^+\| \leq \kappa \|(J_*^T J_*)^+\|$.

Proof This is just a simple application of the previous lemma, using $A = J_*^T J_*$ and $E = J(x)^T J(x) - J_*^T J_*$. In fact, if $\text{rank}(J(x)^T J(x)) = \text{rank}(J_*^T J_*) = p \geq 1$, and

$$\|(J_*^T J_*)^+\| \|J(x)^T J(x) - J_*^T J_*\| \leq \left(1 - \frac{1}{\kappa}\right), \tag{17}$$

for a given $\kappa > 1$, then Lemma 2 applies, and it follows that $\|(J(x)^T J(x))^+\| \leq \kappa \|(J_*^T J_*)^+\| = \kappa / \lambda_p^*$, where λ_p^* is the smallest positive eigenvalue of $J_*^T J_*$. $\square$

Since $J(x)^T J(x)$ is continuous, it is clear that for $r \in (0,1)$, and sufficiently small, condition (17) is satisfied for all $x \in B(x_*, r)$.

The next results bound the step length by the distance of the current iterate to the solution set, under distinct possibilities for (i) the rank of the Jacobian around the solution; (ii) the definition of the LM sequence $\{\mu_k\}$; (iii) the assumptions upon the error in the incomplete linearization for the gradient.

Lemma 3. *Let $x_* \in X^*$ and $r \in (0,1)$ such that A1 and (17) hold. If $x_k \in B(x_*, r)$ and $rank(J(x)) = p \geq 1$ for all $x \in B(x_*, r)$, then there exists $c_1 > 0$ such that $\|d_k\| \leq c_1 \, dist(x_k, X^*)$.*

Proof Recall that $J_k^T F_k \in \mathcal{R}(J_k^T) \perp \mathcal{N}(J_k) = \mathcal{N}(J_k^T J_k)$. Let $J_k^T J_k = Q \Lambda Q^T$ be an eigendecomposition where $\lambda_1 \geq \cdots \geq \lambda_p > \lambda_{p+1} = \cdots = \lambda_n = 0$. Thus

$$d_k = -(J_k^T J_k + \mu_k I)^{-1} J_k^T F_k = -\sum_{i=1}^{n} \frac{q_i^T J_k^T F_k}{\lambda_i + \mu_k} q_i = -\sum_{i=1}^{p} \frac{q_i^T J_k^T F_k}{\lambda_i + \mu_k} q_i. \tag{18}$$

Therefore

$$\|d_k\|^2 = \sum_{i=1}^{p} \frac{(q_i^T J_k^T F_k)^2}{(\lambda_i + \mu_k)^2} \leq \sum_{i=1}^{p} \frac{(q_i^T J_k^T F_k)^2}{(\lambda_p)^2} = \frac{1}{(\lambda_p)^2} \sum_{i=1}^{n} (q_i^T J_k^T F_k)^2 = \frac{\|J_k^T F_k\|^2}{\lambda_p^2}. \tag{19}$$

Then, by (8), we get

$$\|d_k\| \leq (1/\lambda_p) \|J_k^T F_k\| \leq (1/\lambda_p) L_3 \text{dist}(x_k, X^*) \leq (\kappa/\lambda_p^*) L_3 \text{dist}(x_k, X^*),$$

where the last inequality follows from Corollary 1, and $\kappa > 1$. Thus $\|d_k\| \leq c_1 \, \text{dist}(x_k, X^*)$, with $c_1 = \kappa L_3 / \lambda_p^*$. □

Lemma 4. *Let $x_* \in X^*$, $rank(J(x_*)) \geq 1$ and assume A1 and A2 hold. For a sufficiently small $r > 0$, if $x_k \in B(x_*, r)$, and either*

a) A4 holds with $\mu_k = \|J_k^T F_k\|^\delta$, for $\delta \in (0,1)$, or

b) A5 holds with $\mu_k = \|J_k^T F_k\|$,

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

Proof Suppose now that the rank of $J(x)$ decreases as x approaches the solution set X^*. Following ideas from [14], and assuming $\text{rank}(J_*) \geq 1$, let

$$J_*^T J_* = (Q_{*,1} \quad Q_{*,2}) \begin{pmatrix} \Lambda_{*,1} & 0 \\ 0 & 0 \end{pmatrix} (Q_{*,1} \quad Q_{*,2})^T,$$

where $\Lambda_{*,1} \succeq 2\underline{\lambda} I \succ 0$, and

$$J_k^T J_k = (Q_{k,1} \quad Q_{k,2}) \begin{pmatrix} \Lambda_{k,1} & 0 \\ 0 & \Lambda_{k,2} \end{pmatrix} (Q_{k,1} \quad Q_{k,2})^T.$$

By the continuity of $J(x)^T J(x)$ and its eigenvalues, then $\Lambda_{k,1} \to \Lambda_{*,1}$ and $\Lambda_{k,2} \to 0$ as $x_k \to x_*$. Thus, for $r > 0$ small enough, $\Lambda_{k,1} \succeq \underline{\lambda} I \succ 0$.

Multiplying (3) by $Q_{k,1}^T$, we obtain

$$(\Lambda_{k,1} + \mu_k I) Q_{k,1}^T d_k = -Q_{k,1}^T J_k^T F_k,$$

and hence,

$$\|Q_{k,1}^T d_k\| = \|(\Lambda_{k,1} + \mu_k I)^{-1} Q_{k,1}^T J_k^T F_k\| \leq \frac{L_3}{\underline{\lambda}} \|x_k - \bar{x}_k\|. \tag{20}$$

From the triangle inequality, (9) and because $\|(\Lambda_{k,2} + \mu_k I)^{-1} \Lambda_{k,2}\| \leq 1$ and $\|Q_{k,2}^T\| \leq 1$, we have

$$\begin{aligned}
\|Q_{k,2}^T d_k\| &= \|(\Lambda_{k,2} + \mu_k I)^{-1} Q_{k,2}^T J_k^T F_k\| \\
&\leq \|(\Lambda_{k,2} + \mu_k I)^{-1} Q_{k,2}^T (J_k^T F_k - J(\bar{x}_k)^T F(\bar{x}_k) - J_k^T J_k (x_k - \bar{x}_k))\| + \\
&\quad \|(\Lambda_{k,2} + \mu_k I)^{-1} Q_{k,2}^T J_k^T J_k (x_k - \bar{x}_k)\| \\
&\leq \frac{L_4 \|x_k - \bar{x}_k\|^2 + \|(J(x_k) - J(\bar{x}_k))^T F(\bar{x}_k)\|}{\mu_k} + \\
&\quad \|(\Lambda_{k,2} + \mu_k I)^{-1} \Lambda_{k,2} Q_{k,2}^T (x_k - \bar{x}_k)\| \\
&\leq L_4 \frac{\|x_k - \bar{x}_k\|^2}{\mu_k} + \frac{\|J(x_k)^T F(\bar{x}_k)\|}{\mu_k} + \|x_k - \bar{x}_k\|.
\end{aligned} \tag{21}$$

In case A4 holds and $\mu_k = \|J_k^T F_k\|^\delta$, for $\delta \in (0,1)$, from (21), and (12), we obtain

$$\|Q_{k,2}^T d_k\| \leq \left(\frac{L_4}{\omega^\delta} + \frac{C}{\omega^\delta} + 1 \right) \|x_k - \bar{x}_k\|,$$

and hence

$$\|d_k\|^2 = \|Q_{k,1}^T d_k\|^2 + \|Q_{k,2}^T d_k\|^2 \leq \left(\frac{L_3^2}{\underline{\lambda}^2} + \left(\frac{L_4 + C}{\omega^\delta} + 1 \right)^2 \right) \|x_k - \bar{x}_k\|^2,$$

i.e., $\|d_k\| \leq c_1 \operatorname{dist}(x_k, X^*)$, where $c_1^2 = \dfrac{L_3^2}{\underline{\lambda}^2} + \left(\dfrac{L_4 + C}{\omega^\delta} + 1 \right)^2$.

