Mathematical Programming manuscript No.
(will be inserted by the editor)

The effect of calmness on the solution set of systems of nonlinear equations

Roger Behling · Alfredo Iusem

Received: date / Accepted: date

Abstract We address the problem of solving a continuously differentiable nonlinear system of equations under the condition of calmness. This property, also called upper Lipschitz-continuity in the literature, can be described by a local error bound and is being widely used as a regularity condition in optimization. Indeed, it is known to be significantly weaker than classic regularity assumptions that imply that solutions are isolated. We prove that under this condition, the rank of the Jacobian of the function that defines the system of equations must be locally constant on the solution set. As a consequence, we prove that locally, the solution set must be a differentiable manifold. Our results are illustrated by examples and discussed in terms of their theoretical relevance and algorithmic implications.

Keywords Calmness · upper-Lipschitz continuity · nonlinear equations · error bound · Levenberg-Marquardt.

Mathematics Subject Classification (2000) 65H10 · 90C30

1 Introduction

Let us consider the following system of nonlinear equations

$$H(x) = 0, \tag{1}$$

R. Behling
Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil
E-mail: rogerb@impa.br
The work of this author was supported by CNPq

A. Iusem
Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil
E-mail: iusp@impa.br
The work of this author was partially supported by CNPq grant no. 301280/86

where $H : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function. We denote by X^* the set of solutions of (1) and suppose it is nonempty.

We will deal with the notion of *calmness* of a system like (1), which we define next.

The results we are going to present do not require a specific norm. Nevertheless, throughout the paper, let $\|\cdot\|$ denote the Euclidean norm or the associated matrix norm and $\mathcal{B}(x, \delta)$, the closed ball centered in x with radius δ.

We now formally present the definition of a calm problem.

Definition 1 We say that Problem (1) is calm at $x^* \in X^*$ if there exist $\omega > 0$ and $\delta > 0$ so that

$$\omega \operatorname{dist}[x, X^*] \leq \|H(x)\|,$$

for all $x \in \mathcal{B}(x^*, \delta)$, where $\operatorname{dist}[a, A]$ denotes the Euclidean distance from a point a to a set $A \subset \mathbb{R}^n$.

The notion of calmness (see [15], Chapter 8, Section F) should be understood as a regularity condition which extends the classical concept of a *regular* problem at a solution point x^* in the smooth and square case, i.e. with $m = n$, meaning that the Jacobian matrix of H at x^* is nonsingular. When the problem is regular at x^* in this sense, x^* is necessarily an isolated solution. Besides its applicability for nonsmooth and rectangular systems, the notion of calmness encompasses situations with non-isolated solutions. It is easy to check that in the particular case of smooth systems with $m = 1$, calmness implies that directional derivatives in directions normal to the solution set are nonnull. At the same time, the notion of calmness is powerful enough as to allow the extension of a large array of results which were previously known to hold only for regular systems, in the classical sense, and hence it became quite popular. This condition, that is also called upper-Lipschitz continuity (e.g. [14]), is described by a local error bound. If H is affine, Hoffman's Lemma (see [8]) guarantees that this error bound always holds and is global.

One area where this notion turned out to be quite useful is the study of the convergence properties of iterative methods for solving systems of nonlinear equations, e.g. Levenberg-Marquardt type methods (see [11],[12], and Section 4). Among the papers which use the notion of calmness with this purpose we mention [1], [3], [4], [6], [7], [10], [17] and [18]. Calmness also has implications in connection with constraint qualifications, Karush-Kuhn-Tucker systems and second order optimality conditions (see [9]).

For equivalent definitions of calm problems see [15]. Thanks to the Implicit Function Theorem we know that full rank systems are calm. We will see in this paper that in some sense, calm problems do not go that much beyond systems of equations with full rank Jacobians.

In this paper we will establish that the local error bound described by calmness, together with the continuous differentiability of H, imply that the rank of the Jacobian is locally constant on the solution set of system (1).

The paper is organized as follows. In Section 2 we prove our main result. In Section 3 we present some corollaries and examples. We end up with some remarks on the algorithmical relevance of our theorem in Section 4.

