

Global and Local Convergence of a Levenberg-Marquadt Algorithm for Inverse Problems

E. Bergou * Y. Diouane[†] V. Kungurtsev[‡]

May 11, 2017

Abstract

The Levenberg-Marquardt algorithm is one of the most popular algorithms for the solution of nonlinear least squares problems. In this paper, we propose and analyze the global and local convergence results of a novel Levenberg-Marquardt method for solving general nonlinear least squares problems. The proposed algorithm enjoys strong convergence properties (global convergence as well as quadratic local convergence) for least squares problems which do not necessarily have a zero residual solution, all without any additional globalization strategy. Preliminary numerical experiments confirm the theoretical behavior of our proposed algorithm.

Keywords: Nonlinear least squares problem, inverse problems, Levenberg-Marquardt method, global and local convergence, quadratic convergence.

1 Introduction

In this paper we consider the general nonlinear least squares problem

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

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a (deterministic) vector-valued function, assumed continuously differentiable. We do not assume that there is a solution with zero residual, or that we seek such a solution. In fact, problems of this nature arise in several important practical contexts. One example is inverse problems [16] (e.g., data assimilation [4, 17], full-waveform inversion [18]), where typically an ill-posed nonlinear continuous problem is solved through a discretization. Other examples appear in parameter estimation when a mathematical model approximating a true distribution is fit to given (noisy) data [16, 18]. In all these cases, the resulting least squares problems do not necessarily have a zero residual at any point.

*MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France (elhoucine.bergou@inra.fr).

[†]Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), Université de Toulouse, 31055 Toulouse Cedex 4, France (youssef.diouane@isae.fr).

[‡]Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. Support for this author was provided by the Czech Science Foundation project 17-26999S (vyacheslav.kungurtsev@fel.cvut.cz).

Recall that the Gauss-Newton method is an iterative procedure for solving (1) and where at each iterate x_j a step is computed as a solution to the linearized least squares subproblem

$$\min_{s \in \mathbb{R}^n} \frac{1}{2} \|F_j + J_j s\|^2,$$

where $F_j = F(x_j)$ and $J_j = J(x_j)$ denotes the Jacobian of F at x_j . The subproblem has a unique solution if J_j has full column rank, and in that case the step is a descent direction for f .

The Levenberg-Marquardt method [11, 12] (see also [15]) was developed to deal with the rank deficiency of J_j and to provide a globalization strategy for Gauss-Newton. At each iteration a step is considered of the form $-(J_j^\top J_j + \gamma_j I)^{-1} J_j^\top F_j$, corresponding to the unique solution of

$$\min_{s \in \mathbb{R}^n} m_j(x_j + s) = \frac{1}{2} \|F_j + J_j s\|^2 + \frac{1}{2} \gamma_j \|s\|^2, \quad (2)$$

where $\gamma_j > 0$ is an appropriately chosen regularization parameter.

In this paper, we will present and analyze the global and local convergence results of a novel Levenberg-Marquardt method for solving general nonlinear least squares problems. In particular we present a Levenberg-Marquardt updating strategy that carefully balances the opposing objectives of ensuring global convergence and stabilizing a Newton local convergence regime.

This is a novel contribution in two senses. First, to our knowledge, it is the first local convergence result for a Levenberg-Marquardt method for problems with non-zero residual. The strongest results for local convergence of Levenberg-Marquardt are given in a series of papers beginning with [19] (such as [9] and [5], see also [6]), wherein it is assumed that the solution satisfies $F(x) = 0$. In the case of non-zero residuals, it has been found that the standard implementations of Levenberg-Marquardt converge locally at a linear rate, and even then only if the parameter γ_j goes to zero [10]. Second, in general the goals of encouraging global and local convergence compete against each other. Namely, the regularization parameter appearing in the subproblem should be allowed to become arbitrarily large in order to encourage global convergence, to ensure the local accuracy of the linearized subproblem, but the parameter must approach zero in order to function as a stabilizing regularization that encourages stable local convergence.

For example, in the original presentation of the Levenberg-Marquardt method in [11, 12], γ_j is not permitted to go to zero, and only global convergence is considered. By contrast, in [19], for instance, there is a two-phase method where quadratic decline in the residual is tested with each step that is otherwise globalized by a line-search procedure. Two phase methods, while also less mathematically elegant, are practically inefficient and challenging to implement in the sense that it can be difficult to properly ascertain when the region of local convergence is reached.

Our parameter updating strategy is inspired by [8], which presents a Levenberg-Marquardt method inspired from trust-region algorithm for zero residual least squares problems. The trust-region radius is updated as $\Delta_{j+1} = \mu \|F_{j+1}\|^2$, with μ updated according to classical global convergence updating strategies, and the residual is included to enforce local convergence. They show global and superlinear local convergence properties for their method. We extend the results outlined above in scope in the sense of showing the strong convergence properties for residual problems which do not necessarily have a zero residual solution, as well as in elegance in that the method is purely a Levenberg-Marquardt method, with no additional globalization strategies, and thus is seamless and is an extension of the standard classical approach to least-squares problems.

The outline of this paper is as follows. In Section 2 we present the proposed Levenberg-Marquardt algorithm for solving general nonlinear least squares problems. Section 3 addresses the inexact solution of the linearized least squares subproblems arising within the Levenberg-Marquardt method. In Section 4, we show the global convergence of our algorithm. In Section 5 we derive the overall local convergence analysis of the proposed algorithm. In Section 6, preliminary numerical experiments with basic implementations are presented that show the good behavior of our novel algorithm. Finally, in Section 7 we draw some perspectives and conclusions.

Throughout this paper $\|\cdot\|$ will denote the vector or matrix l_2 -norm.

2 A novel Levenberg-Marquardt algorithm

In deciding whether to accept a step s_j generated by the subproblem (2), the Levenberg-Marquardt method can be seen as precursor of the trust-region method [3]. In fact, it seeks to determine when the Gauss-Newton step is applicable (in which case the regularization parameter is set to zero) or when it should be replaced by a slower but safer steepest descent step (corresponding to a sufficiently large regularization parameter). For that purpose, one considers the ratio between the actual reduction $f(x_j) - f(x_j + s_j)$ attained in the objective function and the reduction $m_j(x_j) - m_j(x_j + s_j)$ predicted by the model, given by

$$\rho_j = \frac{f(x_j) - f(x_j + s_j)}{m_j(x_j) - m_j(x_j + s_j)}.$$

Then, if ρ_j is sufficiently above zero, the step is accepted and γ_j is possibly decreased. Otherwise the step is rejected and γ_j is increased.

