Comput Optim Appl (2016) 65:723–751
DOI 10.1007/s10589-016-9845-x

Algebraic rules for computing the regularization parameter of the Levenberg–Marquardt method

Elizabeth W. Karas[1] · **Sandra A. Santos**[2] · **Benar F. Svaiter**[3]

Received: 18 May 2015 / Published online: 25 April 2016

Abstract This paper presents a class of Levenberg–Marquardt methods for solving the nonlinear least-squares problem. Explicit algebraic rules for computing the regularization parameter are devised. In addition, convergence properties of this class of methods are analyzed. We prove that all accumulation points of the generated sequence are stationary. Moreover, q-quadratic convergence for the zero-residual problem is obtained under an error bound condition. Illustrative numerical experiments with encouraging results are presented.

Keywords Nonlinear least-squares problems · Levenberg–Marquardt method · Regularization · Global convergence · Local convergence · Computational results

Mathematics Subject Classification 90C30 · 65K05 · 49M37

1 Introduction

Given $F : \mathbb{R}^n \to \mathbb{R}^m$, the nonlinear least-squares (NLS) problem is as follows:

$$\min_{x \in \mathbb{R}^n} \|F(x)\|^2,$$

✉ Elizabeth W. Karas
ewkaras@gmail.com

1 Department of Mathematics, Federal University of Paraná, CP 19081, Curitiba, PR 81531-980, Brazil

2 Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil

3 IMPA, Estrada Dona Castorina, 110, Rio de Janeiro 22460-320, Brazil

where $\|\cdot\|$ is the Euclidean norm. This problem is important from both a theoretical and practical viewpoint [3,18]. The seminal works of Levenberg [12], Morrison [15], and Marquardt [13] provided a regularization strategy for improving the convergence of the Gauss–Newton (GN) method. The latter is the first and simplest method to address the NLS problem. In the GN method, the nonlinear residual function F is replaced by its linear approximation at the current iterate, so that the NLS problem is solved by means of a sequence of quadratic problems, without any additional parameter. In the Levenberg–Marquardt (LM) method, a sequence of quadratic problems is also generated, and a regularization term is introduced that essentially depends on the so-called LM parameter.

For very large problems, as those arising in data assimilation, it may be necessary to make further approximations in the GN method in order to reduce computational costs. Gratton et al. [10] point out conditions that ensure the convergence of truncated and perturbed GN methods, deriving rates of convergence for the iterations. Inexact versions of the LM method have also been considered under an error bound condition. Such a condition is weaker than assuming nonsingularity of the appropriate matrix at a solution, namely the Jacobian $JF \in \mathbb{R}^{m\times n}$ if $m = n$ or the square matrix $JF^T JF$, otherwise [2,8]. The local convergence of the LM method has been studied also under an error bound condition. Yamashita and Fukushima [17] proved q-quadratic convergence for the LM method with the LM parameter sequence set as $\mu_k = \|F(x_k)\|^2$. Later, this q-quadratic rate was extended by Fan and Yuan [6] for the setting $\mu_k = \|F(x_k)\|^\delta$, with $\delta \in [1, 2]$. Fan and Pan [5] have enlarged upon the analysis for the variation of the exponent for $\delta \in (0, 1)$ within the aforementioned update of the LM parameter, showing local superlinear convergence. Accelerated versions of the LM method have been proposed recently by Fan [4], in which cubic local convergence was reached. When it comes to complexity analysis, Ueda and Yamashita [16] have investigated a global complexity bound for the LM method.

In this work we have devised algebraic rules for computing the LM parameter, so that μ_k belongs to an interval that is proportional to $\|F(x_k)\|$. Under the Lipschitz continuity of the Jacobian, the rules were generated in order to accept the full LM step by the Armijo sufficient decrease condition. On the one hand, under the availability of the Lipschitz constant, we have proved that the full steps are acceptable. On the other hand, for unknown Lipschitz constants, we have proposed a scheme for dynamically updating an estimate of such a constant within a well defined and globally convergent algorithm for solving the NLS problem. The q-quadratic rate of convergence for the zero-residual problem is obtained under an error bound condition. Numerical results illustrate the performance of the proposed algorithm for a set of test problems from the `CUTEst`, with promising results in terms of both efficiency and robustness.

The text is organized as follows. The basic results that have generated the proposed algorithm are presented in Sect. 2. The algorithm and its underlying properties are given in Sect. 3. The global convergence analysis is provided in Sect. 4, whereas the local convergence result is developed in Sect. 5. The numerical experiments are presented and examined in Sect. 6. A summary of the work is given in Sect. 7 and the tables with the complete numerical results compose the Appendix.

2 Technical results

We consider $F : \mathbb{R}^n \to \mathbb{R}^m$ of class $\mathcal{C}^1$ with an L-Lipschitz continuous Jacobian $JF \in \mathbb{R}^{m\times n}$, that is, for all $x, y \in \mathbb{R}^n$,

$$\|JF(x) - JF(y)\| \le L\|x - y\|. \tag{1}$$

On the left hand-side of the inequality above we have the operator norm induced by the canonical norms in $\mathbb{R}^n$ and $\mathbb{R}^m$. More specifically, we are concerned about choosing the regularization parameter μ in the Levenberg–Marquardt method,

$$s = -\big(JF(x)^T JF(x) + \mu I\big)^{-1} JF(x)^T F(x), \quad x_+ = x + ts, \quad 0 < t \le 1$$

where x_+ is the new iterate and t is the step length.

Initially, a classical result is recalled [3, Lemma 4.1.12].

Lemma 2.1 (Linearization error) *For any x, s in $\mathbb{R}^n$,*

$$\|F(x + s) - (F(x) + JF(x)s)\| \le \frac{L}{2}\|s\|^2.$$

From now on, in this section, s is the Levenberg–Marquardt step at $x \in \mathbb{R}^n$, with regularization parameter $\mu>0$, that is,

$$s = \operatorname{argmin}\|F(x)+JF(x)s\|^2+\mu\|s\|^2 = -\big(JF(x)^T JF(x)+\mu I\big)^{-1} JF(x)^T F(x). \tag{2}$$

Let $\phi : \mathbb{R} \to \mathbb{R}$ be

$$\phi(t) = \|F(x + ts)\|^2. \tag{3}$$

Note that

$$\begin{aligned}\phi'(0) &= 2\langle F(x), JF(x)s\rangle \\ &= -2\langle JF(x)^T F(x), \big(JF(x)^T JF(x) + \mu I\big)^{-1} JF(x)^T F(x)\rangle \le 0.\end{aligned} \tag{4}$$

First we analyze the norm of the linearization of F along the direction s.

Lemma 2.2 (Linearization's norm) *For any $t \in [0, 1]$,*

$$\begin{gathered}\|F(x) + tJF(x)s\|^2 = \|F(x)\|^2 + t\langle JF(x)^T F(x), s\rangle + \big(t^2 - t\big)\|JF(x)s\|^2 \\ -t\mu\|s\|^2 \le \|F(x)\|^2.\end{gathered}$$

Proof

$$\begin{aligned}\|F(x) + tJF(x)s\|^2 &= \|F(x)\|^2 + 2t\langle JF(x)^T F(x), s\rangle + t^2\|JF(x)s\|^2 \\ &= \|F(x)\|^2 + t\langle F(x), JF(x)s\rangle\end{aligned}$$

$$
\begin{aligned}
&+ t\langle JF(x)^T F(x), s\rangle + t^2\|JF(x)s\|^2 \\
&= \|F(x)\|^2 + t\langle F(x), JF(x)s\rangle \\
&\quad + t\langle -(JF(x)^T JF(x) + \mu I)s, s\rangle + t^2\|JF(x)s\|^2 \\
&= \|F(x)\|^2 + t\langle F(x), JF(x)s\rangle \\
&\quad + (t^2 - t)\|JF(x)s\|^2 - t\mu\|s\|^2.
\end{aligned}
$$

Using (4) and the fact that $t \in [0, 1]$, we conclude the proof. □

Now we analyze the norm of F along the direction s.

Lemma 2.3 *For any* $t \in [0, 1]$,

$$
\begin{aligned}
\|F(x + ts)\|^2 &\leq \|F(x)\|^2 + t\langle F(x), JF(x)s\rangle + (t^2 - t)\|JF(x)s\|^2 \\
&\quad + t\|s\|^2 \left[\frac{L^2}{4} t^3 \|s\|^2 + Lt\|F(x) + tJF(x)s\| - \mu\right] \\
&\leq \|F(x)\|^2 + t\langle JF(x)^T F(x), s\rangle + (t^2 - t)\|JF(x)s\|^2 \\
&\quad + t\|s\|^2 \left[\frac{L^2}{4} t^3 \|s\|^2 + Lt\|F(x)\| - \mu\right].
\end{aligned}
$$

Proof Let

$$
R(t) = F(x + ts) - (F(x) + tJF(x)s).
$$

Since $F(x + ts) = F(x) + tJF(x)s + R(t)$,

$$
\begin{aligned}
\|F(x + ts)\|^2 &= \|F(x) + tJF(x)s\|^2 + \|R(t)\|^2 + 2\langle F(x) + tJF(x)s, R(t)\rangle \\
&\leq \|F(x) + tJF(x)s\|^2 + \|R(t)\|^2 + 2\|F(x) + tJF(x)s\|\|R(t)\| \\
&\leq \|F(x) + tJF(x)s\|^2 + \|R(t)\|^2 + 2\|F(x)\|\|R(t)\|
\end{aligned}
$$

where the first inequality follows from Cauchy–Schwarz inequality and the second one from Lemma 2.2. From the definition of $R(t)$ and the assumption of JF being L-Lipschitz continuous, we have that $\|R(t)\| \leq Lt^2\|s\|^2/2$. Therefore,

$$
\|F(x + ts)\|^2 \leq \|F(x) + tJF(x)s\|^2 + \frac{L^2}{4} t^4 \|s\|^4 + Lt^2\|F(x)\|\|s\|^2.
$$

To end the proof, use Lemma 2.2 again. □

In view of Lemma 2.3 we need a bound for $\|s\|$ in order to choose, at a non-stationary point, a regularization parameter μ. The next technical lemma will be useful for obtaining such a bound.