2 Our main result

We assume from now on that H is continuously differentiable. $J_H(x) \in \mathbb{R}^{m\times n}$ will denote the Jacobian matrix of H evaluated at a point x, which has as columns the gradients of the component functions of H. Before proving our theorem, we present a result by Fischer ([5], Corollary 2), that will play a crucial role in our proof.

Lemma 1 *Assume that Problem* (1) *is calm at* x^*. *Then, there exist* $\bar{\delta} > 0$ *and* $\bar{w} > 0$ *so that*

$$\bar{\omega}\,\mathrm{dist}[x, X^*] \leq \|J_H(x)H(x)\|,$$

for all $x \in \mathcal{B}(x^*, \bar{\delta})$.

This lemma says that the problem $J_H(x)H(x) = 0$ inherits the calmness of $H(x) = 0$, and also that these systems must be equivalent in a neighborhood of x^*.

We now arrive at the main result of this paper.

Theorem 1 *Assume that Problem* (1) *is calm at* x^*. *Then, there exists* $\delta^\diamond > 0$ *so that* $\mathrm{rank}(J_H(x)) = \mathrm{rank}(J_H(x^*))$ *for all* $x \in \mathcal{B}(x^*, \delta^\diamond) \cap X^*$.

Proof Suppose that there exists a sequence $\{x^k\} \subset X^*$ with $x^k \to x^*$ so that

$$\mathrm{rank}\left(J_H(x^k)\right) \neq \mathrm{rank}\left(J_H(x^*)\right),$$

for all k. From the continuity of the Jacobian we can assume, without loss of generality, that

$$\mathrm{rank}\left(J_H(x^k)\right) > \mathrm{rank}\left(J_H(x^*)\right).$$

Let us consider now the singular value decomposition (see, e.g., [2], p. 109) of $J_H(x)$,

$$J_H(x)^\top = U_x \Sigma_x V_x^\top,$$

where $U_x \in \mathbb{R}^{m\times m}$ and $V_x \in \mathbb{R}^{n\times n}$ are orthogonal and Σ_x is the m by n diagonal matrix $\mathrm{diag}(\sigma_1(x), \sigma_2(x), ..., \sigma_{r_x}(x), 0, ..., 0)$ with positive singular values $\sigma_1(x) \geq \sigma_2(x) \geq ... \geq \sigma_{r_x}(x) > 0$. We recall that these singular values are the square roots of the nonzero eigenvalues of the matrix $J_H(x)J_H(x)^\top$. Note that $r_x \geq 0$ indicates the rank of $J_H(x)$. According to these definitions we have that $r_{x^*} < r_{x^k}$. In order to facilitate the notation we omit x in some manipulations and set $r := r_{x^*}$ and $r_k := r_{x^k}$. Now define $v^k := V_k e_{r+1}$, where

$$e_{r+1} := [0...0\underbrace{1}_{r+1}0...0]^\top \in \mathbb{R}^n.$$

Then, for all k we have that $\|v^k\| = 1$ and

$$v^k \perp \text{Kernel}\left(J_H(x^k)^\top\right) = \text{Span}\{V_k e_{r_k+1}, ..., V_k e_n\}. \tag{2}$$

We introduce now an auxiliary operator. Let $\Sigma_* := \Sigma_{x^*}$ and define $T : \mathbb{R}^n \to \mathbb{R}^{n\times m}$ as

$$T(x) := V_x \Sigma_*^\top U_x^\top.$$

Using the notation $U_k := U_{x^k}, V_k := V_{x^k}$, we get, for all k,

$$T(x^k)J_H(x^k)^\top v^k = V_k\Sigma_*^\top U_k^\top U_k\Sigma_k V_k^\top v^k = V_k\Sigma_*^\top \Sigma_k e_{r+1} = V_k\Sigma_k^\top \Sigma_* e_{r+1} = 0. \tag{3}$$

