Optim Lett (2012) 6:927–940
DOI 10.1007/s11590-011-0321-3

ORIGINAL PAPER

A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods

Roger Behling · Andreas Fischer

Received: 17 March 2010 / Accepted: 29 March 2011 / Published online: 13 April 2011

Abstract The Levenberg–Marquardt method is a regularized Gauss–Newton method for solving systems of nonlinear equations. If an error bound condition holds it is known that local quadratic convergence to a non-isolated solution can be achieved. This result was extended to constrained Levenberg–Marquardt methods for solving systems of equations subject to convex constraints. This paper presents a local convergence analysis for an inexact version of a constrained Levenberg–Marquardt method. It is shown that the best results known for the unconstrained case also hold for the constrained Levenberg–Marquardt method. Moreover, the influence of the regularization parameter on the level of inexactness and the convergence rate is described. The paper improves and unifies several existing results on the local convergence of Levenberg–Marquardt methods.

Keywords Constrained equation · Levenberg–Marquardt method · Convergence rate · Inexactness · Non-isolated solution

1 Introduction

Let a sufficiently smooth mapping $H : \mathbb{R}^n \to \mathbb{R}^m$ and a closed convex nonempty set $\Omega \subseteq \mathbb{R}^n$ be given. In this paper we consider the following system of nonlinear

R. Behling
Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110,
Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: rogerb@impa.br

A. Fischer (✉)
Department of Mathematics, Institute of Numerical Mathematics,
Technische Universität Dresden, 01062 Dresden, Germany
e-mail: Andreas.Fischer@tu-dresden.de

equations subject to constraints

$$H(x) = 0, \quad x \in \Omega. \tag{1}$$

In the last decade, Levenberg–Marquardt methods [15,17] and related techniques (see [8,11]) have turned out to be a valuable tool for solving such problems, in particular for the case when problem (1) has non-isolated solutions. For results on the local convergence related to this case we refer to [7,8,10,14,18,19]. Several globalization techniques have been suggested, see [1,11,14,16,20] for example. In [4,9,14] applications of Levenberg–Marquardt techniques to other problems are dealt with. For problems (1) where H is only semismooth at a solution Levenberg–Marquardt type algorithms have been developed in [5,12,13]. However, the conditions used for proving their local superlinear convergence imply the local uniqueness of the solution.

In the present paper we are interested in extending and unifying results on the local convergence of inexact Levenberg–Marquardt methods for smooth problems (1) that can have non-isolated solutions. We are now going to discuss this in more detail.

Levenberg–Marquardt methods guarantee a local quadratic (or superlinear) convergence under an error bound condition (see [18] for the first result in this direction). Even in the classical case when $\Omega = \mathbb{R}^n$ and $m = n$ this condition is weaker than requiring that the Jacobian of H at a solution x^* is nonsingular. The latter however implies the isolatedness of x^*. Under such a nonsingularity condition a well known theory for the level of inexactness in Newton's method exists [3].

The level of inexactness in the subproblems of Levenberg–Marquardt methods that is possible without loosing a given superlinear convergence rate was investigated in [2,6,10] for the unconstrained case, i.e., if $\Omega = \mathbb{R}^n$. Then, the Levenberg–Marquardt subproblems read as follows

$$\frac{1}{2}\|H(s) + \nabla H(s)^\top (x - s)\|^2 + \frac{1}{2}\alpha(s)\|x - s\|^2 + \pi(s)^\top (x - s) \to \min_x, \tag{2}$$

where $s \in \mathbb{R}^n$ can be understood as the current iterate, $\alpha(s) > 0$ denotes the regularization parameter, and $\pi(s) \in \mathbb{R}^n$ is used to formally describe the inexactness that may result from approximate data, truncated solution algorithms, or numerical errors. To obtain a large level for the inexactness under a given convergence rate the regularization parameter $\alpha(s)$ plays a crucial role.

To explain this let us assume that a quadratic convergence rate should be guaranteed. Then, it is shown in [2] that $\alpha(s) \sim \|H(s)\|^2$ enables an inexactness level of (at least) $\|\pi(s)\| \sim \|H(s)\|^4$. The inexactness level of $\|\pi(s)\| \sim \|H(s)\|^3$ was obtained in [6] if a significantly larger regularization parameter is used, namely if

$$\alpha(s) \sim \|H(s)\|. \tag{3}$$

For this choice of $\alpha(s)$ the inexactness level can be further improved to $\|\pi(s)\| \sim \|H(s)\|^2$, see [10]. This result may serve as a bench mark if we consider the constrained case. Levenberg–Marquardt methods for this case were suggested in [14] and in [19] (the latter for Ω defined by nonnegativity constraints only). The subproblems of the constrained Levenberg–Marquardt method in [14] read as follows

$$\frac{1}{2}\|H(s)+\nabla H(s)^\top(x-s)\|^2+\frac{1}{2}\alpha(s)\|x-s\|^2\to\min_x\quad\text{s.t.}\quad x\in\Omega.$$

The constrained Levenberg–Marquardt method is known to converge locally with a quadratic rate if $\alpha(s)\sim\|H(s)\|^2$ is assumed, see [14] (and [19] under slightly different conditions). It seems that nothing is known on the behavior of inexact versions of the constrained Levenberg–Marquardt method. The reason for this may lie in the fact that it is not even clear whether a quadratic rate is possible if the larger value (3) is used for the regularization parameter $\alpha(s)$.