Similarly, in case A5 holds and $\mu_k = \|J_k^T F_k\|$, then $\|d_k\| \leq c_1 \operatorname{dist}(x_k, X^*)$ holds with $c_1^2 = \dfrac{L_3^2}{\underline{\lambda}^2} + \left(\dfrac{L_4 + K}{\omega} + 1 \right)^2$. □

Lemma 5. *Let $x_* \in X^*$, and assume that A1 and A2 hold. For a small enough $r > 0$, if $x_k \in B(x_*, r)$, A3 holds, and $\sigma + L_4 \|x_k - \bar{x}_k\| \leq \mu_k$ then there exists $c_1 > 0$ such that $\|d_k\| \leq c_1 \, dist(x_k, X^*)$.*

Proof Using the same reasoning in the proof of the previous lemma, we have $\|Q_{k,1}^T d_k\| \leq \dfrac{L_3}{\underline{\lambda}} \|x_k - \bar{x}_k\|$, and from (21), by using A3, we obtain

$$\|Q_{k,2}^T d_k\| \leq \left(\frac{L_4 \|x_k - \bar{x}_k\| + \sigma}{\mu_k} \right) \|x_k - \bar{x}_k\| + \|x_k - \bar{x}_k\|.$$

For $\mu_k \geq L_4\|x_k - \bar{x}_k\| + \sigma$, it follows that $\|Q_{k,2}^T d_k\| \leq 2\|x_k - \bar{x}_k\|$. Therefore $\|d_k\| \leq c_1 \text{dist}(x_k, X^*)$, with $c_1^2 = \dfrac{L_3^2}{\underline{\lambda}^2} + 4$. □

4 Local convergence

We start our local convergence analysis with an auxiliary result that comes from the error bound assumption.

Lemma 6. *Under A1 and A2, let $\{x_k\}$ be the LM sequence. If $x_{k+1}, x_k \in B(x_*, r/2)$ and $\|d_k\| \leq c_1\|x_k - \bar{x}_k\|$, then*

$$\begin{aligned} \omega\, dist(x_{k+1}, X^*) \quad &\leq \quad (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1 \|x_k - \bar{x}_k\| + \\ &\quad\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\|, \end{aligned} \tag{22}$$

where $L_5 := L_0 L_2 (2 + c_1)(1 + c_1)$.

Proof From the error bound condition (11), the LM iteration (3)-(4), inequality (9), along with the reverse triangle inequality, and assuming $x_{k+1}, x_k \in B(x_*, r/2)$, we have

$$\begin{aligned} \omega\, \text{dist}(x_{k+1}, X^*) \quad &\leq \quad \|J(x_{k+1})^T F(x_{k+1})\| = \|J(x_k + d_k)^T F(x_k + d_k)\| \\ &\leq \quad \|J_k^T F_k + J_k^T J_k d_k\| + L_4\|d_k\|^2 + \\ &\quad\quad \|(J(x_k) - J(x_{k+1}))^T F(x_{k+1})\| \\ &\leq \quad \mu_k\|d_k\| + L_4\|d_k\|^2 + \|(J(x_k) - J(x_{k+1}))^T F(x_{k+1})\| \\ &\leq \quad (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1\|x_k - \bar{x}_k\| + \\ &\quad\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \end{aligned}$$

where the last inequality follows from $\|d_k\| \leq c_1\|x_k - \bar{x}_k\|$, $\|x_{k+1} - \bar{x}_k\| \leq (1 + c_1)\|x_k - \bar{x}_k\|$, and inequality (10). □

Henceforth, the convergence analysis is divided in two cases, according to the behavior of the rank of $J(x)$ in a neighborhood of x_*, namely, constant rank, and diminishing rank. These cases are discussed in the following subsections.

4.1 Constant rank

In this section we consider that $1 \leq \text{rank}(J(x_*)) = p \leq n$ and $\text{rank}(J(x)) = \text{rank}(J(x_*))$ for every x in $B(x_*, r)$. From Lemma 3, recall that $\|d_k\| \leq c_1 \,\text{dist}(x_k, X^*)$, with $c_1 = \kappa L_3/\lambda_p^*$, independently on whether A3, A4 or A5 holds.

4.1.1 Under Assumption 3

In this subsection, we assume A1, A2 and A3 to hold. Additionally, if we suppose that the constant $\sigma > 0$ in A3 is sufficiently small, we derive local linear convergence of the sequence $\{dist(x_k, X^*)\}$ to 0 and convergence of the LM sequence $\{x_k\}$ to a solution in X^*, with the LM parameter being chosen as

$\mu_k = ||J_k^T F_k||$. This assertion is formally stated in Theorem 1. In preparation, two auxiliary results are proved.

Lemma 7. *Assume that A1, A2, A3 hold, and rank*$(J(x)) =$ *rank*$(J(x_*)) \geq 1$ *in* $B(x_*, r)$. *If* $\eta\omega > (2 + c_1)\sigma$ *in (13), for some* $\eta \in (0, 1)$, $\mu_k = \|J_k^T F_k\|$, x_k, $x_{k+1} \in B(x_*, r/2)$ *and dist*$(x_k, X^*) < \varepsilon$, *where*

$$\varepsilon = \min\left\{\frac{r/2}{1 + \dfrac{c_1}{1-\eta}}, \frac{\eta\omega - (2 + c_1)\sigma}{L_4c_1^2 + L_5 + L_3c_1}\right\}, \tag{23}$$

then dist$(x_{k+1}, X^*) \leq \eta$ *dist*(x_k, X^*).

Proof If $J(x)$ has constant rank in $B(x_*, r)$, $\mu_k = \|J_k^T F_k\|$, and A3 holds, then from (22) and (12), we have

$$\begin{aligned}
\omega\,\mathrm{dist}(x_{k+1}, X^*) &\leq (L_4c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1\|x_k - \bar{x}_k\| + \\
&\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \\
&\leq (L_4c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + L_3c_1\|x_k - \bar{x}_k\|^2 + \\
&\quad \sigma\|x_k - \bar{x}_k\| + \sigma\|x_{k+1} - \bar{x}_k\| \\
&\leq (L_4c_1^2 + L_5 + L_3c_1)\|x_k - \bar{x}_k\|^2 + (2 + c_1)\sigma\|x_k - \bar{x}_k\| \\
&\leq \left[(L_4c_1^2 + L_5 + L_3c_1)\varepsilon + (2 + c_1)\sigma\right]\|x_k - \bar{x}_k\|.
\end{aligned}$$

Let us denote $L_6 := L_4c_1^2 + L_5 + L_3c_1$. Thus, for $\varepsilon \leq \dfrac{\eta\omega - (2 + c_1)\sigma}{L_6}$, we obtain $\mathrm{dist}(x_{k+1}, X^*) \leq \eta\,\mathrm{dist}(x_k, X^*)$. □

Remark 6. *The condition* $\sigma < \eta\dfrac{\omega}{2 + c_1} < \eta\dfrac{\omega}{c_1} = \dfrac{\eta}{\kappa}\dfrac{\omega}{L_3}\lambda_p^* < \lambda_p^*$ *resembles the hypothesis* $\sigma < \lambda_n^*$ *used in [1] for the full rank case.*

If x_k, $x_{k+1} \in B(x_*, r/2)$, and $\mathrm{dist}(x_k, X^*) \leq \varepsilon$ for all k, it follows that the sequence $\{\mathrm{dist}(x_k, X^*)\}$ converges linearly to zero. This occurs whenever $x_0 \in B(x_*, \varepsilon)$, as shows the next lemma.

Lemma 8. *Suppose the assumptions of Lemma 7 hold and let* ε *be given by (23). If* $x_0 \in B(x_*, \varepsilon)$, *then* $x_{k+1} \in B(x_*, r/2)$ *and dist*$(x_k, X^*) \leq \varepsilon$ *for all* $k \in \mathbb{N}$.