Using Lemma 1 we conclude that for all $x \in \mathcal{B}(x^*, \bar{\delta})$ it holds that

$$\begin{aligned}\text{dist}[x, X^*] &\le \bar{\omega}\|J_H(x)H(x)\| \\ &\le \bar{\omega}\|\left(J_H(x) - T(x^k)\right)H(x)\| + \bar{\omega}\|T(x^k)H(x)\| \\ &\le \bar{\omega}\|J_H(x) - T(x^k)\|\|H(x)\| + \bar{\omega}\|T(x^k)H(x)\|. \end{aligned}\tag{4}$$

From the differentiability of H we know that there exist $\check{\delta} > 0$ and a Lipschitz constant $L > 0$ so that

$$\|H(x) - H(y)\| \le L\|x - y\|,$$

for all $x, y \in \mathcal{B}(x^*, \check{\delta})$. Let $\bar{x} \in X^*$ denote a solution that satisfies $\|x - \bar{x}\| = \text{dist}[x, X^*]$ for an arbitrary point x. Then, there exists a positive constant $\delta' \le \check{\delta}$ so that for all $x \in \mathcal{B}(x^*, \delta')$ we have $\bar{x} \in \mathcal{B}(x^*, \delta')$ and

$$\|H(x)\| = \|H(x) - H(\bar{x})\| \le L\|x - \bar{x}\| = L\text{dist}[x, X^*]. \tag{5}$$

The continuity of J_H implies that for some positive $\tilde{\delta} \le \delta'$ we have that

$$\|J_H(x) - T(x^k)\| \le \frac{1}{2\bar{\omega}L}, \tag{6}$$

whenever $x, x^k \in \mathcal{B}(x^*, \tilde{\delta})$. Thus, in this ball, (4), (5) and (6) lead to

$$\text{dist}[x, X^*] \le 2\bar{\omega}\|T(x^k)H(x)\|. \tag{7}$$

Using Taylor's formula,

$$\|H(x^k + tv^k) - H(x^k) - tJ_H(x^k)^\top v^k\| = o(t),$$

with $\lim_{t\to 0} o(t)/t = 0$. Then, in view of (3) and (7), there exist $\bar{k} > 0$ and $\bar{t} > 0$, so that for all $k > \bar{k}$ and $0 < t < \bar{t}$ we have that

$$\begin{aligned}\frac{1}{2\bar{\omega}}\text{dist}[x^k + tv^k, X^*] &\le \|T(x^k)H(x^k + tv^k)\| \\ &= \|T(x^k)\left(H(x^k + tv^k) - H(x^k) - tJ_H(x^k)^\top v^k\right)\| \\ &\le \|T(x^k)\|\|H(x^k + tv^k) - H(x^k) - tJ_H(x^k)^\top v^k\| \\ &= \sigma_1(x^*)o(t).\end{aligned}$$

On the other hand, taking (5) into account and using the previous inequality, we conclude that

$$\begin{aligned}\|J_H(x^k)^\top v^k\| &\le \frac{o(t)}{t} + \frac{\|H(x^k+tv^k) - H(x^k)\|}{t}\\ &\le \frac{o(t)}{t} + \frac{L}{t}\mathrm{dist}[x^k+tv^k, X^*]\\ &\le (1+2\bar{\omega}\sigma_1(x^*)L)\,\frac{o(t)}{t}.\end{aligned}$$

Taking the limit when $t \to 0^+$ we get

$$J_H(x^k)^\top v^k = 0,$$

which contradicts (2). □

3 Some related results

In this section we present some results related to Theorem 1 and discuss examples that illustrate the relevance of the assumption of calmness, and the sharpness of the conclusions that we obtain. Our first example shows that under calmness, the rank of the Jacobian must be locally constant only on the solution set, but not at other points.