The present paper gives answers to these problems. It will turn out that the regularization parameter $\alpha(s)$ in terms of $\|H(s)\|^\beta$ with $\beta\in(0,2]$ is responsible for the rate of convergence of the exact constrained Levenberg–Marquardt method. Moreover, inexactness does not worsen this rate if it is at most proportional to $\|H(s)\|^{\beta+1}$. It will also be shown in Sect. 5 that this level of inexactness is sharp. The results in this paper improve or extend previous results in particular those in [2,6,10,14,19].

In Sect. 2 problem (1) is reformulated as an optimization problem and its necessary optimality conditions are represented as a generalized equation. Based on this an auxiliary lemma on the upper Lipschitz-continuity for a perturbation of the generalized equation is proved. In Sect. 3 the subproblems of the inexact constrained Levenberg–Marquardt method and the resulting method are formally defined. The local convergence analysis is given in Sect. 4. By an example the sharpness of the inexactness level is demonstrated in Sect. 5.

The norm $\|\cdot\|$ always denotes the Euclidean vector norm. By $\mathscr{B}$ the closed unit ball in $\mathbb{R}^n$ and by $\mathscr{B}(z,\rho)$ the closed ball around z with radius ρ is denoted.

2 An upper Lipschitz-continuity result

Throughout this paper we assume that the solution set of problem (1) is nonempty, i.e.,

$$X^*:=\{x\in\Omega\,|\,H(x)=0\}\neq\emptyset.$$

Then, any solution of (1) solves the minimization problem

$$\frac{1}{2}\|H(x)\|^2\to\min_x\quad\text{s.t.}\quad x\in\Omega,\tag{4}$$

and vice versa. If H is differentiable then every solution of the minimization problem satisfies the necessary optimality condition

$$0\in\nabla H(x)H(x)+\mathrm{N}_\Omega(x),\tag{5}$$

where

$$\mathrm{N}_\Omega(x) := \begin{cases} \{y \in \mathbb{R}^n \mid y^\top(z-x) \le 0 \quad \text{for all } z \in \Omega\} & \text{if } x \in \Omega, \\ \emptyset & \text{if } x \notin \Omega \end{cases}$$

denotes the normal cone to the set Ω at x. In addition to the generalized Eq. (5) we also consider the perturbed generalized equation

$$p \in \nabla H(x)H(x) + \mathrm{N}_\Omega(x) \tag{6}$$

for perturbation parameters $p \in \mathbb{R}^n$. Let $X(p)$ denote the solution set of (6). Obviously, $X^* \subseteq X(0)$ holds.

Assumption 1 The function $H : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable and $\nabla H : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ is locally Lipschitz continuous.

To formulate the error bound condition (Assumption 2 below) let $x^* \in X^*$ be an arbitrary but fixed solution. Assumption 1 could be restricted to some neighborhood of x^* which, however, is avoided for simplicity.

Assumption 2 There are $w > 0$ and $\delta \in (0, 1]$ so that

$$w\, \mathrm{dist}[x, X^*] \le \|H(x)\| \quad \text{for all } x \in \mathscr{B}(x^*, \delta) \cap \Omega.$$

To enable subproblems simpler than those we use in this paper (see Sect. 3), projected Levenberg–Marquardt steps can be employed, see [14] and [16] (within a local phase). Then, however, the corresponding local convergence analysis requires that the inequality in Assumption 2 holds for all $x \in \mathscr{B}(x^*, \delta)$ although the set Ω in problem (1) is a proper subset of $\mathbb{R}^n$. This implies that, in a neighborhood of x^*, problem (1) is equivalent to $H(x) = 0$. Therefore, depending on the problem, Assumption 2 can be significantly weaker than assuming that the inequality holds for all $x \in \mathscr{B}(x^*, \delta)$. It is shown in [10, Section 3] that such projected Levenberg–Marquardt steps can be regarded as solutions of unconstrained Levenberg–Marquardt subproblems (2) with an appropriate definition of the perturbation π. Due to this, the advantage from enlarging the regularization parameter, i.e., that the level of inexactness increases without destroying quadratic convergence (see the discussion in Sect. 1), applies to the projected steps in [14,16].

The following lemma shows that the mapping $p \mapsto X(p)$ is upper Lipschitz continuous at x^* under the previous assumptions.