In this paper we consider the choice of the regularization parameter as $\gamma_j = \mu \|\nabla f(x_j)\|^2$, where μ is updated according to the ratio ρ_j . The considered Levenberg-Marquardt algorithm is described below.

Algorithm 1: Levenberg-Marquardt algorithm.

Initialization

Choose the constants $\eta \in (0, 1)$, $\mu_{\min} > 0$ and $\lambda > 1$. Select x_0 and $\mu_0 \geq \mu_{\min}$. Set $\gamma_0 = \mu_0 \|\nabla f(x_0)\|^2$ and $\bar{\mu} = \mu_0$.

For $j = 0, 1, 2, \dots$

1. Solve (or approximately solve) (2), and let s_j denote such a solution.
 2. Compute $\rho_j = \frac{f(x_j) - f(x_j + s_j)}{m_j(x_j) - m_j(x_j + s_j)}$.
 3. If $\rho_j \geq \eta$, then set $x_{j+1} = x_j + s_j$ and $\mu_{j+1} \in [\max(\mu_{\min}, \bar{\mu}/\lambda), \bar{\mu}]$ and $\bar{\mu} = \mu_{j+1}$.
Otherwise, set $x_{j+1} = x_j$ and $\mu_{j+1} = \lambda\mu_j$.
 4. Compute $\gamma_{j+1} = \mu_{j+1} \|\nabla f(x_{j+1})\|^2$.
-

A brief remark is warranted regarding the new step acceptance criteria. Note that we have an auxiliary parameter $\bar{\mu}$, that represents the last good parameter. This is necessary in order to

balance the requirements of global and local convergence. If the model is relatively inaccurate, then μ_j is driven higher, however, when we reach a region of local convergence, we need the parameter γ_j to encourage local convergence, and thus bound μ_j so that the component $\|\nabla f(x)\|$ dominates the behavior of the parameter γ_j . It can be seen that the convergence theory does not change with some heuristic modifications of the procedure to prevent an unnecessarily large number of iterations out of an inaccurate approximation far from the solution to increase μ after a lucky step that brings it back down to $\bar{\mu}$, e.g., by saving the last high value of μ_j upon a successful step, and coming back to it if the next step is indeed poor.

3 Inexact solution of the linearized subproblems

Step 1 of Algorithm 1 requires the approximate solution of subproblem (2). As in trust-regions methods, there are different techniques to approximate the solution of this subproblem yielding a globally convergent step. For the purposes of global convergence it is sufficient to compute a step s_j that provides a reduction in the model as good as the one produced by the so-called Cauchy step (defined as the minimizer of the model along the negative gradient).

The Cauchy step is defined by minimizing $m_j(x_j - t\nabla f(x_j))$ when $t > 0$ and is given by

$$s_j^c = -\frac{\|\nabla f(x_j)\|^2}{\nabla f(x_j)^\top (J_j^\top J_j + \gamma_j I) \nabla f(x_j)} \nabla f(x_j). \quad (3)$$

The corresponding Cauchy decrease of the model is

$$m_j(x_j) - m_j(x_j + s_j^c) = \frac{1}{2} \frac{\|\nabla f(x_j)\|^4}{\nabla f(x_j)^\top (J_j^\top J_j + \gamma_j I) \nabla f(x_j)}.$$

Since $\nabla f(x_j)^\top (J_j^\top J_j + \gamma_j I) \nabla f(x_j) \leq \|\nabla f(x_j)\|^2 (\|J_j\|^2 + \gamma_j)$, we conclude that

$$m_j(x_j) - m_j(x_j + s_j^c) \geq \frac{1}{2} \frac{\|\nabla f(x_j)\|^2}{\|J_j\|^2 + \gamma_j}.$$

The Cauchy step (3) is cheap to calculate as it does not require any system solve. Moreover, the Levenberg-Marquadt method will be globally convergent if it uses a step that attains a reduction in the model as good as a multiple of the Cauchy decrease. Thus we will impose the following assumption on the step calculation:

Assumption 3.1 *For every step j ,*

$$m_j(x_j) - m_j(x_j + s_j) \geq \frac{\theta_{fcd}}{2} \frac{\|\nabla f(x_j)\|^2}{\|J_j\|^2 + \gamma_j}$$

for some constant $\theta_{fcd} > 0$.

Despite providing a sufficient reduction in the model and being cheap to compute, the Cauchy step is a particular form of steepest descent. In practice, a version of Algorithm 1 solely based on the Cauchy step would suffer from the same drawbacks as the steepest descent algorithm on ill-conditioned problems. One can see that the Cauchy step depends on $J_j^\top J_j$ only in the step length. Faster convergence can be expected if the matrix $J_j^\top J_j$ also influences the step direction.

Since the Cauchy step is the first step of the conjugate gradient method (CG) when applied to the minimization of the quadratic $s \rightarrow m_j(x_j + s)$, it is natural to propose to run the CG further and stop only when the residual becomes relatively small. The truncated-CG step is of the form:

$$s_j^{\text{CG}} = V_j \left(V_j^\top (J_j^\top J_j + \gamma_j I) V_j \right)^{-1} V_j^\top \nabla f(x_j), \quad (4)$$

where V_j is a given unitary matrix whose first column is given by $-\nabla f(x_j)/\|\nabla f(x_j)\|$.

Since the CG generates iterates by minimizing the quadratic model over nested Krylov subspaces, and the first subspace is the one generated by $\nabla f(x_j)$ (see, e.g., [14, Theorem 5.2]), the decrease attained at the first CG iteration (i.e., by the Cauchy step) is kept by the remaining iterations. Thus Assumption 3.1 holds for all the iterates s_j^{CG} generated by the truncated-CG.

The following Lemma is similar to [1, Lemma 5.1] and will be useful for our global convergence analysis.