Lemma 2.4 *For any $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $\mu>0$*

$$\left\|\left(A^TA+\mu I\right)^{-1}A^Tb\right\| \le \frac{1}{2\sqrt{\mu}}\|P_{\mathcal{R}(A)}b\|,$$

where $\mathcal{R}(A)$ is the range of A and $P_{\mathcal{R}(A)}$ is the orthogonal projection onto this subspace.

Proof Let $b' = P_{\mathcal{R}(A)}b$ and $s = (A^TA+\mu I)^{-1}A^Tb$. Observe that

$$\left(A^TA+\mu I\right)s = A^Tb'. \tag{5}$$

We will use an SVD of A. There exist U and V unitary matrices $m\times m$ and $n\times n$, respectively, such that

$$U^TAV = D, \quad d_{i,j} = \begin{cases}\sigma_i, & i=j\\ 0, & \text{otherwise},\end{cases}$$

where $\sigma_i \ge 0$ for all $i = 1,\ldots,\min\{m,n\}$. Note that $V^T = V^{-1}$, $U^T = U^{-1}$ and

$$V^TA^TAV = (U^TAV)^TU^TAV = D^TD.$$

Therefore, pre-multiplying both sides of (5) by V^T and using the substitutions

$$\tilde{s} = V^Ts, \quad \tilde{b} = U^Tb'$$

in (5) we conclude that

$$D^T\tilde{b} = V^T\left(A^TA+\mu I\right)V\tilde{s} = \left(D^TD+\mu I\right)\tilde{s}.$$

It follows from this equation that if $n \le m$ then

$$\tilde{s}_i = \frac{\sigma_i}{\sigma_i^2+\mu}\tilde{b}_i, \quad i = 1,\ldots,n;$$

while if $n>m$ then

$$\tilde{s}_i = \begin{cases}\dfrac{\sigma_i}{\sigma_i^2+\mu}\tilde{b}_i, & 1\le i\le m\\ 0, & m<i\le n.\end{cases}$$

Since $t/(t^2+\mu) \le 1/(2\sqrt{\mu})$ for $\mu>0$ and $t\ge 0$, we have

$$\|\tilde{s}\| \le \frac{1}{2\sqrt{\mu}}\|\tilde{b}\|.$$

To end the proof, note that $\|\tilde{s}\| = \|s\|$ and $\|\tilde{b}\| = \|b'\|$. □

In the following result, the main ingredients for the devised algebraic rules are established.

Theorem 2.5 *For any* $\mu > 0$,

$$\|s\| \le \frac{1}{2\sqrt{\mu}} \|P(F(x))\| \le \frac{1}{2\sqrt{\mu}} \|F(x)\|$$

where P *is the orthogonal projection onto the range of* $JF(x)$. *Moreover, if*

$$\mu \ge \frac{L}{4}\left(2\|F(x)\| + \sqrt{4\|F(x)\|^2 + \|P(F(x))\|^2}\right), \tag{6}$$

then

$$\|F(x+s)\|^2 \le \|F(x)\|^2 + \langle F(x), JF(x)s\rangle.$$

Proof The first part of the theorem, which are the bounds for $\|s\|$, follows from (2), Lemma 2.4 and the metric properties of the orthogonal projection. Combining the first part of the theorem with Lemma 2.3, we conclude that

$$\|F(x+s)\|^2 \le \|F(x)\|^2 + \langle F(x), JF(x)s\rangle + \frac{\|s\|^2}{\mu}\left[\frac{L^2}{16}\|P(F(x))\|^2 + L\|F(x)\|\mu - \mu^2\right].$$

To end the proof, note that the right hand-side of inequality (6) is the largest root of the concave quadratic $\mu \mapsto \frac{L^2}{16}\|P(F(x))\|^2 + L\|F(x)\|\mu - \mu^2$. □

Remark 2.6 In view of Theorem 2.5, possible choices for μ at a point x are

$$\mu = \frac{L}{4}\left(2\|F(x)\| + \sqrt{4\|F(x)\|^2 + \|P(F(x))\|^2}\right), \tag{7}$$

or

$$\mu = \frac{L}{4}(4\|F(x)\| + \|P(F(x))\|), \tag{8}$$

or

$$\mu = \frac{2+\sqrt{5}}{4} L\|F(x)\|. \tag{9}$$

The first value is the smallest one and the first two values require the computation of $\|P(F(x))\|$, although an upper bound for such a norm can also be used.

3 The algorithm

Now we propose and analyze the Levenberg–Marquard algorithm with algebraic rules for computing the regularization parameter.

Algorithm 1.

Input : $x_0 \in \mathbb{R}^n$, $\beta \in (0,1)$, $\eta \in [0,1)$, $L_0 > 0$ and $\delta \geq 0$ with $L_0 \geq \delta$.

1. $k \leftarrow 0$
2. **while** $JF(x_k)^T F(x_k) \neq 0$ **do**
3.   Set $F_k = F(x_k)$, $J_k = JF(x_k)$
4.   Choose $\mu_k \in [\mu_k^-, \mu_k^+]$ where P_k is the projection onto the range of J_k,

$$\mu_k^- = \frac{L_k}{4}\left(2\|F_k\| + \sqrt{4\|F_k\|^2 + \|P_k(F_k)\|^2}\right), \quad \mu_k^+ = \frac{2+\sqrt{5}}{4} L_k \|F_k\|$$

5.   Compute $s_k = -(J_k^T J_k + \mu_k I)^{-1} J_k^T F_k$
6. $t \leftarrow 1$
7.   **while** $\|F(x_k + ts_k)\|^2 > \|F_k\|^2 + \beta t \langle s_k, J_k^T F_k\rangle$ **do**
8. $t \leftarrow t/2$
9.   **end while**
10. $t_k = t$
11. $x_{k+1} = x_k + t_k s_k$
12.   **if** $t_k < 1$ **then**
13. $L_{k+1} = 2L_k$
14.   **else**
15. $\text{Ared} = \|F_k\|^2 - \|F(x_{k+1})\|^2$
16. $\text{Pred} = \|F_k\|^2 - \|F_k + J_k s_k\|^2 - \mu_k\|s_k\|^2 = -\langle s_k, J_k^T F_k\rangle$
17.   **if** $\text{Ared} > \eta\,\text{Pred}$ **then**
18. $L_{k+1} = \max\{L_k/2, \delta\}$
19.   **else**
20. $L_{k+1} = L_k$
21.   **end if**
22.   **end if**
23. $k \leftarrow k+1$
24. **end while**

Concerning Algorithm 1, it is worth noticing that

(i) Iteration ℓ begins with $k = \ell - 1$, and ends with $k = \ell$ if $JF(x_{\ell-1})^T F(x_{\ell-1}) \neq 0$.
(ii) If the algorithm does not end at iteration $k+1$, then $\mu_k > 0$, s_k is well defined and it is a descent direction for $\|F(\cdot)\|^2$. Therefore, the Armijo line search in Step 6–10 of Algorithm 1 has finite termination. Altogether, Algorithm 1 is well defined and either it terminates after ℓ steps with $JF(x_{\ell-1})^T F(x_{\ell-1}) = 0$, or it generates infinite sequences (x_k), (s_k), (t_k), (μ_k), (L_k).
(iii) L_k plays the role of the Lipschitz constant of JF. If $\delta > 0$, this parameter also plays the role of a safeguard which prevents L_k from becoming too small.

From now on, we assume that Algorithm 1 with input $x_0 \in \mathbb{R}^n$, $\beta \in (0,1)$, $\eta \in [0,1)$, $L_0 > 0$ and $\delta \geq 0$ does not stop at Step 2 and that (x_k), (s_k), (t_k), (μ_k), (L_k) are the (infinite) sequences generated by it.

Next we analyze the basic properties of Algorithm 1.

Proposition 3.1 *If $L_k \geq L$, then $t_k = 1$ and $L_{k+1} = \max\{L_k/2, \delta\}$.*

Proof Suppose that $L_k \geq L$. From this assumption and the definition of μ_k (in Step 4) we have that

$$\mu_k \geq \frac{L}{4}\left(2\|F(x_k)\| + \sqrt{4\|F(x_k)\|^2 + \|P(F(x_k))\|^2}\right),$$

where P is the orthogonal projection onto the range of $JF(x_k)$. The first equality of the proposition follows from the above inequality; the definition of s_k (in Step 5); Theorem 2.5 with $\mu = \mu_k$, $x = x_k$, $s = s_k$; and Steps 6–10. The second equality comes from the first one and Steps 12–22. □

Proposition 3.2 *For all k,*

$$\delta \leq L_k \leq \max\{L_0, 2L\}; \tag{10}$$

and, for infinitely many k, $t_k = 1$.

Proof Since $L_0 \geq \delta$ and $L_{k+1} \geq \max\{L_k/2, \delta\}$ for all k, the first inequality in (10) also holds for all k.

We will prove the second inequality by induction in k. This inequality holds trivially for $k = 0$. Assume that it holds for some k. Steps 12–22 of the algorithm imply that if $t_k = 1$, then $L_{k+1} = L_k$ or $L_{k+1} = \max\{\delta, L_k/2\} \leq L_k$ and in both cases the inequality holds for $k+1$. If $t_k < 1$, it follows from Proposition 3.1 that $L_k < L$ and then $L_{k+1} = 2L_k \leq 2L$. So the inequality holds for $k+1$ and the induction proof is complete.

To prove the second part of the proposition, suppose that $t_k < 1$ for any $k \geq k_0$. Then

$$L_k = 2^{k-k_0} L_{k_0}, \; k = k_0, \; k_0 + 1, \ldots$$

in contradiction with (10). □

From Proposition 3.2 and the Step 4 of the algorithm, we have

$$\delta\|F(x_k)\| \leq \mu_k \leq \frac{2+\sqrt{5}}{4}\max\{L_0, 2L\}\|F(x_k)\| \tag{11}$$

for all k.