Proof The proof is by induction. For $k = 0$, we have

$$\mathrm{dist}(x_0, X^*) \leq \|x_0 - x_*\| \leq \varepsilon$$

and

$$\begin{aligned}
\|x_1 - x_*\| &\leq \|x_1 - x_0\| + \|x_0 - x_*\| \\
&\leq \|d_0\| + \varepsilon \\
&\leq c_1\mathrm{dist}(x_0, X^*) + \varepsilon \\
&\leq (1 + c_1)\varepsilon \leq r/2.
\end{aligned}$$

For $k \geq 1$, let us assume that $x_i \in B(x_*, r/2)$ and $\mathrm{dist}(x_{i-1}, X^*) \leq \varepsilon$ for $i = 1, \ldots, k$. Then from Lemma 7 applied to x_{k-1} and x_k, we get

$$\|x_k - \bar{x}_k\| = \mathrm{dist}(x_k, X^*) \leq \eta \, \mathrm{dist}(x_{k-1}, X^*) \leq \eta \, \varepsilon < \varepsilon.$$

Additionally, by Lemma 7 and the induction hypothesis for $i \leq k$, we obtain

$$\mathrm{dist}(x_i, X^*) \leq \eta \, \mathrm{dist}(x_{i-1}, X^*) \leq \cdots \leq \eta^i \, \mathrm{dist}(x_0, X^*) \leq \eta^i \, \varepsilon < \varepsilon. \qquad (24)$$

Thus, using Lemma 3

$$\begin{aligned} \|x_{k+1} - x_*\| &\leq \|x_1 - x_*\| + \sum_{i=1}^{k} \|d_i\| \\ &\leq \|x_1 - x_*\| + \sum_{i=1}^{k} c_1 \mathrm{dist}(x_i, X^*) \\ &\leq (1 + c_1)\varepsilon + c_1 \varepsilon \sum_{i=1}^{k} \eta^i \\ &\leq (1 + c_1)\varepsilon + c_1 \varepsilon \sum_{i=1}^{\infty} \eta^i \\ &= \left(1 + \frac{c_1}{1 - \eta}\right) \varepsilon \leq r/2. \end{aligned}$$

□

From the previous two lemmas, it follows that if $x_0 \in B(x_*, \varepsilon)$, with ε defined by (23), then $\{\mathrm{dist}(x_k, X^*)\}$ converges linearly to zero, where $\{x_k\}$ is generated by the LM method, with $\mu_k = \|J_k^T F_k\|$. In order to prove that $\{x_k\}$ converges to some solution $\bar{x} \in X^* \cap B(x_*, \varepsilon)$, it suffices to show that $\{x_k\}$ is a Cauchy sequence.

Theorem 1. *Suppose that A1, A2, A3 hold and rank$(J(x))$ = rank$(J(x_*)) \geq 1$ in $B(x_*, r)$. Let $\{x_k\}$ be generated by LM with $\mu_k = \|J_k^T F_k\|$, for every k. If $\sigma < \eta\omega/(2 + c_1)$, for some $\eta \in (0, 1)$, and $x_0 \in B(x_*, \varepsilon)$, with $\varepsilon > 0$ given by (23), then $\{dist(x_k, X^*)\}$ converges linearly to zero. Moreover, the sequence $\{x_k\}$ converges to a solution $\bar{x} \in X^* \cap B(x_*, r/2)$.*

Proof The first assertion follows directly from Lemmas 7 and 8. Since $\{\mathrm{dist}(x_k, X^*)\}$ converges to zero and $x_k \in B(x_*, r/2)$ for all k, it suffices to show that $\{x_k\}$ converges. From Lemma 3, Lemma 7 and (24)

$$\|d_k\| \leq c_1 \mathrm{dist}(x_k, X^*) \leq c_1 \eta^k \, \mathrm{dist}(x_0, X^*) \leq c_1 \varepsilon \, \eta^k,$$

for all $k \geq 1$. Thus, for any positive integers ℓ, q such that $\ell \geq q$

$$\|x_\ell - x_q\| \leq \sum_{i=q}^{\ell-1} \|d_i\| \leq \sum_{i=q}^{\infty} \|d_i\| \leq c_1 \varepsilon \sum_{i=q}^{\infty} \eta^i,$$

which implies that $\{x_k\} \subset \mathbb{R}^n$ is a Cauchy sequence, and hence it converges. □

4.1.2 Under Assumption 4

Let us rename $\bar{r}$ as the radius of the ball for which A1, A2 and A4 hold, so that the constants ω, L_3, c_1 and C have been obtained within $B(x_*, \bar{r})$. For a possibly smaller radius $r < \min\left\{\frac{1}{2}\left(\frac{\eta\omega}{C(2+c_1)}\right)^{1/\delta}, \bar{r}\right\}$, it is clear that A4 implies A3 with $\sigma = C(2r)^\delta < \eta\frac{\omega}{2+c_1}$, $\eta \in (0,1)$ within $B(x_*, r)$. Thus, under A4, due to Lemmas 7 and 8, and Theorem 1, the linear convergence is ensured. Moreover, from (22), the triangular inequality, Lemma 3, the fact that $\mu_k = \|J_k^T F_k\|$, $\|x_k - \bar{x}_k\| < 1$, and inequality (12), we obtain

$$\begin{aligned}
\omega\,\mathrm{dist}(x_{k+1}, X^*) &\leq (L_4c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1\|x_k - \bar{x}_k\| + \\
&\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \\
&\leq (L_4c_1^2 + L_5 + L_3c_1)\|x_k - \bar{x}_k\|^2 + \\
&\quad C\|x_k - \bar{x}_k\|^{1+\delta} + C\|x_{k+1} - \bar{x}_k\|^{1+\delta} \\
&\leq L_6\|x_k - \bar{x}_k\|^2 + \\
&\quad C(1 + (1+c_1)^{1+\delta})\|x_k - \bar{x}_k\|^{1+\delta} \\
&\leq \left[L_6 + C(1 + (1+c_1)^{1+\delta}\right]\|x_k - \bar{x}_k\|^{1+\delta} \\
&:= \widetilde{C}\|x_k - \bar{x}_k\|^{1+\delta}.
\end{aligned}$$

Therefore,

$$\omega\frac{\mathrm{dist}(x_{k+1}, X^*)}{\mathrm{dist}(x_k, X^*)} \leq L_6\mathrm{dist}(x_k, X^*) + \frac{o(\mathrm{dist}(x_k, X^*))}{\mathrm{dist}(x_k, X^*)},$$

meaning that $\{\mathrm{dist}(x_k, X^*)\}$ converges to zero superlinearly.

On the other hand, because

$$\omega\,\mathrm{dist}(x_{k+1}, X^*) \leq \widetilde{C}\mathrm{dist}(x_k, X^*)^\delta\,\mathrm{dist}(x_k, X^*) \leq \widetilde{C}\widehat{\varepsilon}^\delta\mathrm{dist}(x_k, X^*),$$

the linear convergence is also ensured by taking $x_0 \in B(x_*, \widehat{\varepsilon})$, with $\widehat{\varepsilon} = \max\{\varepsilon_1, \varepsilon_2\}$ where

$$\varepsilon_1 = \min\left\{\frac{\bar{r}/2}{1 + \dfrac{c_1}{1-\eta}}, \left(\frac{\eta\omega}{\widetilde{C}}\right)^{1/\delta}\right\},$$

and

$$\varepsilon_2 = \min\left\{\frac{r/2}{1 + \dfrac{c_1}{1-\eta}}, \max\left\{\frac{\eta\omega - (2+c_1)\sigma}{L_6}, \left(\frac{\eta\omega}{\widetilde{C}}\right)^{1/\delta}\right\}\right\},$$

with $\sigma = C(2r)^\delta$.