Lemma 3.1 *For the three steps proposed (exact, Cauchy, and truncated-CG), one has that*

$$\|s_j\| \leq \frac{\|\nabla f(x_j)\|}{\gamma_j} = \frac{1}{\mu_j \|\nabla f(x_j)\|}$$

and

$$|s_j^\top (\gamma_j s_j + \nabla f(x_j))| \leq \frac{\|J_j\|^2 \|\nabla f(x_j)\|^2}{\gamma_j^2} = \frac{\|J_j\|^2}{\mu_j^2 \|\nabla f(x_j)\|^2}.$$

Proof. We will omit the indices j in the proof. We note that the truncated-CG step can be seen as a generalized step of both exact and Cauchy steps. In fact, the first CG iteration produces the Cauchy step while the last iteration gives the exact one.

Thus, without loss of generality, for the three proposed steps there exists a unitary matrix V with first column given by $-\nabla f(x)/\|\nabla f(x)\|$ and such that

$$s = V \left(V^\top (J^\top J + \gamma I) V \right)^{-1} V^\top \nabla f(x) = V \left(V^\top J^\top J V + \gamma I \right)^{-1} \|\nabla f(x)\| e_1,$$

where e_1 is the first vector of the canonical basis of \mathbb{R}^n . From the positive semidefiniteness of $V^\top J^\top J V$, we immediately obtain $\|s\| \leq \|\nabla f(x)\|/\gamma$.

To prove the second inequality of Lemma 3.1, we apply the Sherman–Morrisson–Woodbury formula and obtain

$$s = V \left(\frac{1}{\gamma} I - \frac{1}{\gamma^2} (JV)^\top \left(I + \frac{(JV)(JV)^\top}{\gamma} \right)^{-1} (JV) \right) \|\nabla f(x)\| e_1.$$

Since $V e_1 = -\nabla f(x)/\|\nabla f(x)\|$,

$$\gamma s + \nabla f(x) = -\frac{1}{\gamma} V (JV)^\top \left(I + \frac{(JV)(JV)^\top}{\gamma} \right)^{-1} (JV) \|\nabla f(x)\| e_1.$$

Now, from the fact that $(JV)(JV)^\top/\gamma$ is positive semidefinite, the norm of the inverse of $I + (JV)(JV)^\top/\gamma$ is less than one, and thus (since V is unitary)

$$\|\gamma s + \nabla f(x)\| \leq \frac{\|J\|^2 \|\nabla f(x)\|}{\gamma}.$$

Finally,

$$|s^\top(\gamma s + \nabla f(x))| \leq \|s\| \|\gamma s + \nabla f(x)\| \leq \frac{\|J\|^2 \|\nabla f(x)\|^2}{\gamma^2}.$$

■

4 Global convergence

We start by giving some classical assumptions and then state and prove some lemmas that later will appear in the global convergence analysis.

Assumption 4.1 *The function f is continuously differentiable in an open set containing $L(x_0) = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ with Lipschitz continuous gradient on $L(x_0)$ and corresponding constant $\nu > 0$.*

Assumption 4.2 *The function f gradient is uniformly bounded, i.e., there exists $\kappa_g > 0$ such that $\|\nabla f(x_j)\| \leq \kappa_g$ for all j .*

Assumption 4.3 *The function F Jacobian is uniformly bounded, i.e., there exists $\kappa_J > 0$ such that $\|J_j\| \leq \kappa_J$ for all j .*

The next lemma says that, if we suppose that the gradient norm is bounded below by a non zero constant g_{\min} , then for a value of the parameter μ_j sufficiently large, the step is accepted by the acceptance criterion.

Lemma 4.1 *Let Assumptions 3.1, 4.1, 4.2, and 4.3 hold. Suppose that, for all the iterations j , there exists a bound $g_{\min} > 0$ such that $\|\nabla f(x_j)\| \geq g_{\min}$. Then, one has*

$$\lim_{\mu_j \rightarrow \infty} \rho_j = 2.$$

Proof. By applying a Taylor expansion, one has

$$\begin{aligned} 1 - \frac{\rho_j}{2} &= \frac{m(x_j) - f(x_j) + f(x_j + s_j) - m(x_j + s_j) + m(x_j) - m(x_j + s_j)}{2[m(x_j) - m(x_j + s_j)]} \\ &= \frac{R - s_j^\top \nabla f(x_j) - s_j^\top (J_j^\top J_j + \gamma_j I) s_j}{2[m(x_j) - m(x_j + s_j)]} \\ &= \frac{R - s_j^\top (J_j^\top J_j) s_j - s_j^\top (\gamma_j s_j + \nabla f(x_j))}{2[m(x_j) - m(x_j + s_j)]} \quad \text{where } R \leq \frac{\nu}{2} \|s\|^2. \end{aligned}$$

Then, using Lemma 3.1, Assumptions 3.1 and 4.1, one gets

$$\begin{aligned}
\left|1 - \frac{\rho_j}{2}\right| &\leq \frac{\frac{\nu}{2}\|s_j\|^2 + \|J_j\|^2\|s_j\|^2 + |s_j^\top(\gamma_j s_j + \nabla f(x_j))|}{\frac{\theta_{fcd}\|\nabla f(x_j)\|^2}{\|J_j\|^2 + \gamma_j}} \\
&\leq \frac{\left(\frac{\nu}{2} + 2\kappa_j^2\right) \kappa_j^2 + \gamma_j}{\theta_{fcd} \gamma_j^2} \\
&\leq \frac{\left(\frac{\nu}{2} + 2\kappa_j^2\right) \kappa_j^2 + \mu_j \|\nabla f(x_j)\|^2}{\theta_{fcd} \mu_j^2 \|\nabla f(x_j)\|^4} \\
&\leq \frac{\left(\frac{\nu}{2} + 2\kappa_j^2\right) \kappa_j^2 + \mu_j \kappa_g^2}{\theta_{fcd} \mu_j^2 g_{\min}^4}.
\end{aligned}$$