Proposition 3.3 *For each k*

$$\begin{aligned} \|F(x_{k+1})\|^2 &\leq \|F(x_k)\|^2 + \beta t_k \langle JF(x_k)^T F(x_k), s_k\rangle \\ &\leq \|F(x_k)\|^2 - \beta t_k \frac{\|JF(x_k)^T F(x_k)\|^2}{\|JF(x_k)\|^2 + \mu_k}. \end{aligned} \tag{12}$$

As a consequence, the sequence $\left(\|F(x_k)\|^2\right)$ *is strictly decreasing and*

$$\sum_{k=0}^{\infty} \beta t_k \frac{\|JF(x_k)^T F(x_k)\|^2}{\|JF(x_k)\|^2 + \mu_k} \leq \|F(x_0)\|^2.$$

Proof The first inequality follows from the stopping condition for the Armijo line search (Steps 6–10). In view of the definition of s_k and μ_k, and the fact that $\mu_k > 0$,

$$\begin{aligned} -\langle JF(x_k)^T F(x_k), s_k \rangle &= \langle JF(x_k)^T F(x_k), (JF(x_k)^T JF(x_k) + \mu_k I)^{-1} JF(x_k)^T F(x_k) \rangle \\ &\geq \frac{\|JF(x_k)^T F(x_k)\|^2}{\|JF(x_k)^T JF(x_k)\| + \mu_k}, \end{aligned}$$

which trivially implies the second inequality. The last statement of the proposition follows directly from (12). □

4 General convergence analysis

The convergence of Algorithm 1 is examined in the following results.

Proposition 4.1 *If the sequence* (x_k) *is bounded, then it has a stationary accumulation point.*

Proof By Proposition 3.2, $t_k = 1$ for infinitely many k. Since (x_k) is bounded, there exists a subsequence (x_{k_j}) convergent to some $\bar{x}$, such that $t_{k_j} = 1$ for all j. Thus, by Proposition 3.3,

$$\sum_{j=1}^{\infty} \beta \frac{\|JF(x_{k_j})^T F(x_{k_j})\|^2}{\|JF(x_{k_j})\|^2 + \mu_{k_j}} \leq \|F(x_0)\|^2.$$

Moreover, since F and JF are continuous, $\|F(x_k)\|$, $\|JF(x_k)\|$ and μ_k are bounded. Hence the sequence $(JF(x_{k_j})^T F(x_{k_j}))$ converges to 0. To end the proof, use again the continuity of F and JF to conclude that $JF(\bar{x})^T F(\bar{x}) = 0$. □

From the bound for the norm of F along the direction s, obtained in Lemma 2.3, we prove next that the step length is bounded away from zero.

Proposition 4.2 *If* $\delta > 0$*, then*

$$t_k \geq \frac{8\delta^2/L^2}{1 + 16\delta/L}$$

for all k.

Proof For any $t \in [0, 1]$ it holds

$$
\begin{aligned}
\frac{L^2}{4} t^3 \|s_k\|^2 + Lt\|F(x_k)\| - \mu_k &\leq t\left(\frac{L^2}{4}\|s_k\|^2 + L\|F(x_k)\|\right) - \mu_k \\
&\leq t\left(\frac{L^2}{16\mu_k}\|F(x_k)\|^2 + L\|F(x_k)\|\right) - \mu_k \\
&\leq t\left(\frac{L^2}{16\delta}\|F(x_k)\| + L\|F(x_k)\|\right) - \delta\|F(x_k)\| \\
&= \delta\|F(x_k)\|\left[t\left(\frac{L^2}{16\delta^2} + \frac{L}{\delta}\right) - 1\right], \qquad (13)
\end{aligned}
$$

where the first inequality comes from the bound on t, the second from Theorem 2.5 and the third from the fact that $\mu_k \geq \delta\|F(x_k)\|$, as stated in (11).

From inequality (13) and Lemma 2.3, we have that if

$$
0 \leq t \leq \frac{1}{\dfrac{L^2}{16\delta^2} + \dfrac{L}{\delta}} = \frac{16\delta^2/L^2}{1 + 16\delta/L},
$$

then $\|F(x_k + ts_k)\|^2 \leq \|F(x_k)\|^2 + t\langle s_k, JF(x_k)^T F(x_k)\rangle$. So, the result follows from Steps 6–10 of Algorithm 1. □

We now present the global convergence result of Algorithm 1.

Proposition 4.3 *If $\delta>0$, then all accumulation points of the sequence (x_k) are stationary for the function $\|F(x)\|^2$.*

Proof Suppose that (x_{k_j}) converges to some $\bar{x}$. From Propositions 3.3 and 4.2, we have that

$$
\sum_{j=1}^{\infty} \frac{\|JF(x_{k_j})^T F(x_{k_j})\|^2}{\|JF(x_{k_j})\|^2 + \mu_{k_j}} < \infty.
$$

It follows from (11) and from the continuity of F and JF that $\|JF(x_{k_j})\|^2 + \mu_{k_j}$ is bounded. Therefore $JF(x_{k_j})^T F(x_{k_j})$ converges to 0. □

A complexity result concerning the stationarity measure of the Algorithm 1 is given next.

Proposition 4.4 *If $\delta>0$ and (x_k) is bounded, then*

$$
\min_{i=1,\ldots,k} \|JF(x_i)^T F(x_i)\| = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).
$$

Proof Define

$$M = \sup_k \left\{ \|JF(x_k)\|^2 + \frac{2+\sqrt{5}}{4} \max\{L_0, 2L\} \|F(x_k)\| \right\}.$$

Then, by (11) and Propositions 3.3 and 4.2, for any k

$$\begin{aligned} k\, \frac{\beta}{M} \left(\frac{8\delta^2/L^2}{1+16\delta/L} \right) \min_{i=1,\ldots,k} \|JF(x_i)^T F(x_i)\|^2 \\ \leq \sum_{i=1}^{k} \beta t_i \frac{\|JF(x_i)^T F(x_i)\|^2}{\|JF(x_i)\|^2 + \mu_i} \leq \|F(x_0)\|^2 \end{aligned}$$

and the conclusion follows. □

5 Quadratic convergence under an error bound condition

Given $F : \mathbb{R}^n \to \mathbb{R}^m$, consider the system

$$F(x) = 0 \tag{14}$$

and let X^* be its solution set, that is,

$$X^* = \{x \in \mathbb{R}^n \mid F(x) = 0\} \neq \emptyset. \tag{15}$$

Our aim is to prove that (x_k) converges quadratically near solutions of (14) where, locally, $\|F(x)\|$ *provides an error bound* for this system, in the sense of [17, Definition 1]. For completeness, we present this definition in the sequel (see Definition 5.3).

Define, for $\gamma > 0$, $\mathbf{S}_\gamma : \mathbb{R}^n \setminus X^* \to \mathbb{R}^n$

$$\mathbf{S}_\gamma(x) = -\Big(JF(x)^T JF(x) + \gamma \|F(x)\| I \Big)^{-1} JF(x)^T F(x). \tag{16}$$

Auxiliary bounds are established in the next two results.

Proposition 5.1 *If $x \in \mathbb{R}^n \setminus X^*$, $\gamma > 0$, $s = \mathbf{S}_\gamma(x)$, and $x_+ = x + s$ then*

$$\|JF(x_+)^T F(x_+)\| \leq (\gamma + L)\|F(x)\|\|s\| + \frac{L}{2}\|JF(x_+)\|\|s\|^2.$$

Proof Direct algebraic manipulations yield

$$\begin{aligned} JF(x_+)^T F(x_+) = {} & JF(x)^T (F(x) + JF(x)s) + (JF(x_+) \\ & - JF(x))^T (F(x) + JF(x)s) \\ & + JF(x_+)^T [F(x_+) - (F(x) + JF(x)s)]. \end{aligned}$$

It follows from (16) that $JF(x)^T(F(x)+JF(x)s) = -\gamma\|F(x)\|s$. Combining this result with the above equality, and using the triangle inequality, we have

$$\begin{aligned}\|JF(x_+)^T F(x_+)\| &\le \gamma\|F(x)\|\|s\| + \|JF(x_+) - JF(x)\|\|F(x)+JF(x)s\| \\ &\quad + \|JF(x_+)\|\|F(x_+) - (F(x)+JF(x)s)\| \\ &\le \gamma\|F(x)\|\|s\| + L\|F(x)\|\|s\| + \frac{L}{2}\|JF(x_+)\|\|s\|^2,\end{aligned}$$

where the last inequality follows from Assumption (1), Lemmas 2.1 and 2.2. □

Proposition 5.2 *If* $x \in \mathbb{R}^n \setminus X^*$, $\bar{x} \in X^*$, $\gamma > 0$, $s = \mathbf{S}_\gamma(x)$, *and* $\bar{s} = \bar{x} - x$, *then*

1. $\|F(x) + JF(x)\bar{s}\| \le \dfrac{L}{2}\|\bar{s}\|^2$;
2. $\|F(x)+JF(x)s\|^2 + \|JF(x)(\bar{s}-s)\|^2 + \gamma\|F(x)\|(\|s\|^2 + \|\bar{s}-s\|^2) \le \dfrac{L^2}{4}\|\bar{s}\|^4 + \gamma\|F(x)\|\|\bar{s}\|^2$.

Proof Since $F(\bar{x}) = 0$,

$$F(x) + JF(x)\bar{s} = F(x) + JF(x)(\bar{x} - x) - F(\bar{x})$$

and the first inequality follows from this equality and Lemma 2.1.

Define

$$\psi_{\gamma,x} : \mathbb{R}^n \to \mathbb{R}, \quad \psi_{\gamma,x}(u) = \|F(x) + JF(x)u\|^2 + \gamma\|F(x)\|\|u\|^2.$$

From item 1 we have

$$\psi_{\gamma,x}(\bar{s}) \le \frac{L^2}{4}\|\bar{s}\|^4 + \gamma\|F(x)\|\|\bar{s}\|^2.$$

Observe that $s = \operatorname{argmin}_{u\in\mathbb{R}^n}\psi_{\gamma,x}(u)$. Since $\psi_{\gamma,x}$ is a quadratic with Hessian $2(JF(x)^T JF(x) + \gamma\|F(x)\|I)$ and it is minimized by s,

$$\begin{aligned}\psi_{\gamma,x}(\bar{s}) &= \psi_{\gamma,x}(s) + \|JF(x)(\bar{s}-s)\|^2 + \gamma\|F(x)\|\|\bar{s}-s\|^2 \\ &= \|F(x)+JF(x)s\|^2 + \|JF(x)(\bar{s}-s)\|^2 + \gamma\|F(x)\|\left(\|s\|^2 + \|\bar{s}-s\|^2\right).\end{aligned}$$

The second inequality of the proposition follows from the two above relations. □

We will analyze the local convergence of the sequence (x_k) under the local error bound condition, as defined next. Such a condition is weaker than assuming nonsingularity of $JF(x)^T JF(x)$ for x at the solution set X^*.