4.1.3 Under Assumption 5

As before, let $\bar{r}$ denote the radius of the ball for which A1, A2 and A5 hold, so that the constants ω, L_3, c_1 and K have been obtained within $B(x_*, \bar{r})$. For $r < \min\left\{\frac{\eta\omega}{2K(2+c_1)}, \bar{r}\right\}$, A5 implies A3 with $\sigma = K(2r) < \frac{\eta\omega}{2+c_1}, \eta \in (0,1)$ within $B(x_*, r)$. Thus, under A5, due to Lemmas 7 and 8, and Theorem 1, the linear convergence of $\{\mathrm{dist}(x_k, X^*)\}$ to zero is ensured.

Furthermore, from (22), if A5 holds, we also have that

$$\omega\,\mathrm{dist}(x_{k+1}, X^*) \leq \widehat{K}\,\mathrm{dist}(x_k, X^*)^2, \tag{25}$$

with $\widehat{K} = L_6 + K(1 + (1 + c_1)^2)$. Thus, since

$$\omega\,\mathrm{dist}(x_{k+1}, X^*) \leq (\widehat{K}\mathrm{dist}(x_k, X^*))\,\mathrm{dist}(x_k, X^*) \leq \widehat{K}\widehat{\varepsilon}\mathrm{dist}(x_k, X^*),$$

the linear convergence is also ensured by taking $x_0 \in B(x_*, \widehat{\varepsilon})$, with $\widehat{\varepsilon} = \max\{\varepsilon_1, \varepsilon_2\}$, where

$$\varepsilon_1 = \min\left\{\frac{\bar{r}/2}{1 + \dfrac{c_1}{1-\eta}}, \frac{\eta\omega}{\widehat{K}}\right\},$$

and

$$\varepsilon_2 = \min\left\{\frac{r/2}{1 + \dfrac{c_1}{1-\eta}}, \max\left\{\frac{\eta\omega - (2 + c_1)\sigma}{L_6}, \frac{\eta\omega}{\widehat{K}}\right\}\right\},$$

with $\sigma = 2Kr$.

Moreover, (25) also implies that $\mathrm{dist}(x_k, X^*)$ goes to zero quadratically.

4.2 Diminishing rank

In case the rank of $J(x_k)$ decreases as x_k approaches the solution set X^*, the convergence analysis is a bit more challenging, and Lemmas 4 and 5 give *sufficient* conditions to ensure that $\|d_k\| = O(\mathrm{dist}(x_k, X^*))$. Notice that both lemmas directly depend on the bound for $\|J(x_k)^T F(\bar{x}_k)\|$, and on the choice of μ_k, in contrast with the analysis of the previous subsection, in which we used $\mu_k = \|J_k^T F_k\|$ independently of A3, A4 or A5. Moreover, from inequality (22) of Lemma 6, we can observe that the choice of μ_k also impacts on the convergence rate, and on the size of the corresponding convergence region, as detailed next.

4.2.1 Under Assumption 5

Suppose A5 holds. Thus, from Lemma 4, for $\mu_k = \|J_k^T F_k\|$, we have $\|d_k\| \leq c_1\,\mathrm{dist}(x_k, X^*)$, with

$$c_1 = \sqrt{\frac{L_3^2}{\underline{\lambda}^2} + \left(\frac{L_4 + K}{\omega} + 1\right)^2}.$$

From Lemma 6, A5, $\mu_k = \|J_k^T F_k\|$, (8), and assuming $\operatorname{dist}(x_k, X^*) < \varepsilon$ we have

$$\begin{aligned}
\omega \operatorname{dist}(x_{k+1}, X^*) &\leq (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1 \|x_k - \bar{x}_k\| + \\
&\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \\
&\leq (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + L_3 c_1 \|x_k - \bar{x}_k\|^2 + \\
&\quad K\|x_k - \bar{x}_k\|^2 + K(1+c_1)^2\|x_k - \bar{x}_k\|^2 \\
&= \left[L_6 + K(1 + (1+c_1)^2)\right] \|x_k - \bar{x}_k\|^2 \qquad (26) \\
&\leq \left[L_6 + K(1 + (1+c_1)^2)\right] \varepsilon \|x_k - \bar{x}_k\|.
\end{aligned}$$

Hence, for $\varepsilon \leq \dfrac{\eta\,\omega}{L_6 + K(1+(1+c_1)^2)}$, we obtain $\operatorname{dist}(x_{k+1}, X^*) \leq \eta \operatorname{dist}(x_k, X^*)$, thus proving the following result.

Lemma 9. *Let A1, A2 and A5 hold. If $\mu_k = \|J_k^T F_k\|$, x_k, $x_{k+1} \in B(x_*, r/2)$ and dist$(x_k, X^*) < \varepsilon$, where*

$$\varepsilon = \min \left\{ \frac{r/2}{1 + \dfrac{c_1}{1-\eta}}, \ \frac{\eta\,\omega}{L_6 + K(1+(1+c_1)^2)} \right\}, \qquad (27)$$

with $\eta \in (0,1)$, then dist$(x_{k+1}, X^*) \leq \eta$ dist(x_k, X^*).

Lemma 10. *Suppose the assumptions of Lemma 9 hold and let ε be given by (27). If $x_0 \in B(x_*, \varepsilon)$, then $x_{k+1} \in B(x_*, r/2)$ and dist$(x_k, X^*) \leq \varepsilon$ for all k.*

Proof Using $\varepsilon > 0$ given by (27), and Lemma 9, the proof is analogous to the one of Lemma 8. □

It follows from Lemma 9 and Lemma 10 that, for x_0 sufficiently close to x_*, the sequence $\{\operatorname{dist}(x_k, X^*)\}$ converges quadratically to zero.

Theorem 2. *Suppose A1, A2 and A5 hold. Let $\{x_k\}$ be generated by LM with $\mu_k = \|J_k^T F_k\|$. If $x_0 \in B(x_*, \varepsilon)$, with $\varepsilon > 0$ from (27), then $\{$dist$(x_k, X^*)\}$ converges quadratically to zero. Moreover, the sequence $\{x_k\}$ converges to a solution $\bar{x} \in X^* \cap B(x_*, r/2)$.*

Proof The proof follows the same lines of Theorem 1. Since $\{\operatorname{dist}(x_k, X^*)\}$ converges to zero, the quadratic convergence comes from (26). □

4.2.2 Under Assumption 4

Suppose A4 holds. Thus, for $\mu_k = \|J_k^T F_k\|^\delta$, with $\delta \in (0,1)$, from Lemma 4, $\|d_k\| \leq c_1 \operatorname{dist}(x_k, X^*)$ is satisfied with

$$c_1 = \sqrt{\frac{L_3^2}{\underline{\lambda}^2} + \left(\frac{L_4 + C}{\omega^\delta} + 1\right)^2}.$$

From Lemma 6, A4, (8), Lemma 4, and because $\mathrm{dist}(x_k, X^*) \le r/2 < 1/2$,

$$\begin{aligned}
\omega\,\mathrm{dist}(x_{k+1}, X^*) &\le (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1 \|x_k - \bar{x}_k\| + \\
&\quad \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \\
&\le (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + L_3^\delta c_1 \|x_k - \bar{x}_k\|^{1+\delta} + \\
&\quad C(1 + (1+c_1)^{1+\delta})\|x_k - \bar{x}_k\|^{1+\delta} \\
&\le [L_4 c_1^2 + L_5 + L_3^\delta c_1 + C(1 + (1+c_1)^{1+\delta})]\mathrm{dist}(x_k, X^*)^{1+\delta} \\
&\le [L_4 c_1^2 + L_5 + L_3^\delta c_1 + C(1 + (1+c_1)^{1+\delta})]\,\varepsilon^\delta\,\mathrm{dist}(x_k, X^*). \\
&:= \widehat{C}\varepsilon^\delta \mathrm{dist}(x_k, X^*). \qquad (28)
\end{aligned}$$

Hence, for $\varepsilon^\delta \le \dfrac{\eta\,\omega}{\widehat{C}}$, we have $\mathrm{dist}(x_{k+1}, X^*) \le \eta\,\mathrm{dist}(x_k, X^*)$, so that the next result is established.