■

Now, we can show our main global convergence result

Theorem 4.1 *Under Assumptions 3.1, 4.1, 4.2, and 4.3, the sequence $\{x_j\}$ generated by Algorithm 1 satisfies*

$$\liminf_{j \rightarrow \infty} \|\nabla f(x_j)\| = 0.$$

Proof. By contradiction, if the theorem is not true, then there exists a bound $g_{\min} > 0$ such that

$$\|\nabla f(x_j)\| \geq g_{\min}, \quad \forall j \geq 0.$$

Define $\mathbb{S} = \{j \in \mathbb{N} \mid \rho_j \geq \eta\}$ as the set of successful iterations. Hence for $j \in \mathbb{S}$, one has

$$\begin{aligned}
\|F_j\|^2 - \|F_{j+1}\|^2 &\geq \eta(m_j(x_j) - m_j(x_j + s_j)) \\
&\geq \eta \frac{\theta_{fcd} \|\nabla f(x_j)\|^2}{2 (\|J_j\|^2 + \gamma_j)} \quad (\text{using Assumption 3.1}) \\
&\geq \frac{\eta \theta_{fcd}}{2} \frac{\|\nabla f(x_j)\|^2}{\kappa_j^2 + \mu_j \|\nabla f(x_j)\|^2} \\
&\geq \frac{\eta \theta_{fcd}}{2} \frac{g_{\min}^2}{\kappa_j^2 + \mu_j \kappa_g^2}
\end{aligned}$$

If \mathbb{S} is infinite, then since $\sum_{j \in \mathbb{S}} \|F_j\|^2 - \|F_{j+1}\|^2$ is finite, we deduce that

$$\lim_{j \rightarrow \infty} \frac{\eta \theta_{fcd}}{2} \frac{g_{\min}^2}{\kappa_j^2 + \mu_j \kappa_g^2} = 0,$$

hence $\lim_{j \rightarrow \infty} \mu_j = +\infty$.

Otherwise (i.e., \mathbb{S} is finite), from Algorithm 1 we have $\mu_{j+1} = \lambda \mu_j$ for all sufficiently large j . Since $\lambda > 1$, we deduce that $\lim_{j \rightarrow \infty} \mu_j = +\infty$.

Thus, using Lemma 4.1, we have $\lim_{j \rightarrow \infty} \rho_j = 2$. Since when $\rho_j \geq \eta$ we decrease μ_j then, there exists a positive constant μ_{\max} such that $\mu_j \leq \mu_{\max}$ holds for all sufficiently large j . Which leads to a contradiction with the fact that μ_j goes to infinity when j goes to $+\infty$. ■

5 Local convergence

In this section we present the local convergence analysis of Algorithm 1. In particular, we show a quadratic convergence rate without assuming the nonsingularity of $J(x)$. As mentioned in Section 1, the state of the art for problems with zero residual at the solution is that local convergence at a quadratic rate holds under a Lipschitz continuity and error bound assumptions. We will replicate this in a manner appropriate for problems without a zero residual at the solution. In the sequel of this section, the considered step is the exact solution of subproblem (2).

5.1 Assumptions

In this setting, $\|F(x)\|$ is no longer an appropriate measure for the distance to the solution. Stationarity is associated with a zero gradient, and, as can be gleaned from the form of our update for the regularization parameter γ_j in Algorithm 1, this is what we use for the regularization to encourage fast convergence.

In the case of zero-residual problems, there is a set of solutions to $F(x) = 0$ and the purpose of the algorithm is to obtain a point at which the residual is zero. In this case, we seek a stationary solution where $\nabla f(x) = 0$, however there can be multiple sets of stationary points, with varying objective values. As the behavior of the algorithm is such that both descent of $f(x)$ encouraged as well as a solution to stationarity is sought, for a clear picture of the convergence, we instead propose to consider a particular subset with a constant value of the objective.

Assumption 5.1 *There exists a connected isolated set X^* composed of stationary points to (1), and Algorithm 1 generates a sequence with an accumulation point $x^* \in X^*$.*

We shall denote by \bar{F} the value of $\bar{F} = \|F(\bar{x})\|$ at any $\bar{x} \in X^*$. Note that this is unique, as X^* is a connected set of stationary points, so there is no direction of ascent for $f(x)$ among the set of directions feasible within X^* .

Henceforth, from the global convergence analysis, we can assume, without loss of generality, that there exists a subsequence approaching this X^* . This subsequence need not be unique, i.e., there may be more than one subsequence converging to separate connected sets of stationary points. We shall see that eventually, one of these sets shall "catch" the subsequence and result in direct convergence to the solution set at a quadratic rate.

In the sequel $N(x, \delta)$ denotes the closed ball of center x (a given vector) and radii $\delta > 0$. $\text{dist}(x, X^*)$ denotes the distance between the vector x and the set X^* , $\text{dist}(x, X^*) = \min_{y \in X^*} \|x - y\|$.

Assumption 5.2 *It holds that $F(x)$ and $J(x)$ are both locally Lipschitz continuous around $x^* \in X^*$ with x^* satisfying Assumption 5.1. In particular this implies, letting $\bar{x} = \text{argmin}_{y \in X^*} \|x - y\|$, that there exists $\delta_1 > 0$ such that for $x \in N(x^*, \delta_1)$,*

$$\|\nabla f(x)\|^2 = \|F(x)J(x)\|^2 = \|F(x)J(x) - F(\bar{x})J(\bar{x})\|^2 \leq L_1 \text{dist}(x, X^*)^2, \quad (5)$$

$$\|F(x) - F(\bar{x})\| \leq L_2 \text{dist}(x, X^*), \quad (6)$$

and that for all $x, y \in N(x^*, \delta)$,

$$\|F(y) - F(x) - J(x)(y - x)\| \leq L_3 \|y - x\|^2, \quad (7)$$

where L_1, L_2 , and L_3 are positive constants.

From the triangle inequality and assuming (6), we get

$$\|F(x)\| - \bar{F} \leq \|F(x) - F(\bar{x})\| \leq L_2 \text{dist}(x, X^*). \quad (8)$$

In addition, as it would be standard for unconstrained optimization, the gradient gives an error bound for the distance to the solution,

Assumption 5.3 *There exists a $\delta_2 > 0$ and $M_1 > 0$ such that for $x \in N(x^*, \delta_2)$,*

$$\text{dist}(x, X^*) \leq M_1 \|\nabla f(x)\|.$$

However, this is insufficient for local convergence analysis. If the gradient is the only error bound, then in order to ensure local convergence, an accurate second derivative approximation must be available to quadratically reduce the error bound. This can be seen from the careful analysis in [7, 6] of local convergence for a family of Newton methods for problems with nonisolated solutions. Since the Levenberg-Marquadt method is based on the Gauss-Newton approximation eschewing calculation of the second derivatives of $F(x)$, this suggests that the gradient would be unsatisfactory.