Lemma 1 *Let Assumptions 1 and 2 be satisfied. Then, there are $\mu > 0$ and $\delta_* > 0$ so that*

$$X(p) \cap \mathscr{B}(x^*, \delta_*) \subseteq X^* + \mu\|p\|\mathscr{B}$$

for all $p \in \mathbb{R}^n$.

Proof Let us first fix some $\delta_* \in (0, \delta]$ and let $p \in \mathbb{R}^n$ be arbitrarily chosen. If $X(p) \cap \mathscr{B}(x^*, \delta_*)$ is empty, nothing has to be shown. Otherwise, let x_p denote any element of $X(p) \cap \mathscr{B}(x^*, \delta_*)$. Then, there is $y_p \in \mathrm{N}_\Omega(x_p)$ so that

$$p = \nabla H(x_p) H(x_p) + y_p. \tag{7}$$

Since X^* is nonempty and closed there is $\hat{x}_p \in X^*$ such that $\|x_p - \hat{x}_p\| = \mathrm{dist}[x_p, X^*]$. This implies

$$\hat{x}_p \in \mathscr{B}\left(x^*, 2\delta_*\right). \tag{8}$$

Assumption 1 ensures (due to Taylor's formula) that there is $C_1 > 0$ so that, for any $z \in X^* \cap \mathscr{B}(x^*, 2\delta_*)$ and any $x \in \mathscr{B}(x^*, 2\delta_*)$,

$$\|H(x) + \nabla H(x)^\top (z - x)\|^2 \leq C_1 \|z - x\|^4.$$

Therefore, by (8), $\hat{x}_p \in X^*$, and $x_p \in \mathscr{B}(x^*, \delta)$, we have

$$\begin{aligned} r\left(x_p\right) &:= \|H\left(x_p\right) + \nabla H\left(x_p\right)^\top \left(\hat{x}_p - x_p\right)\|^2 \leq C_1 \|\hat{x}_p - x_p\|^4 \\ &= C_1 \,\mathrm{dist}\left[x_p, X^*\right]^4. \end{aligned} \tag{9}$$

For the left hand side we obtain

$$\begin{aligned} r\left(x_p\right) = {} & 2\left(\hat{x}_p - x_p\right)^\top \nabla H\left(x_p\right) H\left(x_p\right) + \left(\hat{x}_p - x_p\right)^\top \nabla H\left(x_p\right) \nabla H\left(x_p\right)^\top \left(\hat{x}_p - x_p\right) \\ & + \|H\left(x_p\right)\|^2 \end{aligned}$$

and

$$2\left(\hat{x}_p - x_p\right)^\top \nabla H\left(x_p\right) H\left(x_p\right) + \|H\left(x_p\right)\|^2 \leq r\left(x_p\right).$$

Since $\hat{x}_p, x_p \in \Omega$ and $y_p \in \mathrm{N}_\Omega(x_p)$ we have $(\hat{x}_p - x_p)^\top y_p \leq 0$ and, with (7), get

$$\begin{aligned} p^\top \left(\hat{x}_p - x_p\right) &= \left(\hat{x}_p - x_p\right)^\top \nabla H\left(x_p\right) H\left(x_p\right) + \left(\hat{x}_p - x_p\right)^\top y_p \\ &\leq \frac{1}{2}\left(r\left(x_p\right) - \|H\left(x_p\right)\|^2\right). \end{aligned}$$

By (9) and Assumption 2, this implies

$$p^\top \left(\hat{x}_p - x_p\right) \leq \frac{1}{2}\left(C_1 \,\mathrm{dist}\left[x_p, X^*\right]^4 - w^2 \,\mathrm{dist}\left[x_p, X^*\right]^2\right).$$

Due to $x_p \in \mathscr{B}(x^*, \delta_*)$, choosing $\delta_* \in (0, \delta]$ sufficiently small leads to $C_1 \,\mathrm{dist}[x_p, X^*]^2 \leq \frac{1}{2} w^2$ and

$$p^\top \left(\hat{x}_p - x_p\right) \leq -\frac{1}{4} w^2 \,\mathrm{dist}\left[x_p, X^*\right]^2$$

follows. Thus, dividing this inequality by $\|\hat{x}_p - x_p\|$ (which is the same as dist$[x_p, X^*]$) yields

$$\frac{1}{4}w^2 \operatorname{dist}\left[x_p, X^*\right] \leq \|p\|.$$

Setting $\mu := 4w^{-2}$ completes the proof. □

Lemma 1 implies that $X(0) \cap \mathscr{B}(x^*, \delta_*) \subseteq X^*$. This means that, although the generalized Eq. (5) is only a necessary optimality condition for the minimization problem (4), all vectors that satisfy this necessary condition and are not too far away from x^* also solve the minimization problem.