Lemma 11. *Let A1, A2 and A4 hold. If $\mu_k = \|J_k^T F_k\|^\delta$, with $\delta \in (0,1)$, x_k, $x_{k+1} \in B(x_*, r/2)$ and dist$(x_k, X^*) < \varepsilon$, where*

$$\varepsilon = \min\left\{ \frac{r/2}{1 + \dfrac{c_1}{1-\eta}}, \left(\frac{\eta\,\omega}{\widehat{C}}\right)^{1/\delta} \right\}, \qquad (29)$$

with $\eta \in (0,1)$, then dist$(x_{k+1}, X^*) \le \eta$ dist(x_k, X^*).

Lemma 12. *Suppose the assumptions of Lemma 11 hold and let $\varepsilon > 0$ be given by (29). If $x_0 \in B(x_*, \varepsilon)$, then $x_{k+1} \in B(x_*, r/2)$ and dist$(x_k, X^*) \le \varepsilon$ for all k.*

Proof Using $\varepsilon > 0$ given by (29), and Lemma 11, the proof is analogous to the one of Lemma 8. □

It follows from Lemma 11 and Lemma 12 that, for x_0 sufficiently close to x_*, the sequence $\{\mathrm{dist}(x_k, X^*)\}$ converges superlinearly to zero.

Theorem 3. *Suppose A1, A2 and A4 hold. Let $\{x_k\}$ be generated by LM with $\mu_k = \|J_k^T F_k\|^\delta$, for $\delta \in (0,1)$. If $x_0 \in B(x_*, \varepsilon)$, with $\varepsilon > 0$ from (29), then $\{dist(x_k, X^*)\}$ converges superlinearly to zero. Moreover, the sequence $\{x_k\}$ converges to a solution $\bar{x} \in X^* \cap B(x_*, r/2)$.*

Proof The proof follows the same lines of Theorem 1. Since $\{\mathrm{dist}(x_k, X^*)\}$ converges to zero, the superlinear convergence comes from (28)

$$\lim_{k\to\infty} \frac{\mathrm{dist}(x_{k+1}, X^*)}{\mathrm{dist}(x_k, X^*)} \le \lim_{k\to\infty} \frac{\widehat{C}}{\omega}\,\mathrm{dist}(x_k, X^*)^\delta = 0.$$

□

4.2.3 Under Assumption 3

When A3 holds, we only have $\|J(x)^T F(\bar{x})\| \leq \sigma\|x-\bar{x}\|$. In this case, according to Lemma 5, for $\sigma + L_4\|x_k - \bar{x}_k\| \leq \mu_k$ we obtain $\|d_k\| \leq c_1 \text{dist}(x_k, X^*)$, where $c_1 = \left(\frac{L_3^2}{\underline{\lambda}^2} + 4\right)^{1/2}$. The analogous of Lemma 7 is given below.

Lemma 13. *Let A1, A2 and A3 hold.*

If

$$\sigma + L_4\|x_k - \bar{x}_k\| \leq \mu_k \leq \theta(\sigma + L_4\|x_k - \bar{x}_k\|), \tag{30}$$

for $\theta > 1$, *and*

$$\eta\,\omega > ((1+\theta)c_1 + 2)\sigma \tag{31}$$

for some $\eta \in (0,1)$, x_k, $x_{k+1} \in B(x_*, r/2)$ *and* $dist(x_k, X^*) < \varepsilon$, *where*

$$\varepsilon = \min\left\{\frac{r/2}{1 + \frac{c_1}{1-\eta}}, \frac{\eta\,\omega - ((1+\theta)c_1 + 2)\sigma}{L_4 c_1(c_1+\theta) + L_5}\right\}, \tag{32}$$

then $dist(x_{k+1}, X^*) \leq \eta\, dist(x_k, X^*)$.

Proof If (30) and A3 holds, then from Lemma 6, Lemma 5, and $\text{dist}(x_k, X^*) < \varepsilon$, we have

$$\begin{aligned}
\omega\,\text{dist}(x_{k+1}, X^*) &\leq (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + \mu_k c_1\|x_k - \bar{x}_k\| \\
&\quad + \|J(x_k)^T F(\bar{x}_k)\| + \|J(x_{k+1})^T F(\bar{x}_k)\| \\
&\leq (L_4 c_1^2 + L_5)\|x_k - \bar{x}_k\|^2 + (\theta\sigma + \theta L_4\|x_k - \bar{x}_k\|)c_1\|x_k - \bar{x}_k\| \\
&\quad + (2 + c_1)\sigma\|x_k - \bar{x}_k\| \\
&\leq (L_4 c_1(c_1+\theta) + L_5)\|x_k - \bar{x}_k\|^2 + (2 + (1+\theta)c_1)\sigma\|x_k - \bar{x}_k\| \\
&\leq [(L_4 c_1(c_1+\theta) + L_5)\varepsilon + (2 + (1+\theta)c_1)\sigma]\,\text{dist}(x_k, X^*),
\end{aligned}$$

so that $\text{dist}(x_{k+1}, X^*) \leq \eta\,\text{dist}(x_k, X^*)$ as long as $\varepsilon \leq \dfrac{\eta\omega - ((1+\theta)c_1 + 2)\sigma}{L_4 c_1(c_1+\theta) + L_5}$. □

Lemma 14. *Suppose the assumptions of Lemma 13 hold and let* $\varepsilon > 0$ *be given by (32). If* $x_0 \in B(x_*, \varepsilon)$, *then* $x_{k+1} \in B(x_*, r/2)$ *and* $dist(x_k, X^*) \leq \varepsilon$ *for all* k.

Proof Using $\varepsilon > 0$ given by (32), and Lemma 13, the proof follows analogously to the one of Lemma 8. □

Theorem 4. *Suppose A1, A2 and A3 hold. Let* $\{x_k\}$ *be generated by LM, with* μ_k *satisfying (30) for all* k, *for some* $\theta > 1$. *If* $x_0 \in B(x_*, \varepsilon)$, *with* $\varepsilon > 0$ *from (32), and there exists* $\eta \in (0,1)$ *such that (31) is satisfied, then* $\{dist(x_k, X^*)\}$ *converges linearly to zero. Moreover, the sequence* $\{x_k\}$ *converges to a solution* $\bar{x} \in X^* \cap B(x_*, r/2)$.

Proof It follows directly from Lemma 13 and Lemma 14, similarly to the proof of Theorem 1. □

We highlight that in the case of diminishing rank, under assumption A3, we cannot choose $\{\mu_k\}$ freely, and this sequence must be bounded away from zero. Besides, if the sequence $\{\mu_k\}$ satisfies (30), the constant ω from the error bound condition (11) must be sufficiently larger than σ, i.e., $\omega > (c_1(1+\theta)+2)\sigma$, in order to achieve linear convergence. Roughly speaking, from the definitions of θ and c_1 in this subsection, ω should be at least six times larger than σ for the above results to hold.

Remark 7. *An alternative convergence analysis would be devised by viewing the LM iteration as an inexact regularized Newton method, and then applying the results of [14]. The LM iteration can be written as*

$$(\nabla^2 f(x_k) + \mu_k I)d_k + g_k = r_k,$$

where $r_k = -S_k(J_k^T J_k + \mu_k I)^{-1} g_k$. *However, for* $\mu_k = \|g_k\|$, *we have* $\|r_k\| \leq \|S_k\|$, *and the condition* $\|r_k\| = O(dist(x_k, X^*)^2)$ *required in [14] turns out to be stronger than assumptions A4 and A3.*

5 Illustrative examples

We have devised a few simple nonzero-residue examples with two variables, so that, besides exhibiting the constants associated with the assumptions, some geometric features can also be depicted. These examples cover different scenarios with constant or diminishing rank, and isolated or non-isolated stationary points.