Thus we introduce an additional assumption,

Assumption 5.4 *There exists a $\delta_3 > 0$ and $M_2 > 0$ such that for $x \in N(x^*, \delta_3)$,*

$$\text{dist}(x, X^*) \leq M_2 (\|F(x)\| - \bar{F}).$$

Note the resemblance to the zero residual case [19, 9, 5, 6]. If $\bar{F} = 0$ then the usual error bound assumption is recovered. In addition, locally, nonsingularity of $J(x)$ or standard second order sufficient optimality conditions imply this error bound.

5.2 Convergence Proof

From the global convergence results, we have established that there is a subsequence of successful iterations converging to a solution set X^* . In this section, we begin by considering the subsequence of iterations that succeed the successful iterations, i.e., we consider the subsequence $\mathcal{K} = \{j + 1 : j \in \mathbb{S}\}$. We shall present the results with a slight abuse of notation that simplifies the presentation without sacrificing accuracy or generality: in particular every time we denote a quantity a_j , the index j corresponds to an element of this subsequence \mathcal{K} denoted above, thus when we say a particular statement holds eventually, this means that it holds for all $j \in \mathbb{S} + 1$ with j sufficiently large.

We shall denote $\hat{\mu}$ as an upper bound for $\bar{\mu}$. Note that this exists for $\{\mu_j\}_{j \in \mathcal{K}}$ by the formulation of Algorithm 1. In addition we shall denote δ as $\delta = \min(\delta_1, \delta_2, \delta_3)$, with $\{\delta_i\}_{i=1,2,3}$ defined in the Assumptions.

In the proof we follow the structure of the local convergence proof in [19], with the additional point that the step is accepted by the globalization procedure. We define,

$$\bar{x}_j = \operatorname{argmin}_{y \in X^*} \|x_j - y\|.$$

This Lemma is similar to [19, Lemma 2.1].

Lemma 5.1 *Suppose Assumptions 5.2, 5.3, and 5.4.*

If $x_j \in N(x^, \frac{\delta}{2})$, then the solution s_j to (2) satisfies,*

$$\|J_j s_j + F_j\| - \bar{F} \leq C_1 \text{dist}(x_j, X^*)^2, \quad (9)$$

where C_1 is a positive constant independent of j .

Proof. We have,

$$\begin{aligned} \|J_j s_j + F_j\|^2 &\leq 2m_j(s_j) \leq 2m_j(\bar{x}_j - x_j) = \|J_j(\bar{x}_j - x_j) + F_j\|^2 + \gamma_j \|\bar{x}_j - x_j\|^2 \\ &\leq \|J_j(\bar{x}_j - x_j) + F_j\|^2 + \mu_j L_1 \|x_j - \bar{x}_j\|^4 \quad (\text{using Assumption 5.2}) \\ &\stackrel{(a)}{\leq} \left((\bar{F} + L_3 \|x_j - \bar{x}_j\|^2)^2 + \mu_j L_1 \|x_j - \bar{x}_j\|^4 \right) \\ &= \left((L_3^2 + \mu_j L_1) \|x_j - \bar{x}_j\|^4 + 2L_3 \bar{F} \|x_j - \bar{x}_j\|^2 + \bar{F}^2 \right) \\ &\leq \left((L_3^2 + \mu_j L_1) \|x_j - \bar{x}_j\|^4 + 2L_3 \sqrt{\frac{L_3^2 + \mu_j L_1}{L_3^2}} \bar{F} \|x_j - \bar{x}_j\|^2 + \bar{F}^2 \right) \\ &= \left(\sqrt{L_3^2 + \hat{\mu} L_1} \|x_j - \bar{x}_j\|^2 + \bar{F} \right)^2 \end{aligned}$$

where (a) is from the triangle inequality as well as Assumption 5.2, i.e.,

$$\begin{aligned} \|J_j(\bar{x}_j - x_j) + F_j\| &\leq \bar{F} + \|J_j(\bar{x}_j - x_j) + F_j - F(\bar{x})\| \\ &\leq \bar{F} + L_3 \|x_j - \bar{x}_j\|^2. \end{aligned}$$

■

Lemma 5.2 *Suppose Assumptions 5.2, 5.3, and 5.4.*

If $x_j \in N(x^, \frac{\delta}{2})$, then the solution s_j to (2) satisfies,*

$$\|s_j\| \leq C_2 \text{dist}(x_j, X^*), \quad (10)$$

where C_2 is a positive constant independent of j .

Proof. The solution of the classical Levenberg-Marquadt subproblem, for zero-residual problems proposed in [19], when solving $F(x) - F(\bar{x}) = 0$, satisfies

$$\begin{aligned} &\text{argmin}_s \frac{1}{2} \|F_j - F(\bar{x}) + J_j s\|^2 + \frac{1}{2} \mu_j \|F_j - F(\bar{x})\|^2 \|s\|^2, \\ &= \text{argmin}_s \frac{1}{2} \left(2F_j^T J_j s - 2F(\bar{x})^T J_j s + s^T J_j^T J_j s + \mu_j \|F_j - F(\bar{x})\|^2 s^T s \right) \end{aligned}$$

From [19, Lemma 2.1] it holds that the solution to this problem, \hat{s}_j , satisfies $\|\hat{s}_j\| \leq \hat{C} \text{dist}(x_j, X^*)$, where \hat{C} is a positive constant independent of j .

Now define the function

$$G(s, u^{(1)}, u^{(2)}) = \frac{1}{2} \left(2F_j^T J_j s + (u^{(1)})^T s + s^T J_j^T J_j s + u^{(2)} s^T s \right). \quad (11)$$

We may consider the model in (2) as a perturbation of (11) if we set in the expression of G , $u^{(1)}$ and $u^{(2)}$ to $u_0^{(1)} = 0$ and $u_0^{(2)} = \gamma_j$, respectively. The classical Levenberg-Marquadt model is the

same as (11) with $u_1^{(1)} = -2F(\bar{x})^T J_j$ and $u_1^{(2)} = \mu_j \|F_j - F(\bar{x})\|^2$. In other words, if we change in the expression of G , $u_1^{(1)}$ and $u_1^{(2)}$ by $u_0^{(1)}$ and $u_0^{(2)}$, respectively, we modify the model from the classical Levenberg-Marquadt subproblem for zero residual problems to the one defined in (2).