Definition 5.3 ([17, Definition 1]) Let V be an open subset of $\mathbb{R}^n$ such that $V \cap X^* \ne \emptyset$, where X^* is as in (15). We say that $\|F(x)\|$ *provides a local error bound* on V for the system (14) if there exists a positive constant c such that

$$c \ \operatorname{dist}(x, X^*) \le \|F(x)\|$$

for all $x \in V$.

The next lemma, which will be instrumental in the main result of this section, was proved in [7, Corollary 2]. A proof is provided here for the sake of completeness, where the function F is simply continuously differentiable.

Lemma 5.4 *If $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and $\|F(x)\|$ provides an error bound for (14) in a neighborhood V of $x^* \in X^*$ as in Definition 5.3, then $\|JF(x)^T F(x)\|$ also provides an error bound for (14) in some neighborhood of x^*.*

Proof Suppose that $V \subset \mathbb{R}^n$ is a neighborhood of x^* where $\|F(x)\|$ provides an error bound for (14), that is,

$$c \text{ dist}(x, X^*) \leq \|F(x)\| \quad \forall x \in V.$$

Define, for $x, y \in \mathbb{R}^n$

$$R(y, x) = F(y) - (F(x) + JF(x)(y - x)).$$

Since F is continuously differentiable, there exists $r>0$ such that $B(x^*, r) \subseteq V$ and

$$\|R(y, x)\| \leq c\|y - x\|/2 \quad \forall y, x \in B(x^*, r).$$

Take $x \in B(x^*, r/2)$ and $\bar{x} \in \arg\min_{z \in X^*} \|z - x\|$. Then $\text{dist}(x, X^*) = \|\bar{x} - x\| < r/2$, $\bar{x} \in B(x^*, r)$ and, in view of the above assumptions,

$$c\,\|\bar{x} - x\| \leq \|F(x)\|, \quad \|R(\bar{x}, x)\| \leq c\|\bar{x} - x\|/2. \tag{17}$$

Since $F(\bar{x}) = 0$, $-JF(x)(\bar{x} - x) = F(x) + R(\bar{x}, x)$ and

$$-\langle JF(x)^T F(x), \bar{x} - x\rangle = -\langle F(x), JF(x)(\bar{x} - x)\rangle = \|F(x)\|^2 + \langle F(x), R(\bar{x}, x)\rangle.$$

Using the above equalities, Cauchy-Schwarz inequality and (17) we conclude that

$$\begin{aligned} \|JF(x)^T F(x)\|\|\bar{x} - x\| &\geq \|F(x)\|\big(\|F(x)\| - \|R(\bar{x}, x)\|\big) \\ &\geq \|F(x)\|\big(c\,\|\bar{x} - x\| - (c/2)\|\bar{x} - x\|\big) \geq c^2\|\bar{x} - x\|^2/2. \end{aligned}$$

Therefore, $\|JF(x)^T F(x)\| \geq (c^2/2) \text{ dist}(x, X^*)$ for any $x \in B(x^*, r/2)$. □

From now on, in this section, $x^* \in \mathbb{R}^n$, $r, c>0$ are such that

$$F(x^*) = 0, \qquad c \text{ dist}(x, X^*) \leq \|F(x)\| \qquad \forall x \in B(x^*, r). \tag{18}$$

Another auxiliary result is provided next.

Lemma 5.5 *Consider $x^* \in \mathbb{R}^n$, $r, c>0$ satisfying (18). If $x \in B(x^*, r) \setminus X^*$, $\bar{x} \in \arg\min_{u \in X^*} \|u - x\|$, $\gamma>0$, $s = \mathbf{S}_\gamma(x)$ and $\bar{s} = \bar{x} - x$, then*

1. $\|\bar{s}\| = \mathrm{dist}(x, X^*) \le \|F(x)\|/c$;
2. $\|s\| \le \left(1 + \frac{L^2}{4c\gamma}\|\bar{s}\|\right)^{1/2} \|\bar{s}\|$;
3. $\|F(x) + JF(x)s\| \le \left(\frac{L^2}{4c^2\gamma^2}\|\bar{s}\|^2 + \frac{1}{c\gamma}\|\bar{s}\|\right)^{1/2} \gamma\|F(x)\|$.

Moreover, if $\|\bar{s}\| \le \frac{c\gamma}{2L^2}$ *then* $\|F(x+s)\|^2 \le \|F(x)\|^2 + \langle s, JF(x)^T F(x)\rangle$.

Proof The first relation in item 1 follows trivially from the definition of $\bar{x}$ while the second relation comes from (18).

From item 2 of Proposition 5.2 and item 1 of this lemma, we have that

$$\gamma\|F(x)\|\|s\|^2 \le \frac{L^2}{4}\|\bar{s}\|^4 + \gamma\|F(x)\|\|\bar{s}\|^2 \le \frac{L^2}{4c}\|F(x)\|\|\bar{s}\|^3 + \gamma\|F(x)\|\|\bar{s}\|^2$$

and

$$\begin{aligned}\|F(x) + JF(x)s\|^2 &\le \frac{L^2}{4}\|\bar{s}\|^4 + \gamma\|F(x)\|\|\bar{s}\|^2 \\ &\le \left(\frac{L^2}{4c^2\gamma^2}\|\bar{s}\|^2 + \frac{1}{c\gamma}\|\bar{s}\|\right)\gamma^2\|F(x)\|^2,\end{aligned}$$

which trivially imply items 2 and 3, respectively.

To prove the last part of the lemma, suppose that $\|\bar{s}\| \le \frac{c\gamma}{2L^2}$ and define

$$a = \frac{L^2}{4}\|s\|^2 + L\|F(x) + JF(x)s\| - \gamma\|F(x)\|, \qquad w = \frac{L^2}{c\gamma}\|\bar{s}\|.$$

From items 2 and 3, we have

$$\begin{aligned} a &\le \frac{L^2}{4}\left(1 + \frac{L^2}{4c\gamma}\|\bar{s}\|\right)\|\bar{s}\|^2 + L\left(\frac{L^2}{4c^2\gamma^2}\|\bar{s}\|^2 + \frac{1}{c\gamma}\|\bar{s}\|\right)^{1/2}\gamma\|F(x)\| - \gamma\|F(x)\| \\ &\le \left[\frac{L^2}{4c\gamma}\|\bar{s}\|\left(1 + \frac{L^2}{4c\gamma}\|\bar{s}\|\right) + L\left(\frac{L^2}{4c^2\gamma^2}\|\bar{s}\|^2 + \frac{1}{c\gamma}\|\bar{s}\|\right)^{1/2} - 1\right]\gamma\|F(x)\| \\ &= \left[\frac{w}{4}\left(1 + \frac{w}{4}\right) + w^{1/2}\left(1 + \frac{w}{4}\right)^{1/2} - 1\right]\gamma\|F(x)\|, \end{aligned}$$

where the second inequality follows from item 1, and the equality comes from the definition of w. Observe that since $w \le 1/2$ it follows that $a < 0$. To end the proof, use the first inequality in Lemma 2.3 with $t = 1$ and $\mu = \gamma\|F(x)\|$. □

Note that in Algorithm 1, $s_k = \mathbf{S}_\gamma(x_k)$ for $\gamma = \mu_k/\|F(x_k)\|$. In order to simplify the proofs, define

$$D = \frac{2+\sqrt{5}}{4}\max\{L_0, 2L\}, \qquad \gamma_k = \frac{\mu_k}{\|F(x_k)\|}, \qquad k \in \mathbb{N}. \tag{19}$$

In view of (11) and the definition of s_k in Algorithm 1, for all $k \in \mathbb{N}$,

$$\delta \leq \gamma_k \leq D, \qquad s_k = \mathbf{S}_{\gamma_k}(x_k). \tag{20}$$

Assuming that the residual function F provides an error bound for the solution set of the NLS zero-residual problem, the local convergence of Algorithm 1 is established as follows.

Theorem 5.6 *Consider $x^* \in \mathbb{R}^n$, $r, c>0$ satisfying (18). There exists $\tilde{r}>0$ such that if $x_{k_0} \in B(x^*, \tilde{r})$ for some k_0, then either Algorithm 1 stops at some x_k solution of (14) or (x_k) converges q-quadratically to some $\hat{x}$ solution of (14).*

Proof In view of Lemma 5.4, there exist $0 < r_1 \leq r$ and $c_1>0$ such that

$$c_1 \operatorname{dist}(x, X^*) \leq \|JF(x)^T F(x)\|, \qquad \forall x \in B(x^*, r_1). \tag{21}$$

Define

$$M_1 = \max\left\{\|JF(x)\| \mid x \in B(x^*, r_1)\right\}, \quad M_2 = \frac{3M_1}{2\sqrt{2}c_1}\left(D + L\left(1 + \frac{3}{4\sqrt{2}}\right)\right)$$

and

$$\rho = \min\left\{\frac{c\,\delta}{2L^2}, \frac{\sqrt{2}}{3}r_1, \frac{1}{2M_2}\right\}.$$

We claim that if

$$\begin{aligned} &x \in B(x^*, \rho) \setminus X^*,\ \delta \leq \gamma \leq D,\ s = \mathbf{S}_\gamma(x),\ x_+ = x + s, \\ &\bar{x} \in \arg\min_{u \in X^*} \|u - x\|,\ \bar{s} = \bar{x} - x, \end{aligned} \tag{22}$$

then

$$\|s\| \leq \frac{3}{2\sqrt{2}} \operatorname{dist}(x, X^*) \leq \frac{3}{2\sqrt{2}}\rho, \tag{23}$$

$$\operatorname{dist}(x_+, X^*) \leq M_2 \operatorname{dist}(x, X^*)^2 \leq \operatorname{dist}(x, X^*)/2, \tag{24}$$

$$\|F(x_+)\|^2 \leq \|F(x)\|^2 + \langle s, JF(x)^T F(x)\rangle. \tag{25}$$

The first inequality in (23) follows from item 2 of Lemma 5.5 and the definitions of $\bar{s}$ and ρ. The second inequality comes from the inclusions $x^* \in X^*$ and $x \in B(x^*, \rho)$.