Example 1: constant rank and non-isolated minimizers

The elements of this example are

$$F(x) = \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1^2 + x_2^2 - 9 \end{bmatrix} \quad \text{and} \quad J(x) = \begin{bmatrix} 2x_1 & 2x_2 \\ 2x_1 & 2x_2 \end{bmatrix}. \tag{33}$$

It is not hard to see that $J(x)$ has rank one everywhere, except for the origin. The set of minimizers X^* consists of points verifying $x_1^2 + x_2^2 - 5 = 0$, so that $\text{dist}(x, X^*) = |\,\|x\| - \sqrt{5}\,|$. Also, for any $x_* \in X^*$, $f(x_*) = 16$. Figure 1 shows geometric aspects of this instance. Let us consider $x_* = (0, \sqrt{5})$ and the ball $B(x_*, r)$ with $r = 1/2$. Since

$$\|J(x) - J(y)\| = 2\sqrt{2}\|x - y\|,$$

it is clear that A1 is fulfilled with $L_0 = 2\sqrt{2}$.

Moreover, as

$$\|\nabla f(x)\| = 4\,\|x\|\,|\,\|x\|^2 - 5\,| = \left(4\,\|x\|\,|\,\|x\| + \sqrt{5}\,|\right)\text{dist}(x, X^*),$$

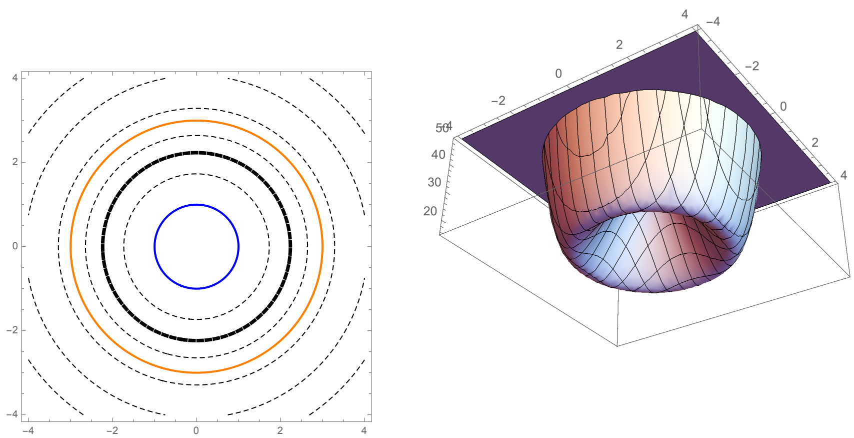

Figure 1: Curves $F_1(x) = 0$ and $F_2(x) = 0$ of (33). The level sets of $f(x) = \frac{1}{2}\|F(x)\|^2$ (dashed curves), on the left; a view of the graph of the least-squares function f, on the right.

Table 1: Example 1, with starting point $x_0 = (0, \sqrt{5} + 0.03)$, and stopping criterion $\|J_k^T F_k\| < 10^{-8}$.

k	$\mathrm{dist}(x_k, X^*)$	$\|J_k^T F_k\|$
0	3.00000000E−02	1.22425753E+00
1	4.74608414E−03	4.24704095E−02
2	6.15190439E−06	5.50243395E−05
3	1.03534958E−11	9.26050348E−11

for $x \in B(x_*, r)$, then A2 is verified with $\omega = 41 - 6\sqrt{5} \approx 27.5836$.

Notice that

$$\|(J(x) - J(z))^T F(z)\| = 4\|x - z\||z_1^2 + z_2^2 - 5|$$

for any $z, x \in \mathbb{R}^2$. In particular, for $z \in X^*$, $\|(J(x) - J(z))^T F(z)\| = 0$, for any $x \in B(x_*, r)$. Therefore, A5 is fulfilled with $K = 0$, and consequently A4 and A3 are verified due to the unified analysis of Section 4.1.

Following the reasoning of Section 4.1, since $L_3 = 4\left(3\left(\sqrt{5} + \frac{1}{2}\right)^2 - 5\right) \approx 69.8328$, and $\lambda_p^* = 40$, for $\kappa = 1.001$ we have $c_1 \approx 1.7475$. Thus, by using $\eta = 1/2$, $\mu_k = \|J_k^T F_k\|$, starting from $x_0 \in B(x_*, \widehat{\varepsilon})$, where $\widehat{\varepsilon} < 0.03621$, at least linear convergence of $\mathrm{dist}(x_k, X^*)$ to zero is ensured, but due to (25) we actually obtain quadratic convergence (see Tables 1 and 2).

Table 2: Example 1, with starting point $x_0 = (0.01, \sqrt{5} - 0.01)$, and stopping criterion $\|J_k^T F_k\| < 10^{-8}$.

k	dist(x_k, X^*)	$\|J_k^T F_k\|$
0	9.97753898E−03	3.96434291E−01
1	3.42773490E−04	3.06575421E−03
2	2.03970840E−08	1.82437070E−07
3	8.88178420E−16	7.10548157E−15

Example 2: diminishing rank and non-isolated stationary points

The residual function and its Jacobian are given by

$$F(x) = \begin{bmatrix} x_1^3 - x_1 x_2 + 1 \\ x_1^3 + x_1 x_2 + 1 \end{bmatrix} \quad \text{and} \quad J(x) = \begin{bmatrix} 3x_1^2 - x_2 & -x_1 \\ 3x_1^2 + x_2 & x_1 \end{bmatrix}. \tag{34}$$

Problem (1) with the function F of (34) has an isolated global minimizer at $(-1, 0)^T$ and a non-isolated set of local minimizers given by $\{x \in \mathbb{R}^2 \mid x_1 = 0\}$. We consider the non-isolated set of minimizers

$$X^* = \{(0, \xi), \xi \in \mathbb{R}\},$$

so that dist$(x, X^*) = |x_1|$.

The rank of the Jacobian varies along $\mathbb{R}^2$. It is 0 at the origin, it is 1 along the axis $x_1 = 0$ for $x_2 \neq 0$, and rank$(J(x_1, x_2)) = 2$ wherever $x_1 \neq 0$. Figure 2 depicts geometric features of the problem.

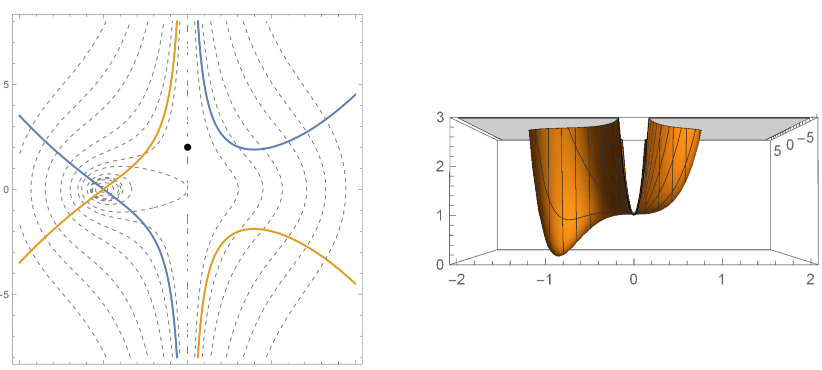

Figure 2: Curves $F_1(x) = 0$ and $F_2(x) = 0$ of (34), a local non-isolated minimizer at $(0, 2)^T$, and the level sets of $f(x) = \frac{1}{2}\|F(x)\|^2$ (dashed curves), on the left; a view of the graph of the least-squares function f, on the right.

By considering the local non-isolated minimizer $x_* = (0, 2)^T$, setting $r = 0.1$

Table 3: Example 2, with starting point $x_0 = (0.008, 2.0)$, and stopping criterion $\|J_k^T F_k\| < 10^{-8}$

k	$\text{dist}(x_k, X^*)$	$\|J_k^T F_k\|$
0	8.00000000E−03	6.43845091E−02
1	1.62858018E−05	1.30286953E−04
2	6.63076730E−11	5.30457012E−10
3	1.08989247E−17	0.00000000E+00

as the radius of the neighborhood, we can verify that $\|J(x) - J(y)\| \leq L_0\|x - y\|$, for all $x, y \in B(x_*, r)$, with $L_0 \approx 1.482$.