Since the function G is quadratic with respect to s and linear with respect to u , the conditions of [2, Proposition 4.36] are satisfied, and thus,

$$\begin{aligned}
\|s_j - \hat{s}_j\| &\leq \hat{C}_2 \|u_1^{(1)} - u_0^{(1)}\| + \hat{C}_3 \|u_1^{(2)} - u_0^{(2)}\| \\
&\leq 2\hat{C}_2 \|F(\bar{x})^T J_j\| + \hat{C}_3 \mu_j \| \|F_j - F(\bar{x})\|^2 - \|F_j J_j\|^2 \| \\
&\leq 2\hat{C}_2 (\|F_j J_j\| + \|F_j - F(\bar{x})\| \|J_j\|) + \hat{C}_3 \max(L_1, L_2^2) \hat{\mu} \text{dist}(x_j, X^*)^2 \\
&\leq 4\hat{C}_2 \max(\sqrt{L_1}, B_J L_2) \text{dist}(x_j, X^*) + \hat{C}_3 \max(L_1, L_2^2) \hat{\mu} \text{dist}(x_j, X^*)^2 \\
&\leq \hat{C}_4 \text{dist}(x_j, X^*),
\end{aligned}$$

where \hat{C}_2 , \hat{C}_3 and \hat{C}_4 are positive constants independent of j . The second inequality follows from Assumption 5.2, the triangle and Cauchy-Schwartz inequalities, and the final inequality again follows from Assumption 5.2 and the boundedness of J_j by B_J (due to the convergence of x_j).

Thus, from the triangle inequality,

$$\|s_j\| \leq \|\hat{s}_j\| + \|s_j - \hat{s}_j\| \leq (\hat{C} + \hat{C}_4) \text{dist}(x_j, X^*).$$

■

Lemma 5.3 *For j sufficiently large, $\rho_j \geq \eta$.*

Proof. Consider first the case of $\bar{F} > 0$.

It holds that,

$$\begin{aligned}
m_j(x_j) - m_j(x_j + s_j) &= \|F_j\|^2 - \|F_j + J_j s_j\|^2 - \gamma_j \|s_j\|^2 \\
&\stackrel{(a)}{\geq} \|F_j\|^2 - \|F_j + J_j(\bar{x}_j - x_j)\|^2 - \gamma_j \|\bar{x}_j - x_j\|^2 \\
&= (\|F_j\| + \|F_j + J_j(\bar{x}_j - x_j)\|) (\|F_j\| - \|F_j + J_j(\bar{x}_j - x_j)\|) - \gamma_j \|\bar{x}_j - x_j\|^2 \\
&\geq (\|F_j\| + \|F_j + J_j(\bar{x}_j - x_j)\|) (\|F_j\| - \bar{F} - L_3 \|\bar{x}_j - x_j\|^2) - \gamma_j \|\bar{x}_j - x_j\|^2 \\
&\stackrel{(b)}{\geq} \|F_j\| (L_2 \|\bar{x}_j - x_j\| - L_3 \|\bar{x}_j - x_j\|^2) - \gamma_j \|\bar{x}_j - x_j\|^2 \\
&\geq \bar{F} L_2 \|\bar{x}_j - x_j\| - (F_0 L_3 + \gamma_j) \|\bar{x}_j - x_j\|^2 \\
&= \bar{F} L_2 \|\bar{x}_j - x_j\| - (\tilde{L}_3 + \gamma_j) \|\bar{x}_j - x_j\|^2,
\end{aligned}$$

where (a) arises from the optimality of s_j for m_j , and for (b) we note that for j sufficiently large $\|\bar{x}_j - x_j\|^2 \ll \|\bar{x}_j - x_j\|$ and thus $L_2 \|\bar{x}_j - x_j\| - L_3 \|\bar{x}_j - x_j\|^2 \geq 0$.

Now we write,

$$\begin{aligned}
|1 - \rho_j| &= \left| \frac{m(x_j) - f(x_j) + f(x_j + s_j) - m(x_j + s_j)}{m(x_j) - m(x_j + s_j)} \right| \\
&= \left| \frac{\|F(x_j + s_j)\|^2 - \|F_j + J_j s_j\|^2 - \gamma_j \|s_j\|^2}{m(x_j) - m(x_j + s_j)} \right| \\
&= \left| \frac{(\|F(x_j + s_j)\| - \|F_j + J_j s_j\|)(\|F(x_j + s_j)\| + \|F_j + J_j s_j\|) - \gamma_j \|s_j\|^2}{m(x_j) - m(x_j + s_j)} \right| \\
&\leq \frac{L_3 \|s_j\|^2 (\|F_j\| + \|J_j\| \|s_j\|) + \gamma_j \|s_j\|^2}{\bar{F} L_2 \|\bar{x}_j - x_j\| - (\tilde{L}_3 + \gamma_j) \|\bar{x}_j - x_j\|^2} \\
&\rightarrow 0 \quad \text{when } j \text{ goes to } +\infty.
\end{aligned}$$

The last limit is zero because $\|s_j\| = O(\|x_j - \bar{x}_j\|) \rightarrow 0$ by Lemma 5.2, γ_j is bounded, and $\bar{F} > 0$.

Now, if $\bar{F} = 0$, we have, from the same derivation,

$$m_j(x_j) - m_j(x_j + s_j) \geq \|F_j\| L_2 \|\bar{x}_j - x_j\| - (\tilde{L}_3 + \gamma_j) \|\bar{x}_j - x_j\|^2,$$

and, using Assumptions 5.4 and 5.2 as well as Lemma 5.2,

$$\begin{aligned}
|1 - \rho_j| &\leq \frac{L_3 \|s_j\|^2 (\|F_j\| + \|J_j\| \|s_j\|) + \gamma_j \|s_j\|^2}{\|F_j\| L_2 \|\bar{x}_j - x_j\| - (\tilde{L}_3 + \gamma_j) \|\bar{x}_j - x_j\|^2} \\
&\leq \frac{C_5 \|x_j - \bar{x}_j\|^2 (\|x_j - \bar{x}_j\| + \|J_j\| \|x_j - \bar{x}_j\|) + \gamma_j \|x_j - \bar{x}_j\|^2}{\tilde{L}_2 \|x_j - \bar{x}_j\| \|\bar{x}_j - x_j\| - (\tilde{L}_3 + \gamma_j) \|\bar{x}_j - x_j\|^2},
\end{aligned}$$

and thus the power for $\|x_j - \bar{x}_j\|$ is larger in the numerator and the fraction converges to zero. ■

This next Lemma is similar to [19, Lemma 2.2].