To prove (24) first observe that

$$\|x_+ - x^*\| \leq \|x - x^*\| + \|s\| \leq \rho + \frac{3}{2\sqrt{2}}\rho < r_1$$

which comes from the definition of x_+, the triangle inequality, (23) and the definition of ρ. Consequently, from (21) for x_+ and Proposition 5.1,

$$\begin{aligned}\operatorname{dist}(x_+, X^*) &\leq \frac{1}{c_1}\|JF(x_+)^T F(x_+)\| \\ &\leq \frac{1}{c_1}\left((\gamma + L)\|F(x)\| + \frac{L}{2}\|JF(x_+)\|\|s\|\right)\|s\|.\end{aligned}$$

Since JF is continuous, $F(x) = F(\bar{x}) - \int_0^1 JF(\bar{x} - t\bar{s})\,\bar{s}\,dt$. As $F(\bar{x}) = 0$, we have

$$\|F(x)\| \leq \left(\max_{x \in B(x^*,\rho)} \|JF(x)\|\right)\|\bar{s}\| \leq M_1\|\bar{s}\|.$$

Using the two relations displayed above, the bounds for γ in (22), (23), and the definitions of M_1 and M_2 we have

$$\operatorname{dist}(x_+, X^*) \leq \frac{3M_1}{2\sqrt{2}c_1}\left(D + L\left(1 + \frac{3}{4\sqrt{2}}\right)\right)\|\bar{s}\|^2 = M_2 \operatorname{dist}(x, X^*)^2,$$

which proves the first inequality of (24). The second inequality in (24) comes from the fact that $\operatorname{dist}(x, X^*) \leq \rho$ and from the definition of ρ.

Our third claim, (25), follows directly from the last result of Lemma 5.5.

Next we define a family of sets on $\mathbb{R}^n$ which are, in some sense, well behaved with respect to Algorithm 1. Define

$$W_\tau = \left\{x \in \mathbb{R}^n \;\middle|\; \|x - x^*\| \leq \frac{3\tau + \sqrt{2}}{3 + \sqrt{2}}\rho,\ \operatorname{dist}(x, X^*) \leq \frac{(1-\tau)\sqrt{2}}{3 + \sqrt{2}}\rho\right\},$$
$$W = \bigcup_{0 \leq \tau < 1} W_\tau,$$

and let

$$\tilde{r} = \frac{\sqrt{2}}{3 + \sqrt{2}}\rho.$$

Since $x^* \in X^*$, we have $W_0 = B(x^*, \tilde{r})$. Hence

$$B\left(x^*, \tilde{r}\right) \subset W \subset B(x^*, \rho) \subset B(x^*, r_1).$$

Let $x_k \in W$. If $JF(x_k)^T F(x_k) = 0$ then the algorithm stops at x_k and, in view of (21), $F(x_k) = 0$. Next we analyze the case $JF(x_k)^T F(x_k) \neq 0$. It follows from (20) and (25) that

$$x_{k+1} = x_k + \mathbf{S}_{\gamma_k}(x_k) = x_k + s_k.$$

As $x_k \in W$, we have $x_k \in W_\tau$ for some $0 \le \tau < 1$. Therefore, from the triangle inequality, the definition of W_τ, and (23)

$$\begin{aligned}\|x_{k+1} - x^*\| &\le \|x_k - x^*\| + \|s_k\| \\ &\le \frac{3\tau + \sqrt{2}}{3 + \sqrt{2}}\rho + \frac{3}{2\sqrt{2}} \operatorname{dist}(x_k, X^*) \\ &\le \frac{3\tau + \sqrt{2}}{3 + \sqrt{2}}\rho + \frac{3}{2\sqrt{2}} \frac{(1-\tau)\sqrt{2}}{3 + \sqrt{2}}\rho = \frac{3\left(\frac{1+\tau}{2}\right) + \sqrt{2}}{3 + \sqrt{2}}\ \rho.\end{aligned}$$

Additionally, from (24) and the definition of W_τ we also have

$$\operatorname{dist}(x_{k+1}, X^*) \le \frac{1}{2} \operatorname{dist}(x_k, X^*) \le \frac{(1-\tau)\sqrt{2}}{2(3+\sqrt{2})}\rho = \frac{\left(1 - \frac{1+\tau}{2}\right)\sqrt{2}}{3 + \sqrt{2}}\rho.$$

Altogether we proved that

$$0 \le \tau < 1,\ x_k \in W_\tau,\ JF(x_k)^T F(x_k) \ne 0 \Rightarrow\ x_{k+1} \in W_{\frac{1+\tau}{2}}.$$

Suppose that $x_{k_0} \in B(x^*, \tilde{r})$. We have just proved that in this case either the algorithm stops at some $x_k \in X^*$ or an infinite sequence is generated and $x_k \in B(x^*, \rho)$ for $k \ge k_0$. Assume that an infinite sequence is generated and define

$$d_k = \operatorname{dist}(x_k, X^*).$$

It follows from (23) and (24) that for $k \ge k_0$,

$$\|s_k\| \le \frac{3}{2\sqrt{2}} d_k, \quad d_{k+1} \le M_2 d_k^2 \le \frac{d_k}{2}. \tag{26}$$

Hence, $\sum_{j=k_0}^{\infty} \|x_{j+1} - x_j\| \le \frac{3}{2\sqrt{2}} \sum_{j=k_0}^{\infty} d_j \le \frac{3}{\sqrt{2}} d_{k_0}$. As (x_k) is a Cauchy sequence, it converges to some $\hat{x}$. Since $d_k = \operatorname{dist}(x_k, X^*)$ converges to 0, $F(\hat{x}) = 0$, that is, $\hat{x} \in X^*$. By the triangle inequality and the first inequality in (26),

$$\|x_k - \hat{x}\| \le \|s_k\| + \sum_{j=k+1}^{\infty} \|s_j\| \le \frac{3}{2\sqrt{2}} \left[d_k + \sum_{j=k+1}^{\infty} d_j \right].$$

From the last two inequalities in (26) we have

$$\sum_{j=k+1}^{\infty} d_j \le \sum_{j=k}^{\infty} M_2 d_j^2 \le M_2 d_k \sum_{j=k}^{\infty} d_j \le 2M_2 d_k^2 \le d_k.$$

Therefore, from the two relations displayed above,

$$d_k \leq \|x_k - \hat{x}\| \leq \frac{3}{\sqrt{2}} d_k,$$

where the first inequality comes from the inclusion $\hat{x} \in X^*$. Consequently, using (26) again and the definition of d_k, we conclude that

$$\|x_{k+1} - \hat{x}\| \leq \frac{3}{\sqrt{2}} d_{k+1} \leq \frac{3M_2}{\sqrt{2}} d_k^2 \leq \frac{3M_2}{\sqrt{2}} \|x_k - \hat{x}\|^2,$$

ensuring the q-quadratic convergence and completing the proof. □

6 Numerical experiments

In this section we describe numerical experiments to illustrate the practical performance of the algorithm. We start with the algorithmic and computational choices for obtaining the numerical results.

6.1 On the choice of μ_k

Similarly to the analysis developed for the unconstrained minimization problem (cf. [11, Prop.1.2]), as the positive scalar μ is a lower bound for the smallest eigenvalue of $JF(x)^T JF(x) + \mu I$, it follows that

$$\mu\|s\|^2 \leq \langle s, \big(JF(x)^T JF(x) + \mu I\big)s\rangle = -\langle s, JF(x)^T F(x)\rangle \leq \|s\|\|JF(x)^T F(x)\|$$

and so $\|s\| \leq \frac{\|JF(x)^T F(x)\|}{\mu}$.

By Lemma 2.3, with $t = 1$,

$$\|F(x+s)\|^2 \leq \|F(x)\|^2 + \langle F(x), JF(x)s\rangle + \|s\|^2 \left[\frac{L^2}{4}\|s\|^2 + L\|F(x)\| - \mu\right].$$

For obtaining μ that ensures $\frac{L^2}{4}\|s\|^2 + L\|F(x)\| - \mu \leq 0$, the previous condition concerning the step gives

$$L^2\|JF(x)^T F(x)\|^2 + 4L\|F(x)\|\mu^2 - 4\mu^3 \leq 0. \tag{27}$$

Let $\mu^2 = z$, and the concave function $\psi : \mathbb{R}_+ \to \mathbb{R}$ given by

$$\psi(z) = L^2\|JF(x)^T F(x)\|^2 + 4L\|F(x)\|z - 4z^{3/2}.$$

Consider $z_0 = L^2\|F(x)\|^2$. Note that $\psi(z_0) = L^2\|JF(x)^T F(x)\|^2 \geq 0$.

An iteration of Newton's method from z_0 gives us

$$z_1 = z_0 - \frac{\psi(z_0)}{\psi'(z_0)} = L^2\|F(x)\|^2 + \frac{L\|JF(x)^T F(x)\|^2}{2\|F(x)\|}.$$

Then (27) is guaranteed for all

$$\mu \geq \sqrt{L^2\|F(x)\|^2 + \frac{L\|JF(x)^T F(x)\|^2}{2\|F(x)\|}} \stackrel{\text{def}}{=} \mu_J.$$

Observe that $L\|F(x)\| \leq \mu_J$, $\mu^+ = \frac{2+\sqrt{5}}{4}L\|F(x)\|$ and so $\frac{4}{2+\sqrt{5}}\mu^+ \leq \mu_J$. Nevertheless, there is no guarantee that $\mu_J \leq \mu^+$ holds, so we set $\mu_k = \min\{\mu_k^+, \mu_J\}$.