Since

$$\|\nabla f(x)\| = \left(2\sqrt{x_1^2 x_2^2 + (3x_1 + 3x_1^4 + x_2^2)^2}\right)|x_1|,$$

A2 is verified with $\omega \approx 7.016$.

Also $\|(J(x) - J(z))^T F(z)\| = 6x_1^2$, for any $x \in B(x_*, r)$ and $z \in X^*$. Thus, A5 is fulfilled with $K = 6$, and according to Section 4.2.1, convergence is assured for $x_0 \in B(x_*, \varepsilon)$, where $\varepsilon \approx 0.01$. Table 3 exemplifies the quadratic convergence of $\text{dist}(x_k, X^*)$.

Example 3: diminishing rank under A3

The elements of this example are

$$F(x) = \begin{bmatrix} \frac{\cos x_1}{9} - x_2 \sin x_1 \\ \frac{\sin x_1}{9} + x_2 \cos x_1 \end{bmatrix} \quad \text{and} \quad J(x) = \begin{bmatrix} -x_2 \cos x_1 - \frac{\sin x_1}{9} & -\sin x_1 \\ -x_2 \sin x_1 + \frac{\cos x_1}{9} & \cos x_1 \end{bmatrix}. \tag{35}$$

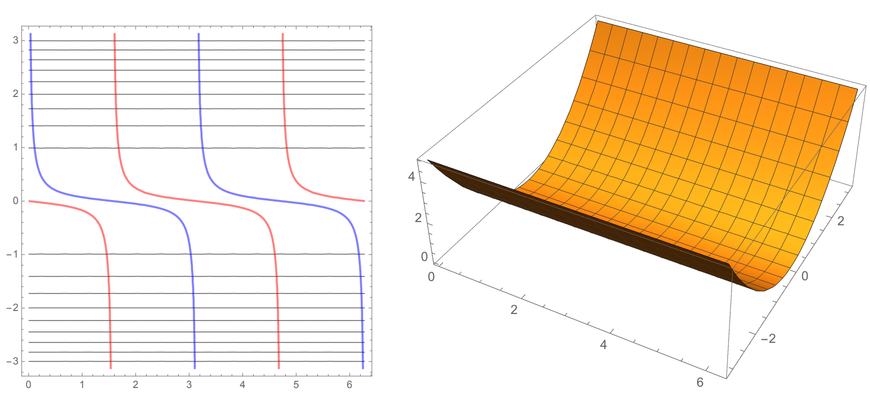

Figure 3: Curves $F_1(x) = 0$ and $F_2(x) = 0$ of (35), non-isolated local minimizers at the line $x_2 = 0$, and the level sets of $f(x) = \frac{1}{2}\|F(x)\|^2$ (dashed curves), on the left; a view of the graph of the least-squares function f, on the right.

Table 4: Example 3, with starting point $x_0 = (\pi, 0.001)$, $\mu_k \equiv 0.2$, and stopping criterion $\|J_k^T F_k\| < 10^{-10}$

k	$\text{dist}(x_k, X^*)$	$\|J_k^T F_k\|$
0	1.00000000E−04	1.00000000E−04
1	1.24236463E−04	1.24236463E−04
2	1.54346729E−05	1.54346729E−05
4	2.38228720E−07	2.38228720E−07
6	3.67697607E−09	3.67697607E−09
8	5.67528256E−11	5.67528255E−11

The set of non-isolated minimizers is given by $X^* = \{x \in \mathbb{R}^2 : x_2 = 0\}$ and $\text{dist}(x, X^*) = |x_2|$ (see Figure 3). Let us consider $x_* = (\pi, 0)$ and $r = 0.1$. For every $\bar{x} \in X^*$, we have $f(\bar{x}) = \frac{1}{2}\frac{1}{81}$. After some algebraic manipulation, we can show that $\|J(x) - J(y)\| \leq L_0\|x - y\|$ holds with $L_0 \approx 1.05839$, for $x, y \in B(x_*, r)$. Moreover, because $\|\nabla f(x)\| = |x_2|$, for $x \in B(x_*, r)$, A2 is verified with $\omega = 1$. Concerning the linearization error, we have

$$\|J(x)^T F(z)\| = \frac{1}{81}\sqrt{81\sin^2(x_1 - z_1) + (9x_2\cos(x_1 - z_1) + \sin(x_1 - z_1))^2},$$

for any $x \in B(x_*, r)$ and $z \in X^*$. In this case, A5 and A4 are not verified, but A3 holds with $\sigma \approx 0.1175$, implying in a convergence radius $\varepsilon \approx 0.003$ (cf. (32)). We remark that the parameters $\eta \in (0, 1)$ and $\theta > 1$ were chosen in order to maximize ε.

As discussed in Section 4.2.3, because the rank of $J(x)$ decreases as x approaches X^*, and only A3 holds, the regularization parameter μ_k must be bounded away from zero, satisfying (30). Therefore, we only obtain linear convergence of $\{\text{dist}(x_k, X^*)\}$ to zero, as shown in Table 4.

Example 4: when LM fails

The purpose of this example is to illustrate that, even in a neighborhood of an isolated minimizer, if the linearization error (the local combined measure of nonlinearity and residue size) is not small enough, then the LM iteration may not converge. Consider

$$F(x) = \begin{bmatrix} x_2 - x_1^2 - 1 \\ x_2 + x_1^2 + 1 \end{bmatrix} \quad \text{and} \quad J(x) = \begin{bmatrix} -2\,x_1 & 1 \\ 2\,x_1 & 1 \end{bmatrix}. \tag{36}$$

The rank of the Jacobian is 1 along the axis $x_1 = 0$ and $\text{rank}(J(x_1, x_2)) = 2$ wherever $x_1 \neq 0$. In Figure 4 we illustrate the geometric features of the problem. The minimizer is $x_* = (0, 0)$, so that $\text{dist}(x, X^*) = \|x\|$.

Since $\|J(x) - J(y)\| = 2\sqrt{2}|x_1 - y_1|$, assumption A1 holds with $L_0 = 2\sqrt{2}$. From the expression $\|\nabla f(x)\| = 2\sqrt{4(x_1 + x_1^3)^2 + x_2^2}$, it follows that A2 is fulfilled with $\omega = 2$, and because $\|J(x)^T F(x_*)\| = 4|x_1|$, it is easy to check that A4 and A5 are not verified, and only A3 holds with $\sigma = 4$. The above constants

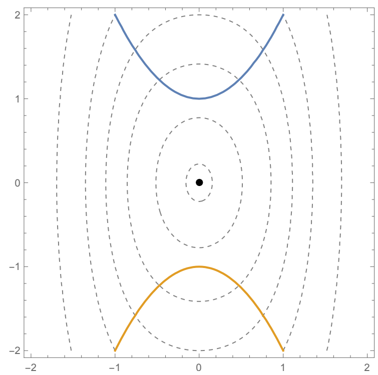

Figure 4: Curves $F_1(x) = 0$ and $F_2(x) = 0$ of (36), the solution $x_* = (0, 0)^T$, and the level sets of $\frac{1}{2}\|F(x)\|^2$ (dashed curves).

are the same, independently of the radius $r > 0$, and then, because $\sigma > \omega$, even with μ_k defined by (30), linear convergence cannot be achieved.

6 Conclusions and future perspectives

We have established the local convergence of the LM method for the NLS problem with possible nonzero residue, under an error bound condition upon the gradient of the nonlinear least-squares function $f(x) = \frac{1}{2}\|F(x)\|^2$. Suitable choices for the LM parameter were made, according to the behavior of the rank of the Jacobian in a neighborhood of x_*, a (possibly non-isolated) stationary point of f. The results hold whenever the error in the linearization of the gradient of the NLS function stays under control.