3 Subproblems and method

Given some $s \in \Omega$ the algorithm we are going to analyze solves the following subproblem

$$0 \in \nabla_x \psi\,(x, s) + \mathrm{N}_\Omega\,(x)\,, \tag{10}$$

where $\psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\psi\,(x, s) := \frac{1}{2}\|H\,(s) + \nabla H\,(s)^\top\,(x - s)\,\|^2 + \frac{1}{2}\alpha\,(s)\,\|x - s\|^2 + \pi\,(s)^\top\,(x - s)$$

and $\pi(s) \in \mathbb{R}^n$ denotes a perturbation that enables inexactness within the subproblem. The level of inexactness is described by Assumption 3 below. It depends on the exponent $\beta \in (0, 2]$ in the following definition of the regularization parameter.

$$\alpha(s) := \begin{cases} \|H(s)\|^\beta & \text{if } s \notin X^*, \\ 1 & \text{if } s \in X^*. \end{cases} \tag{11}$$

Note that the convergence theorem does not change if an appropriate less restrictive rule is used to define $\alpha(s)$ for $s \notin X^*$. For example,

$$\alpha_0\|H(s)\|^\beta \leq \alpha(s) \leq \alpha_1\|H(s)\|^\beta$$

for some constants $\alpha_1 \geq \alpha_0 > 0$ can be employed.

By the convexity of Ω and $\psi(\cdot, s)$ the generalized Eq. (10) is equivalent to the quadratic minimization problem

$$\psi\,(x, s) \to \min_x \quad \text{s.t.} \quad x \in \Omega. \tag{12}$$

Due to $\alpha(s) > 0$ the function $\psi(\cdot, s)$ is strongly convex. Since the convex set Ω is also nonempty and closed problem (12) has a unique solution denoted by $x(s)$.

Assumption 3 There is $c_\pi > 0$ so that

$$\|\pi(s)\| \leq c_\pi \|H(s)\|^{\beta+1} \quad \text{for all } s \in \mathscr{B}(x^*, \delta).$$

Given a starting point $x^0 \in \Omega$ the constrained Levenberg–Marquardt method generates a sequence $\{x^k\}$ of iterates defined by

$$x^{k+1} := x(x^k) \quad k = 0, 1, 2, \ldots \tag{13}$$

Roughly speaking, the following lemma shows that locally the length of a Levenberg–Marquardt step is bounded by a constant times the distance of the current iterate to X^*. The lemma is related to similar results in literature, for example see [2,14]. However, since these results are either proved for the unconstrained case or do not deal with inexactness a proof will be given.

Lemma 2 *Let Assumptions 1–3 be satisfied. Then, there is* $\kappa > 0$ *so that*

$$\|x(s) - s\| \leq \kappa \operatorname{dist}[s, X^*] \qquad \textit{for all } s \in \mathscr{B}(x^*, \delta).$$

Proof Recall that X^* is nonempty and bounded. Thus, for any $s \in \mathbb{R}^n$, there is $\hat{s} \in X^*$ with

$$\|s - \hat{s}\| = \operatorname{dist}[s, X^*]. \tag{14}$$

As $x(s)$ solves the minimization problem (12) we have

$$\frac{1}{2}\alpha(s)\|x(s) - s\|^2 + \pi(s)^\top (x(s) - s) \leq \psi(x(s), s) \leq \psi\left(\hat{s}, s\right). \tag{15}$$

From Assumption 1 and Taylor's formula it follows that there is $L_0 > 0$ so that

$$\|H(s)\| = \|H(s) - H\left(\hat{s}\right)\| \leq L_0\|\hat{s} - s\| \tag{16}$$

and

$$\|H(s) + \nabla H(s)^\top \left(\hat{s} - s\right)\|^2 \leq L_0\|\hat{s} - s\|^4$$

hold for all $s \in \mathscr{B}(x^*, \delta)$. With the latter inequality, (15) implies

$$\|x(s) - s\|^2 \leq \alpha(s)^{-1}\Big(L_0\|\hat{s} - s\|^4 + 2\|\hat{s} - s\|\|\pi(s)\| + 2\|x(s) - s\|\|\pi(s)\|\Big) + \|\hat{s} - s\|^2$$

for all $s \in \mathscr{B}(x^*, \delta)$. Because $\alpha(s)^{-1}\|\pi(s)\| \leq c_\pi \|H(s)\|$ holds by (11) and Assumption 3, we obtain from (11), Assumption 2, (16), and (14) that

$$\|x(s) - s\|^2 \leq \operatorname{dist}\left[s, X^*\right]^2 \left(\operatorname{dist}\left[s, X^*\right]^{2-\beta} L_0 w^{-\beta} + 2c_\pi L_0 + 1\right)$$
$$+ 2c_\pi L_0 \|x(s) - s\| \operatorname{dist}\left[s, X^*\right].$$