6.2 On the computation of the step s_k

The equivalence between the problems

$$\min_s \frac{1}{2}\|J_k s + F_k\|^2 + \frac{\mu}{2}\|s\|^2 \quad \text{and} \quad \min_s \frac{1}{2}\left\|\begin{pmatrix} J_k \\ \sqrt{\mu_k} I \end{pmatrix} s + \begin{pmatrix} F_k \\ 0 \end{pmatrix}\right\|^2$$

implies that the system

$$\left(J_k^T J_k + \mu_k I\right) s + J_k^T F_k = 0 \tag{28}$$

may be handled by the normal equations as follows (cf. [14])

$$\left(J_k^T \;\; \sqrt{\mu_k} I\right)\left[\begin{pmatrix} J_k \\ \sqrt{\mu_k} I \end{pmatrix} s + \begin{pmatrix} F_k \\ 0 \end{pmatrix}\right] = 0.$$

As an alternative to the Cholesky factorization of $J_k^T J_k + \mu_k I$, the QR factorization of $\begin{pmatrix} J_k \\ \sqrt{\mu_k} I \end{pmatrix}$ avoids performing the product $J_k^T J_k$:

$$\begin{pmatrix} J_k \\ \sqrt{\mu_k} I \end{pmatrix} = Q \begin{pmatrix} R_\mu \\ 0 \end{pmatrix},$$

where $Q \in \mathbb{R}^{(m+n)\times(m+n)}$ is orthogonal and $R_\mu \in \mathbb{R}^{n\times n}$ is upper triangular. Indeed, the economic version of the factorization may be used, and just the first n columns of the orthogonal matrix Q are computed. The direction s is thus obtained from the system (28) by means of the relationship

$$J_k^T J_k + \mu_k I = R_\mu^T R_\mu.$$

6.3 About the backtracking scheme

Besides the simple bisection of Steps 6–10 of Algorithm 1, we have implemented a quadratic–cubic interpolation scheme (cf. [3, §6.3.2]). As this scheme performed slightly better in preliminary experiments, it was used in the reported results.

6.4 About the initialization of the sequence (L_k)

We set $L_0 = 10^{-6}$ and $L_{\min} = 10^{-12}$, compute x_1 and define $L_1 = \max\left\{\frac{\|JF(x_1)-JF(x_0)\|_F}{\sqrt{n}\|x_1-x_0\|}, L_{\min}\right\}$. After that, the updating scheme follows Steps 12–22 of Algorithm 1.

6.5 The results

The tests were performed in a notebook DELL Intel Core i7-4510U, CPU @2.00GHz × 4, with 16GB RAM, Inspiron 5000 - i15 5547-A30, 64-bit, using `Matlab` 2014a, v. 8.3.

The set of test problems consists of all 53 problems from the `CUTEst` collection [9] such that the nonlinear equations $F(x) = 0$ are recast from feasibility instances, i.e, problems without objective function, with nonlinear equality constraints and without fixed variables for which the Jacobian matrix is available. The constraint bodies and Jacobians were evaluated in sparse format. The initial point x_0 was always the default of the collection.

To put our approach in perspective, the same problems were addressed by two distinct approaches. The first one is the Self-adaptive Levenberg–Marquardt Algorithm of Fan and Pan in [5], which is close to the scheme proposed in this paper. The second one is the modular code `lsqnonlin` (available within the `Matlab` software), based on the Levenberg–Marquardt Method [12,13,15]. The remaining parameters of the Algorithm 1 were defined by $\beta = 10^{-4}$, $\eta = 10^{-4}$ and $\delta = 10^{-8}$. These parameters are the default values suggested in [5] for the parameters that play a similar role to ours. The choices of Fan and Pan denoted by p_0 and m correspond to our η and δ, respectively. Moreover, the parameter δ of Fan and Pan was set as 1.

Summing up, there are four strategies under analysis:

- `CH`: Algorithm 1 with Cholesky factorization for computing s_k;
- `QR`: Algorithm 1 with QR factorization for computing s_k;
- `FP`: Fan and Pan's algorithm [5, Alg. 2.1], with Cholesky factorization for computing the step and the aforementioned parameters and values;
- `LM`: The Levenberg–Marquardt algorithm (an option of the routine `lsqnonlin` of `Matlab`).

Concerning the implemented stopping criteria, we have adopted the same numbering of the exit flags as the routine `lsqnonlin`. Setting ε_{mac} as the machine precision, the stopping criteria commom to the four strategies were the following:

(1) Convergence to a solution with relative stationarity:
$\|JF(x_k)^T F(x_k))\|_\infty \leq 10^{-10} \max\{\|(JF(x_0)^T F(x_0))\|_\infty, \sqrt{\varepsilon_{\text{mac}}}\}$;

(2) Change in x (too small): $\|x_{k+1} - x_k\| \leq 10^{-9}\left(\sqrt{\varepsilon_{\text{mac}}} + \|x_{k+1}\|\right)$;
(3) Change in the norm of the residue (too small): $|\|F(x_{k+1})\|^2 - \|F(x_k)\|^2| \leq 10^{-6}\|F(x_k)\|^2$;
(4) Computed search direction (too small): $\|s_k\| \leq 10^{-9}$;
(0) Maximum number of functional evaluations exceeded max $= \{2000, 10n\}$.

Besides, as strategy `FP` does not compute the step length t_k, the next criterion was included only for the strategies `CH`, `QR` and `LM`:

(`-4`) Line search failed, as the step length is too small: $|t_k| \leq 10^{-15}$.

Let f^* be the objective function value obtained by a strategy `S` when this strategy is applied to a given problem with the default initial point. We consider that the strategy `S` *has found a solution* (cf. [1]) if

$$\frac{f^* - f_{\min}}{\max\{1, f_{\min}\}} \leq 10^{-4}, \tag{29}$$

where $f_{\min}$ is the smallest function value found among all the strategies under comparison.

All outputs are reported in the Appendix. Tables 1, lists the 53 test problems. The table columns display the name of the problem; the dimensions (n and m); the number of iterations (#Iter); the number of function evaluations (#Fun); the function value at the last iterate (f^*); the CPU time in seconds (CPU), and the reason for stopping (exit). We observe that the stopping criteria `2` and `-4` were never activated during the tests. The results of the four solvers `CH`, `QR`, `FP` and `LM`, respectively, are displayed row by row for each problem.

As it is usual to have some variation of the CPU time from one execution of an algorithm to the other, for each problem we ran nine times all the solvers and we considered the average CPU time of the last eight runs, discarding the CPU time of the first one. The symbol † indicates that the obtained solution does not satisfy (29). It is worth mentioning that `CH` and `FP` performed one Cholesky factorization per iteration. The only exception occurred at a single iteration of the problem `10FOLDTR` for `CH` and of the problems `10FOLDTR` and `ARGLBLE` for `FP`. In these three instances, both strategies required an additional Cholesky factorization with a slight increase in μ_k to ensure the numerically safe positive definiteness of $J_k^T J_k + \mu_k I$.

The results corresponding to the *solved problems* are depicted in the performance profiles of Fig. 1 for the number of iterations, the number of function evaluations and the required CPU time. The logarithmic scale was used in the horizontal axis for better visualization of the results. In terms of efficiency, our strategies slightly outperformed the `FP` and `LM` with regard to the number of iterations and function evaluations. Indeed, 52.8 % of the problems were solved with the fewest number of iterations for both `CH` and `QR`, 35.8 % for `FP` and 49.1 % for `LM`. Moreover, 49.1 % of the problems were solved with the fewest number of function evaluations for both `CH` and `QR`, and 43.4 % for `FP` and `LM`. Now, concerning the CPU time, the shortest time was never reached by `QR`, whereas strategies `CH`, `FP` and `LM` solved respectively 35.8, 41.5 and 22.6 % of the problems in the shortest time. Furthermore, `CH` and `FP` solved 70 % of the problems using no more than twice the best CPU time. Strategies `CH` and `QR` are the most robust,

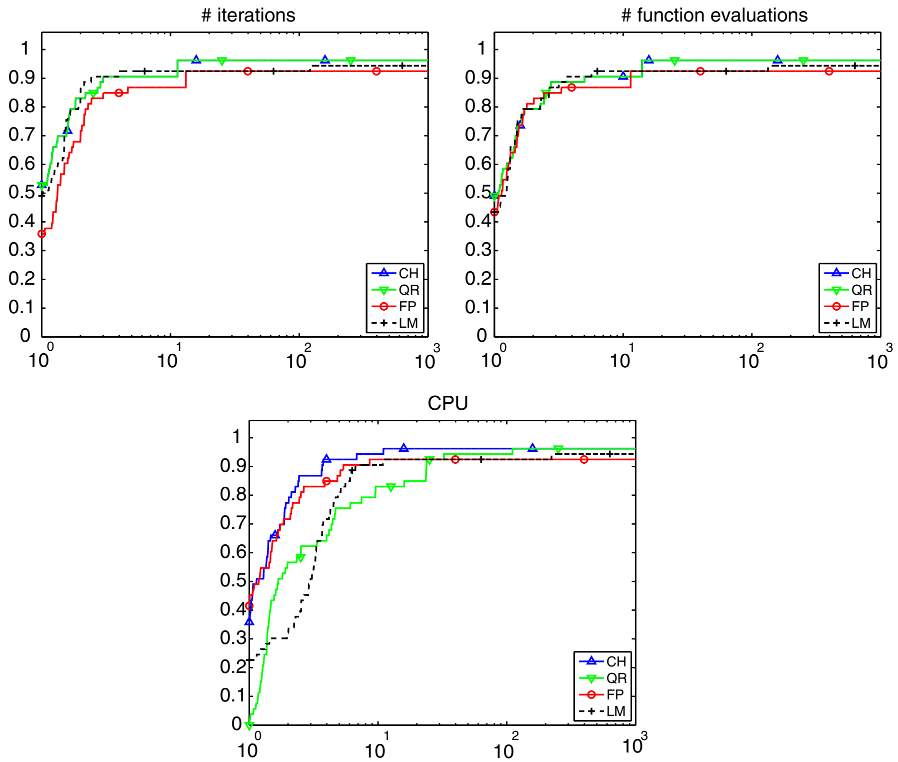

Fig. 1 Performance profiles for the number of iterations, the number of function evaluations and the required CPU time

solving 51 of the 53 problems. Problems `EIGENB` and `YATP2SQ` were considered not solved by `CH`, and `10FOLDTR` and `YATP2SQ` were not solved by strategy `QR`. On the other hand, `LM` did not solve three problems (`10FOLDTR`, `CYCLIC3`, `YATP2SQ`), while `FP` did not solve four problems (`10FOLDTR`, `ARWHDNE`, `CYCLIC3`, `EIGENB`), showing that our approach is competitive.