From the analysis of Section 4, such an error depends on a combined measure of nonlinearity and residue size around x_*. As proved in the local analysis, and illustrated in the examples of Section 5, local quadratic convergence of $\text{dist}(x_k, X^*)$ to zero is possible, even in case of nonzero residue, as long as such a measure goes to zero as fast as $\text{dist}(x_k, X^*)^2$ (see Assumption A5). Since $\|(J(x) - J(z))^T F(z)\| \approx \|S(z)(x - z)\|$, the contribution of $S(x)$ to $\nabla^2 f(x)$ should vanish as x approaches the set X^*, to obtain fast convergence.

Moreover, this measure must be small enough (Assumption A3) in order to guarantee at least linear convergence. We stress that, if only A3 holds in case of diminishing rank, the LM parameter must be bounded away from zero to achieve linear convergence of $\text{dist}(x_k, X^*)$ to zero.

Even for a simple nonzero residue problem satisfying the error bound condition upon the gradient, we have shown that the Levenberg-Marquardt method may fail to converge. This very example (Example 4) clearly reveals where the main difficulties lie on, namely, on the measure of residue size and nonlinearity that determines the quality of the incomplete information used to approximate the actual Jacobian of the vector valued function $J(x)^T F(x)$. In fact, if this measure is not small enough, then the LM iterations may not converge at all. In such a case, when $S(x)$ is not available, first-order methods such as quasi-Newton methods may be more appropriate for minimizing the nonlinear least-squares function.

We believe these theoretical results may shed some light in the development of LM algorithms and inspire new adaptive strategies for choosing the regularization parameter which allow not only the globalization of LM iterations, but also the automatic tuning of such a parameter in order to obtain the best possible local convergence rates. This is subject of future research.

Another closely related idea, currently under investigation, is to consider LP-quasi-Newton methods [25, 26, 27]. For the NLS problem the corresponding subproblem reads

$$\begin{array}{rl} \min\limits_{d,\gamma} & \gamma \\ \text{s.t} & \|\nabla f(x_k) + J_k^T J_k d\| \leq \gamma \|\nabla f(x_k)\| \\ & \|d\| \leq \gamma \|\nabla f(x_k)\|, \end{array}$$

where $\|.\|$ denotes a linearizable norm, and the solution of this subproblem provides the search direction d_k. Notice that, under an error bound condition, such constraints already impose that $\|\nabla f(x_k) + J_k^T J_k d_k\| = O(\text{dist}(x_k, X^*))$ and $\|d_k\| = O(\text{dist}(x_k, X^*))$, which are key intermediate results to prove the local convergence. In this case, the LM parameter is implicitly determined by the Langrange multiplier of the second constraint. Studying the behavior of LP-quasi-Newton methods for nonzero-residue nonlinear least squares problems is ongoing work.

Acknowledgments

This work was partially supported by the Brazilian research agencies CNPq (*Conselho National de Desenvolvimento Científico e Tecnológico*) and FAPESP (*Fundação de Amparo à Pesquisa do Estado de São Paulo*): D. S. Gonçalves grant 421386/2016-9, S. A. Santos grants 302915/2016-8, 2013/05475-7 and 2013/07375-0.

References

[1] Dennis Jr., J.E., Schnabel, R.B.: Numerical methods for unconstrained optimization and nonlinear equations, *Classics in Applied Mathematics*, vol. 16. SIAM, Philadelphia (1996)

[2] Fan, J., Yuan, Y.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Computing **74**(1), 23–29 (2005)

[3] Gratton, S., Lawless, A.S., Nichols, N.K.: Approximate Gauss-Newton methods for Nonlinear Least Squares problems SIAM J. Optim. **18**, 106–132 (2007)

[4] Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. In: A. G., C. X. (eds.) Topics in numerical analysis, pp. 239–249. Springer, Viena (2001)

[5] Bellavia, S., Riccietti, E.: On an elliptical trust-region procedure for ill-posed nonlinear least-squares problems. J. Optim. Theory Appl. **178**(3), 824–859 (2018)

[6] Cornelio, A.: Regularized nonlinear least squares methods for hit position reconstruction in small gamma cameras. Appl. Math. Comput. **217**(12), 5589–5595 (2011)

[7] Deidda, G., Fenu, C., Rodriguez, G.: Regularized solution of a nonlinear problem in electromagnetic sounding. Inverse Probl. **30**(12), 27pp. (2014). URL `http://dx.doi.org/10.1088/0266-5611/30/12/125014`

[8] Henn, S.: A Levenberg-Marquardt scheme for nonlinear image registration. BIT Numer. Math. **43**(4), 743–759 (2003)

[9] Landi, G., Piccolomini, E.L., Nagy, J.G.: A limited memory bfgs method for a nonlinear inverse problem in digital breast tomosynthesis. Inverse Probl. **33**(9), 21pp. (2017). URL `https://doi.org/10.1088/1361-6420/aa7a20`

[10] López, D.C., Barz, T., Korkel, S., Wozny, G.: Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Comput. Chem. Eng. **77**(Supplement C), 24–42 (2015). URL `http://dx.doi.org/10.1016/j.compchemeng.2015.03.002`

[11] Mandel, J., Bergou, E., Gürol, S., Gratton, S., Kasanickỳ, I.: Hybrid Levenberg–Marquardt and weak-constraint ensemble Kalman smoother method Nonlin. Processes Geophys. **23**, 59–73 (2016)

[12] Tang, L.M.: A regularization homotopy iterative method for ill-posed nonlinear least squares problem and its application. In: Advances in Civil Engineering, ICCET 2011, Applied Mechanics and Materials, vol. 90, pp. 3268–3273. Trans Tech Publications (2011). URL `http://www.scientific.net/AMM.90-93.3268`

[13] Dennis Jr., J.E.: Nonlinear least squares and equations, pp. 269–312. The State of the Art in Numerical Analysis. Academic Press, London (1977)

[14] Li, D., Fukushima, M., Qi, L., Yamashita, N.: Regularized Newton methods for convex minimization problems with singular solutions. Comput. Optim. Appl. **28**(2), 131–147 (2004)

[15] Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. **6**(5), 927–940 (2012)

[16] Behling, R., Iusem, A.: The effect of calmness on the solution set of systems of nonlinear equations. Math. Program. **137**(1-2), 155–165 (2013)

[17] Dan, H., Yamashita, N., Fukushima, M.: Convergence properties of the inexact Levenberg-Marquardt method under local error bound conditions. Optim. Methods Softw. **17**(4), 605–626 (2002)

[18] Fan, J.: Convergence rate of the trust region method for nonlinear equations under local error bound condition. Comput. Optim. Appl. **34**(2), 215–227 (2006)

[19] Fan, J., Pan, J.: Convergence properties of a self-adptive Levenberg-Marquardt algorithm algorithm under local error bound condition. Comput. Optim. Appl. **34**(1), 723–751 (2016)

[20] Karas, E.W., Santos, S.A., Svaiter, B.F.: Algebraic rules for computing the regularization parameter of the Levenberg-Marquardt method. Comput. Optim. Appl. **65**(3), 723–751 (2016)

[21] Ipsen, I.C.F., Kelley, C.T., Poppe, S.R.: Rank-deficient nonlinear least squares problems and subset selection. SIAM J. Numer. Anal. **49**(3), 1244–1266 (2011)

[22] Bellavia, S., Cartis, C., Gould, N.I.M., Morini, B., Toint, P.L.: Convergence of a regularized euclidean residual algorithm for nonlinear least-squares. SIAM J. Numer. Anal. **48**(1), 1–29 (2010)

[23] Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**(1), 91–124 (2002)

[24] Lawson, C.L., Hanson, R.J.: Solving Least Squares Problems, *Classics in Applied Mathematics*, vol. 15. SIAM, Philadelphia (1995)

[25] Facchinei, F., Fischer, A., Herrich, M.: A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. Math. Methods Oper. Res. **77**, 433–443 (2013)

[26] Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**(1-2), 1–36 (2014)

[27] Martínez, M.A., Fernández, D.: A quasi-Newton modified LP-Newton method. Optim. Methods Softw. (published online) (2017). URL `https://doi.org/10.1080/10556788.2017.1384955`