Lemma 5.4 *Suppose Assumptions 5.2, 5.3, and 5.4.*

If $x_j, x_{j+1} \in N(x^, \delta/2)$, then,*

$$\text{dist}(x_{j+1}, X^*) \leq C_3 \text{dist}(x_j, X^*)^2, \quad (12)$$

where C_3 is a positive constant independent of j .

Proof. Indeed,

$$\begin{aligned}
\text{dist}(x_{j+1}, X^*)^2 &\leq \text{dist}(x_j + s_j, X^*)^2 \leq M_2^2 (\|F(x_j + s_j)\| - \bar{F})^2 \\
&\leq M_2^2 (\|J(x_j) s_j + F_j\| + L_3 \|s_j\|^2 - \bar{F})^2 \\
&\leq M_2^2 (C_1 \|x_j - \bar{x}_j\|^2 + \bar{F} + L_3 C_2^2 \|x_j - \bar{x}_j\|^2 - \bar{F})^2 \\
&\leq M_2^2 (C_1 + L_3 C_2^2)^2 \text{dist}(x_j, X^*)^4,
\end{aligned}$$

where we used Lemma 5.1 in the second to last inequality. ■

Theorem 5.1 *Algorithm 1 converges locally quadratically to X^* .*

Proof. From the previous results, it can be seen that eventually for $j \in \mathcal{K} = \mathbb{S} + 1$, it holds that there is a step s_j such that $x_j + s_j$ is quadratically closer to the solution, and is accepted. In particular, by the same argument as given in [19, Lemma 2.3] $x_j + s_j$ is always at least as close to x^* as x_j , and thus $x_{j+1} = x_j + s_j$ lies in a ball around x^* for which all of the local assumptions hold as well.

But then this implies that $j + 1 \in \mathcal{K}$ as well, and all of the previous results apply to it. Thus, proceeding inductively we get that for sufficiently large $j \in \mathcal{K}$, it holds that all subsequent iterations are in \mathcal{K} and the entire sequence of iterates $\{x_j\}$ (no longer subsequence) locally converges quadratically to X^* . ■

6 Numerical results

In this section we present some results of numerical experiments performed implementing Algorithm 1. We performed our experiments using the well known 33 Moré/Garbow/Hillstom test problems [13]. All the tested problems are smooth and have a least-squares structure. The residual function F and the Jacobian matrix for all the test problems [13] are implemented in MATLAB. Some of these problems have a non-zero value at the optimum and thus are consistent with the scope of the paper. In what comes next, we refer to the Moré/Garbow/Hillstom test problems by \mathcal{P} .

Among all the test problems in \mathcal{P} three problems can be seen as ill-conditioned (meaning that the condition number of their Jacobian at the optimum is larger than 10^7). Two of them are even nearly singular (i.e., the condition number of their Jacobian at the optimum is larger than 10^{20}). For all the tested problems, we used the proposed starting points x_0 as in the original optimization test problems [13].

A preliminary implementation of Algorithm 1 was written in MATLAB. The initial parameters defining the implemented algorithm were set as follows:

$$\eta = 0.01, \lambda = 2, \mu_0 = 1, \mu_{\min} = 10^{-16}.$$

At each iteration of Algorithm 1 we solved exactly the subproblem using the backslash MATLAB operator. If the iteration successful, we set the parameter μ_{j+1} equals to $\max(\bar{\mu}/\lambda, \mu_{\min})$. The algorithm is stopped when

$$\|\nabla f(x_j)\| \leq \epsilon \max(1, \|\nabla f(x_0)\|) \quad \text{with } \epsilon = 10^{-5}.$$

If it did not converge within a maximum number of iterations $j_{\max} = 10000$, then Algorithm 1 was considered to have failed.

The order of convergence was estimated by the quantity

$$\text{EOC} = \log \left(\frac{\|\nabla f(x_{j_f})\|}{\max(1, \|\nabla f(x_0)\|)} \right) / \log \left(\frac{\|\nabla f(x_{j_f-1})\|}{\max(1, \|\nabla f(x_0)\|)} \right), \quad (13)$$

where j_f denotes the index of the final computed iterate. The obtained values of EOC will help to estimate the speed of convergence. For the problems with an $\text{EOC} \geq 1.8$ Algorithm 1 will be said to be quadratically converging. For test problems with $1.8 > \text{EOC} \geq 1.1$ the algorithm

Table 1: Order of convergence distribution all over \mathcal{P} .

	Order of convergence			Total
	Linear or Sub-linear	Super-linear	Quadratic	
Well-conditioned	3	17	10	30
Poorly conditioned	1	0	2	3
Total	4	17	12	33

Table 2: The complete results for Algorithm 1 on the tested problems.