Aiming at illustrating the rate of convergence of the proposed algebraic rules, Fig. 2 shows the logarithm (base 10) of the squared residual value against the iterations for two typical problems.

7 Final remarks

We have proposed and analyzed a class of Levenberg–Marquardt methods in which algebraic rules for computing the regularization parameter were devised. Under the Lipschitz continuity of the Jacobian of the residual functions, the algebraic rules were proposed to allow the full LM-step to be accepted by the Armijo sufficient decrease condition. In terms of global convergence, all the accumulation points of the sequence generated by the algorithm are stationary. As for the local convergence for the zero-

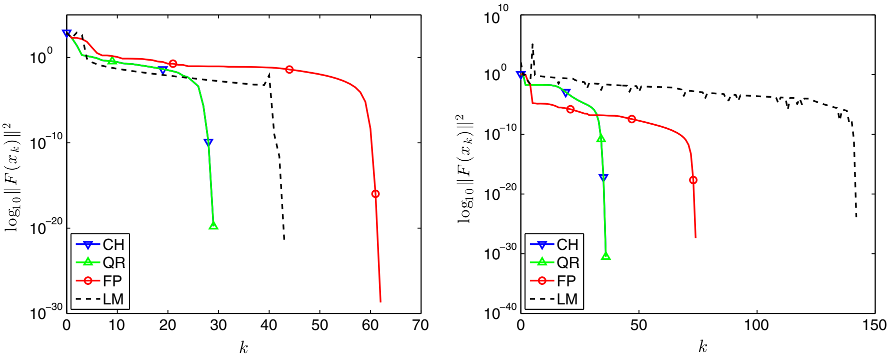

Fig. 2 The residual value $\|F(x_k)\|^2$ against the iterations, for the problems `EIGENC` (*left*) and `POWELLBS` (*right*)

residual problem, we have proved a q-quadratic rate under an error-bound condition. This condition is less restrictive than assuming nonsingularity at the solution. A set of numerical experiments was prepared to illustrate the practical performance of the proposed algorithm. Our approach has shown both efficiency and robustness for the 53 feasibility instances from the `CUTEst`. It has performed slightly better than both the algorithm proposed by Fan and Pan in [5] and the routine `lsqnonlin` of `Matlab`. Obtaining inexact solutions for the linear systems with adequately matching stopping criteria, as well as considering nonzero-residual problems in the local convergence analysis are topics for future investigations.

Acknowledgments The authors are thankful to Abel S. Siqueira for installing the `CUTEst` and preparing our machine to run it within `Matlab`. We are also grateful to the anonymous referees whose suggestions led to improvements in the paper. Elizabeth W. Karas was partially supported by CNPq Grants 477611/2013-3 and 308957/2014-8. Sandra A. Santos was partially supported by CNPq Grant 304032/2010-7, FAPESP Grants 2013/05475-7 and 2013/07375-0 and PRONEX Optimization. Benar F. Svaiter was partially supported by CNPq Grants 302962/2011-5, 474944/2010-7, FAPERJ Grant E-26/102.940/2011 and PRONEX Optimization.

8 Appendix

The complete computational results are presented next. The outcomes of the four solvers `CH`, `QR`, `FP` and `LM`, respectively, are displayed row by row for each problem.

Table 1 Numerical results

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
10FOLDTR	100	100	14	15	2.6403e+17	3.1343e−02	1
			13	14	3.4234e+17	4.7807e−02	1†
			6	7	1.4272e+25	1.4219e−02	3†
			12	13	2.3485e+18	2.2098e−02	1†

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
ARGAUSS	3	15	2	3	1.1279e−08	7.0329e−04	3
			2	3	1.1279e−08	8.0829e−04	3
			2	3	1.1279e−08	6.6125e−04	3
			3	4	1.1279e−08	2.9815e−03	3
ARGLALE	200	400	2	3	2.0000e+02	1.9904e−01	1
			2	3	2.0000e+02	2.0100e−01	1
			3	4	2.0000e+02	2.9091e−01	3
			3	4	2.0000e+02	2.8607e−01	3
ARGLBLE	200	400	1	2	9.9625e+01	1.2199e−01	1
			1	2	9.9625e+01	1.2382e−01	1
			2	3	9.9625e+01	1.9875e−01	1
			1	2	9.9625e+01	1.3872e−01	1
ARGTRIG	200	200	4	5	2.1345e−20	1.5488e−01	1
			4	5	2.1346e−20	1.1674e−01	1
			4	5	2.3426e−26	1.5821e−01	1
			3	4	2.9360e−18	7.2793e−02	1
ARWHDNE	100	198	20	32	2.7662e+01	7.4501e−03	3
			20	32	2.7662e+01	3.4314e−02	3
			11	12	2.7709e+01	3.8385e−03	3†
			11	22	2.7662e+01	9.9535e−03	3
BOOTH	2	2	2	3	0.0000e+00	6.5857e−04	1
			2	3	0.0000e+00	7.4643e−04	1
			2	3	1.9824e−21	6.1887e−04	1
			3	4	1.9564e−18	2.8978e−03	1
BROWNALE	200	200	3	4	1.9879e−09	9.6117e−02	1
			3	4	1.9879e−09	8.8616e−02	1
			3	4	3.9149e−17	1.1809e−01	1
			3	4	2.2300e−09	7.0575e−02	1
BROYDN3D	1000	1000	5	6	1.2079e−29	1.1405e−01	1
			5	6	1.3657e−29	1.1462e+00	1
			5	6	1.1241e−29	8.9167e−02	1
			5	6	1.9529e−28	1.0356e−02	1
BROYDNBD	500	500	7	8	9.2340e−20	4.0205e−02	1
			7	8	9.2340e−20	2.5363e−01	1
			6	7	9.7128e−22	2.8365e−02	1
			6	7	9.7950e−22	1.0694e−02	1
CHANDHEU	500	500	20	21	4.1643e−22	8.4619e+00	1
			20	21	4.1652e−22	6.1886e+00	1
			19	20	7.8066e−22	8.3485e+00	1
			18	19	1.5968e−20	4.5117e+00	1

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
CHNRSBNE	50	98	45	64	6.6325e−25	8.7170e−03	1
			45	64	6.6325e−25	3.8636e−02	1
			55	56	2.7752e−22	8.7747e−03	1
			46	70	4.8262e−25	2.1671e−02	1
CLUSTER	2	2	9	10	7.2338e−19	1.3299e−03	1
			9	10	7.2338e−19	1.6533e−03	1
			9	10	9.5248e−23	1.1657e−03	1
			9	10	1.4396e−22	3.9510e−03	1
COOLHANS	9	9	35	38	2.4541e−22	4.4511e−03	1
			35	38	2.4701e−22	6.2834e−03	1
			22	23	5.3026e−24	2.6813e−03	1
			44	62	1.6044e−24	1.4946e−02	1
CUBENE	2	2	19	25	2.4339e−25	2.4196e−03	1
			19	25	2.4339e−25	3.1379e−03	1
			16	17	9.4250e−27	1.7311e−03	1
			7	10	1.3981e−21	3.9445e−03	1
CYCLIC3	1002	1002	61	85	4.7699e+04	1.3981e+00	1
			61	85	4.7699e+04	1.3381e+01	1
			14	15	1.5918e+06	3.3463e−01	1†
			12	13	2.0870e+08	2.3353e−02	1†
EIGENAU	110	110	3	4	3.6011e−20	3.4029e−03	1
			3	4	3.6011e−20	1.1392e−02	1
			4	5	0.0000e+00	4.0535e−03	1
			5	6	1.7060e−23	1.0722e−02	1
EIGENB	110	110	24	31	6.6876e−02	2.2791e−02	3†
			77	119	2.4501e−25	2.8192e−01	1
			18	19	1.5813e+00	2.1614e−02	4†
			114	172	3.2208e−21	2.6145e−01	1
EIGENC	462	462	29	41	1.6159e−20	4.4347e−01	1
			29	41	1.6158e−20	1.7681e+00	1
			62	63	1.9206e−29	1.1141e+00	1
			42	62	2.5177e−22	1.1227e+00	1
GOTTFR	2	2	14	18	2.7570e−24	1.8471e−03	1
			14	18	2.7562e−24	2.4013e−03	1
			42	43	7.2171e−26	3.9125e−03	1
			9	13	8.5139e−21	4.6638e−03	1
GROWTH	3	12	74	105	1.0040e+00	9.1603e−03	3
			74	105	1.0040e+00	1.3144e−02	3
			96	97	1.0040e+00	8.8942e−03	3
			93	139	1.0040e+00	2.9116e−02	1