Since $\operatorname{dist}[s, X^*] \leq \delta$ is valid for all $s \in \mathscr{B}(x^*, \delta)$ there is $L_1 > 0$ so that

$$\|x(s) - s\|^2 - 2c_\pi L_0 \operatorname{dist}\left[s, X^*\right] \|x(s) - s\| - L_1 \operatorname{dist}\left[s, X^*\right]^2 \leq 0$$

holds for all $s \in \mathscr{B}(x^*, \delta)$. Now, it can easily be seen that the inequalities

$$h^2 - 2c_\pi L_0 \operatorname{dist}\left[s, X^*\right] h - L_1 \operatorname{dist}\left[s, X^*\right]^2 \leq 0, \quad h \geq 0$$

are satisfied both if and only if

$$0 \leq h \leq \left(c_\pi L_0 + \sqrt{c_\pi^2 L_0^2 + L_1}\right) \operatorname{dist}\left[s, X^*\right].$$

Hence, the assertion of the lemma follows for $\kappa := c_\pi L_0 + \sqrt{c_\pi^2 L_0^2 + L_1}$. □

4 Local convergence

The local convergence results obtained below could also be derived by means of the general iterative framework for generalized equations in [8]. For simplicity we however decided to provide proofs that are more instructive for Levenberg–Marquardt methods. To proceed let us first define

$$\Delta(x, s) := \nabla H(x) H(x) - \nabla_x \psi(x, s)$$

for all $s, x \in \mathbb{R}^n$. By the definition of ψ this can be rewritten as

$$\Delta(x, s) = (\nabla H(x) - \nabla H(s)) H(x) + \nabla H(s) \left(H(x) - H(s) - \nabla H(s)^\top (x - s)\right)$$
$$- \alpha(s)(x - s) - \pi(s). \tag{17}$$

Moreover, in the remaining part of the paper

$$\tau := \min\{\beta + 1, 2\}$$

will be used to describe a convergence rate.

Lemma 3 *Let Assumptions 1–3 be satisfied. Then, there is $C > 0$ so that*

$$\|\Delta(x(s), s)\| \leq C \operatorname{dist}\left[s, X^*\right]^\tau \quad \textit{for all } s \in \mathcal{B}(x^*, \delta_0),$$

where $\delta_0 := (\kappa + 1)^{-1}\delta$.

Proof Recall that, for any $s \in \mathbb{R}^n$, $\hat{s} \in X^*$ is defined by (14). From Assumptions 1 and 3, Lemma 2, and (17) we obtain that, for all $s, x \in \mathcal{B}(x^*, \delta)$ with $s \notin X^*$, there is $L \geq L_0 >$ so that

$$\|H(s)\| = \|H(s) - H(\hat{s})\| \leq L \operatorname{dist}\left[s, X^*\right], \tag{18}$$

$$\begin{aligned} \|H(x(s))\| &= \|H(x(s)) - H(\hat{s})\| \\ &\leq L\|x(s) - \hat{s}\| \\ &\leq L\left(\|x(s) - s\| + \|s - \hat{s}\|\right) \\ &\leq L(\kappa + 1) \operatorname{dist}\left[s, X^*\right], \end{aligned} \tag{19}$$

and

$$\begin{aligned} \|\Delta(x, s)\| \leq\ & L\|H(x)\|\|x - s\| + L\|x - s\|^2 \\ & + \|H(s)\|^\beta \|x - s\| + c_\pi \|H(s)\|^{\beta+1}. \end{aligned} \tag{20}$$

By Lemma 2, we have for all $s \in \mathcal{B}(x^*, \delta_0)$ that

$$\begin{aligned} \|x(s) - x^*\| &\leq \|x(s) - s\| + \|s - x^*\| \\ &\leq \kappa \operatorname{dist}[s, X^*] + \delta_0 \leq (\kappa + 1)\delta_0 = \delta. \end{aligned} \tag{21}$$

Therefore, the variable x within (20) can be replaced by $x(s)$. This together with (18), (19), and Lemma 2 leads to

$$\|\Delta(x(s), s)\| \leq C \operatorname{dist}[s, X^*]^\tau \tag{22}$$

for all $s \in \mathcal{B}(x^*, \delta_0) \setminus X^*$, where $C := L^2\kappa(\kappa + 1) + L\kappa^2 + L^\beta\kappa + c_\pi L^{\beta+1}$. Since $x(s) = s$ for $s \in X^*$, it follows by (17) and Assumption 3 that (22) also holds for $s \in X^*$. □

Lemma 4 *Let Assumptions 1–3 be satisfied. Then, there are $\hat{C} > 0$ and $\delta_\diamond > 0$ so that*

$$\operatorname{dist}[x(s), X^*] \leq \hat{C} \operatorname{dist}[s, X^*]^\tau \leq \frac{1}{2} \operatorname{dist}[s, X^*] \quad \textit{for all } s \in \mathcal{B}(x^*, \delta_\diamond).$$