Problem	$f(x_{j_f})$	$\ \nabla f(x_{j_f-1})\ $	$\ \nabla f(x_{j_f})\ $	$\text{Cond}(J_{j_f})$	γ_{j_f}	EOC
rosen	8.366e-14	5.446e-03	6.468e-06	5.008e+01	3.191e-16	1.68
froth	4.898e+01	7.753e-04	5.058e-04	1.370e+07	4.653e-19	1.03
badscp	5.342e-03	1.812e+03	9.084e-03	3.571e+05	4.029e-08	8.14
badscb	1.000e-11	2.392e+03	2.307e-02	1.000e+06	1.513e-17	2.91
beale	1.130e-12	4.841e-04	5.022e-06	1.275e+01	4.927e-14	1.45
jensam	1.244e+02	5.115e-01	2.607e-02	1.147e+05	8.104e-11	1.26
helix	2.777e-10	7.306e-02	1.672e-04	2.222e+01	4.268e-13	1.64
bard	8.215e-03	1.123e-03	3.595e-06	6.687e+01	1.262e-14	1.55
gauss	1.128e-08	3.726e-03	3.260e-08	7.162e+00	5.314e-16	3.08
meyer	1.024e+05	9.623e+06	1.291e+02	3.266e+06	6.064e-08	2.33
gulf	5.794e-12	1.505e-05	4.018e-06	1.217e+05	1.615e-27	1.1
box	5.880e-16	1.796e-04	3.502e-08	9.133e+01	9.355e-21	1.66
sing	1.018e-06	1.019e-03	1.274e-04	1.650e+02	9.901e-13	1.17
wood	7.877e+00	1.027e+00	2.003e-03	3.589e+01	1.531e-11	1.69
kowosb	3.075e-04	1.162e-06	7.750e-07	6.991e+01	1.466e-16	1.03
bd	8.582e+04	1.168e+00	5.906e-01	1.084e+02	3.965e-14	1.05
osb1	5.465e-05	1.773e-03	3.638e-06	4.884e+04	5.049e-17	1.53
biggs	1.983e-19	8.949e-05	1.045e-09	1.134e+03	2.034e-27	2.19
osb2	4.014e-02	2.342e-05	1.564e-06	9.021e+01	5.968e-16	1.23
watson	1.679e-05	8.081e-02	2.587e-05	4.132e+04	1.307e-12	2.15
rosex	7.584e-14	1.120e-02	6.158e-06	5.008e+01	7.233e-17	1.75
singx	1.018e-06	1.019e-03	1.274e-04	1.650e+02	9.901e-13	1.17
pen1	2.250e-05	3.968e-04	4.192e-06	3.162e+02	5.238e-19	1.33
pen2	9.376e-06	8.631e-06	2.833e-06	8.255e+03	8.027e-28	1.08
vardim	5.439e-05	3.300e+00	1.449e-01	1.965e+01	7.822e-11	1.23
trig	5.449e-13	2.768e-04	6.538e-07	3.961e+00	1.336e-14	1.74
bv	2.518e-14	6.502e-05	2.793e-08	3.792e+01	1.950e-16	1.8
ie	3.998e-24	8.507e-06	2.569e-12	1.289e+00	8.248e-25	2.29
trid	1.160e-19	1.695e-04	1.064e-09	3.098e+00	4.423e-21	2.01
band	2.680e+00	2.621e-04	6.755e-05	1.197e+06	6.799e-17	1.1
lin	1.000e+01	2.286e-04	9.497e-14	1.000e+00	3.523e-29	3.11
lin1	4.634e+00	2.423e+04	7.678e-01	9.172e+63	1.757e-08	3.14
lin0	6.135e+00	1.473e+04	1.591e-01	Inf	3.774e-10	3.13

can be seen as super-linearly converging. Otherwise (i.e., $\text{EOC} < 1.1$), Algorithm 1 will be estimated to converge linearly or worse.

Table 1 gives the distribution of the order of convergence over our test set \mathcal{P} . The detailed results are presented in Table 2. The obtained results confirm our theoretical analysis in two senses. First, for all the tested problems, we observe the global convergence of our proposed algorithm. Second, we obtain quadratic or super-linear convergence rates on most of the problems tested. In fact, among the 33 tested problems, only 4 converge linearly or worse, for 12 problems we observe quadratic convergence and 17 are super-linearly converging.

7 Conclusion

In this paper we presented and analyzed a novel Levenberg-Marquardt method for solving non-linear least-squares problems. In particular we were interested in the class of problems for which there is no zero-residual solution (or such a solution is not sought after) and the problem is potentially ill-conditioned. Without the use of ancillary procedures to enforce globalization, we were able to formulate a method that is provably and numerically demonstrably convergent at a quadratic rate for such problems. Future research can include problems with constraints as well as those with noisy data.

References

- [1] E. Bergou, S. Gratton, and L. N. Vicente. Levenberg-Marquardt Methods Based on Probabilistic Gradient Models and Inexact Subproblem Solution, with Application to Data Assimilation. *SIAM/ASA J. Uncertainty Quantification*, 4:924–951, 2016.
- [2] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, 2013.
- [3] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, PA, USA, 2000.
- [4] P. Courtier, J.-N Thépaut, and A. Hollingsworth. A Strategy for Operational Implementation of 4D-Var, Using an Incremental Approach. *Q. J. R. Meteorol. Soc.*, 120:1367–1388, 1994.
- [5] H. Dan, N. Yamashita, and M. Fukushima. Convergence Properties of the Inexact Levenberg-Marquardt Method under Local Error Bound Conditions. *Optim. Methods Softw.*, 17:605–626, 2002.
- [6] F. Facchinei, A. Fischer, and M. Herrich. A Family of Newton Methods for Nonsmooth Constrained Systems with Nonisolated Solutions. *Math. Methods Operations Research*, 77:433–443, 2013.
- [7] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton Method: Nonsmooth Equations, KKT Systems, and Nonisolated Solutions. *Math. Program.*, 146:1–36, 2014.
- [8] J. Fan. Convergence Rate of the Trust Region Method for Nonlinear Equations under Local Error Bound Condition. *Comput. Optim. Appl.*, 34:215–227, 2006.

- [9] J. Fan and Y. Yuan. On the Quadratic Convergence of the Levenberg-Marquardt Method without Nonsingularity Assumption. *Computing*, 74:23–39, 2005.
- [10] I. C. F. Ipsen, C. T. Kelley, and S. R. Pope. Rank-deficient Nonlinear Least Squares Problems and Subset Selection. *SIAM J. Numer. Anal.*, 49:1244–1266, 2011.
- [11] K. Levenberg. A Method for the Solution of Certain Problems in Least Squares. *Quart. Appl. Math.*, 2:164–168, 1944.
- [12] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM J. Appl. Math.*, 11:431–441, 1963.
- [13] J. J. Moré, B. S. Garbow, and K. E. Hillstom. Testing Unconstrained Optimization Software. *ACM Trans. Math. Softw.*, 7:17–41, 1981.
- [14] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, Berlin, second edition, 2006.
- [15] M. R. Osborne. Nonlinear Least Squares – The Levenberg Algorithm Revisited. *J. Austral. Math. Soc. Ser. B*, 19:343–357, 1976.
- [16] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, 2005.
- [17] Y. Trémolet. Model Error Estimation in 4D-Var. *Q. J. R. Meteorol. Soc.*, 133:1267–1280, 2007.
- [18] J. Virieux and S. Operto. An Overview of Full-Waveform Inversion in Exploration Geophysics. *Geophysics*, 74:WCC1–WCC26, 2009.
- [19] N. Yamashita and M. Fukushima. On the Rate of Convergence of the Levenberg-Marquardt Method. In *Topics in numerical analysis*, pages 239–249. Springer, 2001.