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
HATFLDF	3	3	30	42	5.3994e−29	3.5911e−03	1
			30	42	5.4516e−29	5.0013e−03	1
			56	57	4.8148e−35	5.1130e−03	1
			23	32	4.7766e−30	8.0228e−03	1
HATFLDG	25	25	11	17	3.4180e−29	2.2786e−03	1
			11	17	3.3967e−29	3.4887e−03	1
			11	12	1.7482e−19	1.7687e−03	1
			9	12	2.7245e−20	5.1259e−03	1
HEART6	6	6	140	205	4.6561e−27	1.8327e−02	1
			140	205	3.7832e−27	2.3911e−02	1
			140	141	3.7101e−27	1.4222e−02	1
			298	446	4.7970e−26	8.8176e−02	1
HEART8	8	8	35	50	7.7007e−22	4.9433e−03	1
			35	50	7.7024e−22	6.7484e−03	1
			57	58	1.4421e−27	5.9642e−03	1
			54	81	2.7297e−26	1.8570e−02	1
HIMMELBA	2	2	2	3	0.0000e+00	6.8286e−04	1
			2	3	0.0000e+00	7.4800e−04	1
			2	3	1.2534e−19	6.6137e−04	1
			3	4	8.8033e−18	2.8189e−03	1
HIMMELBC	2	2	5	7	4.6689e−22	1.0046e−03	1
			5	7	4.6689e−22	1.1770e−03	1
			7	8	0.0000e+00	1.0929e−03	1
			6	9	1.2168e−28	3.6983e−03	1
HIMMELBD	2	2	35	50	5.9226e+00	4.1591e−03	3
			35	50	5.9226e+00	5.4614e−03	3
			21	22	5.9226e+00	2.1696e−03	3
			16	28	5.9226e+00	7.0777e−03	3
HIMMELBE	3	3	5	6	3.4426e−29	1.0206e−03	1
			5	6	3.4426e−29	1.2193e−03	1
			3	4	1.1063e−22	7.3688e−04	1
			4	5	1.3989e−22	3.0638e−03	1
HYDCAR20	99	99	659	981	4.4058e−24	3.0219e−01	1
			659	981	4.6180e−24	1.5295e+00	1
			403	404	1.9660e−24	1.6077e−01	1
			620	928	4.8696e−07	4.9402e−01	3
HYDCAR6	29	29	115	166	6.8762e−26	2.2417e−02	1
			115	166	1.1268e−25	4.1551e−02	1
			101	102	8.6926e−22	1.6523e−02	1
			245	364	1.1073e−26	1.1031e−01	1

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
HYPCIR	2	2	5	7	2.0954e−31	1.0394e−03	1
			5	7	2.0954e−31	1.2023e−03	1
			8	9	2.9230e−24	1.1021e−03	1
			5	7	5.5524e−21	3.4355e−03	1
KSS	1000	1000	6	7	1.4564e−09	2.0158e+01	1
			6	7	1.4564e−09	1.3076e+01	1
			6	7	6.7372e−22	2.0541e+01	1
			5	6	5.5519e−05	8.3977e+00	1
METHANB8	31	31	15	16	7.6840e−26	2.9310e−03	1
			15	16	7.6195e−26	5.5773e−03	1
			5	6	1.1096e−26	1.2844e−03	1
			8	9	3.1180e−18	4.7629e−03	1
METHANL8	31	31	20	22	4.6850e−22	3.8543e−03	1
			20	22	4.7170e−22	7.4031e−03	1
			7	8	2.4794e−26	1.5854e−03	1
			8	9	1.5002e−17	4.5685e−03	1
MSQRTA	1024	1024	12	14	9.9331e−28	8.1086e−01	1
			12	14	9.6502e−28	6.0304e+00	1
			14	15	6.4271e−28	1.2130e+00	1
			10	14	6.4582e−19	2.1630e+00	1
MSQRTB	1024	1024	11	13	2.4854e−27	9.3771e−01	1
			11	13	2.2154e−27	5.4492e+00	1
			13	14	4.0239e−24	8.9350e−01	1
			10	14	2.2396e−22	1.7536e+00	1
OSCIGRNE	100	100	6	7	1.3793e−10	2.1686e−03	1
			6	7	1.3793e−10	9.0243e−03	1
			6	7	2.0393e−14	2.1793e−03	1
			6	7	2.0393e−14	4.3731e−03	1
OSCIPANE	10	10	1	10	1.0000e+00	9.8714e−04	3
			1	10	1.0000e+00	9.6829e−04	3
			2	3	1.0000e+00	6.7287e−04	3
			1	2	1.0000e+00	2.4082e−03	3
POWELLBS	2	2	36	37	4.4453e−31	3.7107e−03	1
			36	37	3.0893e−31	5.1241e−03	1
			74	75	4.0926e−28	6.4416e−03	1
			142	210	1.2797e−24	4.0419e−02	1
POWELLSQ	2	2	11	15	9.6553e−18	1.6490e−03	1
			11	15	9.6553e−18	2.0507e−03	1
			24	25	7.7265e−23	2.4055e−03	1
			1334	2001	4.9296e−12	3.6657e−01	0

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
RECIPE	3	3	12	13	5.9618e−12	1.5240e−03	1
			12	13	5.9618e−12	2.0644e−03	1
			12	13	2.2208e−12	1.4981e−03	1
			12	13	2.2221e−12	4.4696e−03	1
RSNBRNE	2	2	15	19	0.0000e+00	2.0154e−03	1
			15	19	0.0000e+00	2.5593e−03	1
			45	46	0.0000e+00	4.1940e−03	1
			29	43	1.0227e−25	9.8485e−03	1
SINVALNE	2	2	22	31	4.4191e−25	2.8500e−03	1
			22	31	4.4191e−25	3.6909e−03	1
			47	48	3.7153e−24	4.2986e−03	1
			26	39	1.9461e−23	9.0417e−03	1
SPIN	667	665	11	12	5.4650e−18	1.9316e−01	1
			11	12	5.4650e−18	1.7885e+00	1
			6	7	4.8635e−20	1.1193e−01	1
			6	7	2.7334e−20	1.3700e−01	1
SPIN2	102	100	8	9	1.4970e−27	4.6690e−02	1
			8	9	2.0428e−27	4.7750e−02	1
			5	6	1.9659e−27	3.0468e−02	1
			5	6	1.6395e−27	2.5038e−02	1
SPMSQRT	499	829	7	9	1.6458e−16	3.0316e−02	1
			7	9	1.6458e−16	4.8699e−01	1
			12	13	1.6458e−16	7.2393e−02	1
			7	10	1.6458e−16	1.4985e−02	1
YATP1NE	120	120	68	98	4.9111e−25	3.6879e−02	1
			68	98	1.8968e−24	2.3689e−01	1
			79	80	3.4735e−25	5.1083e−02	1
			6	7	2.2654e−22	1.0024e−02	1
YATP1SQ	120	120	68	98	4.9111e−25	4.9754e−02	1
			68	98	1.8968e−24	1.7366e−01	1
			79	80	3.4735e−25	3.9237e−02	1
			6	7	2.2654e−22	7.2845e−03	1
YATP1SS	120	120	68	98	4.9111e−25	3.6677e−02	1
			68	98	1.8968e−24	2.3648e−01	1
			79	80	3.4735e−25	3.8826e−02	1
			6	7	2.2654e−22	1.0072e−02	1
YATP2SQ	10200	10200	4	6	1.9659e+07	3.0017e+01	3†
			4	6	1.9659e+07	8.0057e+02	3†
			28	29	1.4253e−28	2.2061e+02	1
			41	62	1.7803e+08	1.3012e+03	3†

Table 1 continued

Problem	n	m	#Iter	#Fun	f^*	CPU	Exit
YFITNE	3	17	21	23	6.6697e−13	3.9234e−03	1
			21	23	6.6697e−13	5.5287e−03	1
			37	38	6.6697e−13	3.8822e−03	1
			42	61	6.6697e−13	1.9941e−02	1
ZANGWIL3	3	3	2	3	2.0917e−32	9.9386e−04	1
			2	3	2.1139e−32	1.0361e−03	1
			3	4	3.3173e−29	7.0775e−04	1
			4	5	3.3594e−25	4.2871e−03	1

References

1. Birgin, E.G., Gentil, J.M.: Evaluating bound-constrained minimization software. Comput. Optim. Appl. **53**(2), 347–373 (2012)
2. 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)
3. Dennis Jr., J.E., Schnabel, R.B.: Numerical methods for unconstrained optimization and nonlinear equations. In: Classics in Applied Mathematics, vol 16. Society for Industrial and Applied Mathematics (SIAM). Philadelphia, PA (1996). Corrected reprint of the 1983 original
4. Fan, J.: Accelerating the modified Levenberg-Marquardt method for nonlinear equations. Math. Compd. **83**(287), 1173–1187 (2014)
5. Fan, J., Pan, J.: Convergence properties of a self-adaptive Levenberg-Marquardt algorithm under local error bound condition. Comput. Optim. Appl. **34**(1), 47–62 (2006)
6. Fan, J., Yuan, Y.-X.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Computing **74**(1), 23–39 (2005)
7. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Progr. **94**((1, Ser. A)), 91–124 (2002)
8. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg-Marquardt methods. Optimization **59**(2), 273–287 (2010)
9. Gould, N., Orban, D., Toint, P.: Cutest: a constrained and unconstrained testing environment with safe threads. Technical Report RAL-TR-2013-005. Rutherford Appleton Laboratory, Oxfordshire (2013)
10. Gratton, S., Lawless, A.S., Nichols, N.K.: Approximate Gauss-Newton methods for nonlinear least squares problems. SIAM J. Optim. **18**(1), 106–132 (2007)
11. Karas, E.W., Santos, S.A., Svaiter, B.F.: Algebraic rules for quadratic regularization of Newton's method. Comput. Optim. Appl. **60**, 343–376 (2014)
12. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Quart. Appl. Math. **2**, 164–168 (1944)
13. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**, 431–441 (1963)
14. Moré, J.J.: The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G.A. (ed.), Lecture Notes in Mathematics 630: Numerical Analysis, pp. 105–116. Springer, New York (1978)
15. Morrison, D.D.: Methods for nonlinear least squares problems and convergence proofs. In: Lorell, J., Yagi, F. (eds.), Proceedings of the Seminar on Tracking Programs and Orbit Determination, pp. 1–9. Jet Propulsion Laboratory, Pasadena (1960)
16. Ueda, K., Yamashita, N.: On a global complexity bound on the Levenberg-Marquardt method. J. Optimiz. Theory App. **147**, 443–453 (2010)
17. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. In: Alefeld, G., Chen, X. (eds.) Topics in Numerical Analysis: With Special Emphasis on Nonlinear Problems, Computing Supplement, vol. 15, pp. 239–249. Springer, Vienna (2001)
18. Yuan, Y.-X.: Recent advances in numerical methods for nonlinear equations and nonlinear least squares. Numer. Algebra Control Optim. **1**(1), 15–34 (2011)