Proof Let $s \in \mathbb{R}^n$ be arbitrary but fixed. Then, subproblem (10) is equivalent to the generalized equation

$$\Delta(x, s) \in \nabla H(x) H(x) + \mathrm{N}_\Omega(x).$$

Hence, $x(s)$ is the unique solution of this equation. Therefore, $x(s)$ is also a solution of the perturbed generalized Eq. (6) with $p := \Delta(x(s), s)$, i.e., $x := x(s)$ solves

$$\Delta(x(s), s) \in \nabla H(x) H(x) + \mathrm{N}_\Omega(x).$$

This means

$$x(s) \in X(\Delta(x(s), s)).$$

To apply Lemma 1 we also need that $x(s) \in \mathscr{B}(x^*, \delta_*)$. To this end and for later use let us define

$$\delta_\diamond := \min\left\{\delta_0, \frac{\delta_*}{\kappa+1}, (2\mu C)^{\frac{-1}{\tau-1}}\right\}.$$

Now, for $s \in \mathscr{B}(x^*, \delta_\diamond)$, similar to (21) we obtain from Lemma 2 that $\|x(s)-x^*\| \leq \delta_*$. Thus, $x(s) \in X(\Delta(x(s), s)) \cap \mathscr{B}(x^*, \delta_*)$ is valid for all $s \in \mathscr{B}(x^*, \delta_\diamond)$. Lemma 1, Lemma 3, and the definition of $\delta_\diamond$ imply

$$\operatorname{dist}\left[x(s), X^*\right] \leq \mu \|\Delta(x(s), s)\| \leq \mu C \operatorname{dist}\left[s, X^*\right]^\tau \leq \frac{1}{2} \operatorname{dist}\left[s, X^*\right]$$

for all $s \in \mathscr{B}(x^*, \delta_\diamond)$. With $\hat{C} := \mu C$ the proof is complete. □

Lemma 5 (Lemma 2.9 in [10]) *Let* $\{w^k\} \subset \mathbb{R}^n, r_k \subset [0, \infty)$ *be sequences, and* $r \in [0, 1)$, $R > 0$ *numbers so that, for* $k = 0, 1, 2, \ldots,$

$$\|w^k - w^0\| \leq r_0 \frac{R}{1-r} \tag{23}$$

implies

$$r_{k+1} \leq r\, r_k \quad \textit{and} \quad \|w^{k+1} - w^k\| \leq R r_k. \tag{24}$$

Then, $\{r^k\}$ *converges to* 0 *and* $\{w^k\}$ *converges to some* $\hat{w} \in \mathbb{R}^n$. *If, for some* $t > 1$ *and* $c > 0$,

$$r_{k+1} \leq c r_k^t \quad \textit{and} \quad \|\hat{w} - w^k\| \geq r_k \tag{25}$$

is satisfied for $k = 0, 1, 2, \ldots$ *then* $\{w^k\}$ *converges to* $\hat{w}$ *with the Q-order of* t.

Theorem 1 *Let Assumptions* 1–3 *be satisfied and let* $\{x^k\}$ *be a sequence generated by the Levenberg–Marquardt method* (13). *Then, there is* $\varepsilon > 0$ *so that* $x^0 \in \mathscr{B}(x^*, \varepsilon)$ *implies that the sequence* $\{x^k\}$ *converges to some* $\hat{x} \in X^*$ *with the Q-order* τ.

Proof To apply Lemma 5 we set $w^k := x^k$ and $r_k := \operatorname{dist}[x^k, X^*]$ for $k = 0, 1, 2, \ldots$ and

$$r := \frac{1}{2}, \quad R := \kappa, \quad c := \hat{C}, \quad t := \tau = \min\{\beta + 1, 2\}.$$

Then, assuming (23) provides

$$\|x^k - x^*\| \le \|x^k - x^0\| + \|x^0 - x^*\| \le 2\kappa \operatorname{dist}[x^0, X^*] + \|x^0 - x^*\|.$$

Setting $\varepsilon := (2\kappa + 1)^{-1}\delta_\diamond$ we have

$$\|x^k - x^*\| \le (2\kappa + 1)\|x^0 - x^*\| \le (2\kappa + 1)\varepsilon \le \delta_\diamond \le \delta.$$

Thus, Lemmas 4 and 2 can be applied for $s := x^k$. This leads to

$$\operatorname{dist}\left[x^{k+1}, X^*\right] \le \frac{1}{2} \operatorname{dist}\left[x^k, X^*\right] \quad \text{and} \quad \|x^{k+1} - x^k\| \le \kappa \operatorname{dist}\left[x^k, X^*\right],$$

i.e., (24) is valid. Therefore, by Lemma 5, the sequence $\{\operatorname{dist}[x^k, X^*]\}$ converges to 0 and $\{x^k\}$ converges to some $\hat{x} \in X^*$.

Thanks to Lemma 4 and since $\|\hat{x} - x^k\| \ge \operatorname{dist}[x^k, X^*]$ is obviously valid for $k = 0, 1, 2, \ldots,$ we see that (25) is satisfied. Thus, Lemma 5 guarantees that $\{x^k\}$ converges to $\hat{x}$ with the Q-order τ. □

5 Sharpness of the level of inexactness

In this section we show that, in general, the level of inexactness given by Assumption 3 cannot be increased without reducing the convergence rate $\tau = \min\{\beta + 1, 2\}$ of the Levenberg–Marquardt method (13). To this end let us consider the simple example, where $\Omega := \mathbb{R}^2$ and $H : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$H(x) := \|x\|^2 - 1. \tag{26}$$

Obviously, the solution set X^* of $H(x) = 0$ is the unit sphere and Assumptions 1 is valid. Moreover, since

$$\operatorname{dist}[x, X^*] \le (\|x\| + 1) \operatorname{dist}[x, X^*] = (\|x\| + 1)|\|x\| - 1| = |\|x\|^2 - 1| = |H(x)|,$$

Assumption 2 is satisfied for $w = 1$ and any $\delta > 0$ regardless which solution $x^* \in X^*$ is taken. For later use let, for some $\rho \in (0, \frac{1}{4}]$,

$$S(\rho) := \left\{x \in \mathbb{R}^2 \backslash X^* \mid \operatorname{dist}\left[x, X^*\right] \le \rho\right\}$$

denote a set surrounding the solution set X^*. Then, according to (11), the regularization parameter is

$$\alpha(s) = |H(s)|^{\beta} \tag{27}$$

for any $s \in S(\rho)$ with some $\beta \in (0, 2]$. Let us assume that, for some $\eta \in (0, 1)$, the perturbation vector is given by

$$\pi(s) := -\sigma(s)\, s + \pi(s)_{\perp} \in \mathbb{R}^2 \tag{28}$$

with

$$\sigma(s) := |H(s)|^{\beta+\eta}, \quad \|\pi(s)_{\perp}\| = |H(s)|^{\beta+\eta}, \quad \text{and} \quad s^{\top}\pi(s)_{\perp} = 0. \tag{29}$$

Then, $\|\pi(s)\| \leq \sqrt{5}|H(s)|^{\beta+\eta}$ is valid for all $s \in \mathscr{B}(0, 2)$. Note that for $\eta \geq 1$ Assumption 3 would be satisfied for any $x^* \in X^*$ and the level of inexactness is not increased in comparison with Theorem 1. Therefore, only $\eta \in (0, 1)$ need to be considered.

Now, to estimate the influence of the inexactness on the convergence rate of the Levenberg–Marquardt method, we analyze the ratio

$$\frac{\operatorname{dist}\left[x(s), X^*\right]}{\operatorname{dist}\left[s, X^*\right]^{\nu}} \tag{30}$$

for $\|s\| \to 1$ and $\nu > 1$. By the definition of H in (26) we know that

$$\operatorname{dist}\left[x(s), X^*\right] = |\|x(s)\| - 1| = \frac{\left|\|x(s)\|^2 - 1\right|}{\|x(s)\| + 1}. \tag{31}$$

Recall that, due to $\Omega = \mathbb{R}^2$, the solution $x(s)$ of the subproblem (12) is the unique solution of the linear system

$$\left(\nabla H(s) \nabla H(s)^T + \alpha(s) I\right)(x - s) = -\nabla H(s) H(s) - \pi(s).$$

For our example (26), we obtain

$$\left(4ss^{\top} + \alpha(s) I\right)(x(s) - s) = -2(\|s\|^2 - 1)s - \pi(s). \tag{32}$$

Then, with (32), (28), and (29), a simple calculation shows that, for any $s \in S(\rho)$,

$$x(s) = \frac{2\|s\|^2 + 2 + \alpha(s) + \sigma(s)}{4\|s\|^2 + \alpha(s)} s - \frac{1}{\alpha(s)} \pi(s)_{\perp}$$

and, by (27) and (29),

$$\|x(s)\|^2 = \left(\frac{2\|s\|^2 + 2 + \alpha(s) + \sigma(s)}{4\|s\|^2 + \alpha(s)}\right)^2 \|s\|^2 + |H(s)|^{2\eta}$$

are valid. This implies that $\|x(s)\| + 1$ is bounded above on $S(\frac{1}{4})$. Moreover, taking into account (31), Assumption 2, and (29) as well as $\beta \in (0, 2]$ and $\eta \in (0, 1)$, we obtain after a longer calculation that there are $c_1, c_2 > 0$ and $\hat{\rho} \in (0, \frac{1}{4}]$ so that

$$\text{dist}\left[x(s), X^*\right] \geq c_1 \,\text{dist}\left[s, X^*\right]^{\beta+\eta} + c_2 \,\text{dist}\left[s, X^*\right]^{2\eta}$$

holds for all $s \in S(\hat{\rho})$. Hence, if

$$\nu > \nu(\beta, \eta) := \min\{\beta + \eta, 2\eta\},$$

the ratio (30) tends to ∞ for $\|s\| \to 1$, i.e., in the above example the Levenberg–Marquardt method cannot converge to a solution with an order of ν. Since

$$\nu(\beta, \eta) = \min\{\eta + \beta, 2\eta\} < \min\{1 + \beta, 2\} = \tau,$$

the convergence rate τ in Theorem 1 cannot be guaranteed if the level of inexactness is larger than required in Assumption 3.

6 Final remarks

Let the regularization parameter $\alpha(s)$ be given by (11), i.e., $\alpha(s) = \|H(s)\|^\beta$. For $\beta = 1$, Theorem 1 tells us that the local convergence rate of the constrained Levenberg–Marquardt method (13) is $\tau = 2$, where an inexactness level of $\|H(s)\|^2$ is enabled. For the unconstrained case ($\Omega = \mathbb{R}^n$) a corresponding result has been shown recently in [10]. However, for the constrained Levenberg–Marquardt methods in [14,19], $\beta = 2$ is required to achieve a quadratic rate. With the present paper, the local behavior of an inexact constrained Levenberg–Marquardt method is analyzed for the first time. A sharp maximal level of inexactness depending on $\beta \in (0, 2]$ is derived. In particular, this shows that $\beta = 1$ should be used to allow a maximal inexactness without reducing the quadratic rate. Numerical results in [10] underline this in the unconstrained case. If $\beta = 2$ is considered the level of inexactness given in [2] for the unconstrained case could be improved from $\|H(s)\|^4$ to $\|H(s)\|^3$ and also holds if constraints are present. According to Theorem 1, $\beta \in (0, 1)$ implies a convergence rate less than 2. Nevertheless, this choice of β may be useful to allow larger perturbations.

Acknowledgments The work by Roger Behling was carried out during his one year stay at the Institute of Numerical Mathematics of the Technische Universität Dresden. He was supported by a scholarship of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, and financial aid by the Deutscher Akademischer Austauschdienst (DAAD), Germany. The authors are grateful to the referees for their comments.

References

1. Bellavia, S., Morini, B.: Subspace trust-region methods for large bound-constrained nonlinear equations. SIAM J. Numer. Anal. **44**, 1535–1555 (2006)
2. Dan, H., Yamashita, N., Fukushima, M.: Convergence properties of the inexact Levenberg–Marquardt method under local error bound conditions. Optim. Methods Softw. **17**, 605–626 (2002)
3. Dembo, R.S., Eisenstat, S.C., Steihaug, T.: Inexact Newton methods. SIAM J. Numer. Anal. **19**, 400–408 (1982)
4. Facchinei, F., Fischer, A., Piccialli, V.: Generalized Nash equilibrium problems and Newton methods. Math. Progr. **117**, 163–194 (2009)
5. Facchinei, F., Kanzow, C.: A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. Math. Progr. **76**, 493–512 (1997)
6. Fan, J., Pan, J.: Inexact Levenberg–Marquardt method for nonlinear equations. Discret. Contin. Dyn. System.—Ser. B **4**, 1223–1232 (2004)
7. Fan, J., Yuan, Y.: On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption. Computing. **74**, 23–39 (2005)
8. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Progr. **94**, 91–124 (2002)
9. Fischer, A., Shukla, P.K.: A Levenberg–Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. **36**, 643–646 (2008)
10. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg–Marquardt methods. Optimization. **59**, 273–287 (2010)
11. Floudas, C.A., Pardalos, P.M. (eds.): Encyclopedia of Optimization. 2nd edn. Springer, Berlin (2008)
12. Kanzow, C., Petra, S.: On a semismooth least squares formulation of complementarity problems with gap reduction. Optim. Methods Softw. **19**, 507–525 (2004)
13. Kanzow, C., Petra, S.: Projected filter trust region methods for a semismooth least squares formulation of mixed complementarity problems. Optim. Methods Softw. **22**, 713–735 (2007)
14. Kanzow, C., Yamashita, N., Fukushima, M.: Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. **172**, 375–397 (2004)
15. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. **2**, 164–168 (1944)
16. Macconi, M., Morini, B., Porcelli, M.: Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities. Appl. Numer. Math. **59**, 859–876 (2009)
17. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**, 431–441 (1963)
18. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg–Marquardt method. Computing. **15**(Suppl), 239–249 (2001)
19. Zhang, J.-L.: On the convergence properties of the Levenberg–Marquardt method. Optimization. **52**, 739–756 (2003)
20. Zhu, D.: Affine scaling interior Levenberg–Marquardt method for bound-constrained semismooth equations under local error bound conditions. J. Comput. Appl. Math. **219**, 198–215 (2008)