Example 1 Consider the function $H : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$H(x_1, x_2) = \begin{bmatrix} x_2 \\ x_2^2 \exp(x_1^2) \end{bmatrix}.$$

Thus, the Jacobian is given by

$$J_H(x_1, x_2)^\top = \begin{bmatrix} 0 & 1 \\ 2x_1x_2^2\exp(x_1^2) & 2x_2\exp(x_1^2) \end{bmatrix}.$$

Define the sequence $\{x^k\} \not\subset X^*$ so that $x^k := (1, 1/k)$ with $k > 0$ and consider the solution $x^* := (1, 0)$. Obviously, $x^k \to x^*$ as $k \to \infty$ and $\mathrm{rank}\left(J_H(x^k)\right) = 2 \neq 1 = \mathrm{rank}\left(J_H(x^*)\right)$ for all $k > 0$. Nevertheless, one can easily check that $H(x) = 0$ is calm at x^*.

One may also ask if the converse of Theorem 1 is true, i.e., if constant rank on the solution set implies calmness. The simple example next shows that the answer to this question is negative.

Example 2 Let $H : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$H(x_1, x_2, x_3) = \begin{bmatrix} x_1 + x_2 \\ x_3^2 \end{bmatrix}.$$

The rank of the Jacobian on the solution set is always 1, but it is clear that the second component violates the error bound in Definition 1 around any solution.

In the convergence analysis of the Levenberg-Marquardt methods in [4] and [18] it was assumed, without loss of generality, that the Jacobian of H at x^* had at least one positive singular value. This assumption is rigorously supported by the following result.

Proposition 1 *Assume that Problem* (1) *is calm at* $x^* \in X^*$ *and that* $J_H(x^*) = 0$*. Then, there exists* $\delta_1 > 0$ *so that* $H(x) = 0$ *for all* $x \in \mathcal{B}(x^*, \delta_1)$*.*

Proof Lemma 1 together with the assumption on $J_H(x^*)$ imply that there exists $\bar{\delta}$ such that

$$\begin{aligned} \bar{\omega}\, \mathrm{dist}[x, X^*] &\leq \|J_H(x)H(x)\| \\ &= \|(J_H(x) - J_H(x^*))H(x)\| \\ &\leq \|J_H(x) - J_H(x^*)\| \|H(x)\| \end{aligned} \tag{8}$$

for all $x \in \mathcal{B}(x^*, \bar{\delta})$. For a given x in this ball, take now $x' \in X^*$ such that $\|x - x'\| = \mathrm{dist}[x, X^*]$, and let L be the Lipschitz constant of H. It follows from (8) and the fact that x' belongs to X^* that

$$\bar{\omega}\, \mathrm{dist}[x, X^*] \leq \|J_H(x) - J_H(x^*)\| \|H(x)\| = \|J_H(x) - J_H(x^*)\| \|H(x) - H(x')\| \leq$$

$$L\|J_H(x) - J_H(x^*)\| \|x - x'\| = L\|J_H(x) - J_H(x^*)\| \mathrm{dist}[x, X^*] \tag{9}$$

for all $x \in \mathcal{B}(x^*, \delta^\diamond)$, where $\delta^\diamond \leq \bar{\delta}$ is sufficiently small. By continuity of J_H, there exists $\delta_1 \leq \delta^\diamond$ such that

$$\|J_H(x) - J_H(x^*)\| \leq \frac{\bar{\omega}}{2L}, \tag{10}$$

for all $x \in \mathcal{B}(x^*, \delta_1)$. Combining (9) and (10) we get $(\bar{\omega}/2)\mathrm{dist}[x, X^*] \leq 0$ for all $x \in \mathcal{B}(x^*, \delta_1)$, and hence the whole ball is contained in X^*. □

The next example suggests that complementarity type equations tend not to be calm at points that do not satisfy strict complementarity.

Example 3 Consider $H : \mathbb{R}^2 \to \mathbb{R}$ so that

$$H(x_1, x_2) := x_1 x_2.$$

The solution set correspondent to Problem (1) is $X^* = \{x \in \mathbb{R}^2 | x_1 = 0 \text{ or } x_2 = 0\}$ and the Jacobian is given by

$$J_H(x_1, x_2)^\top = \begin{bmatrix} x_2 \; x_1 \end{bmatrix}.$$

Note that the rank of the Jacobian is 0 at $x^* := (0, 0)$ but it is equal to 1 at any other solution. Since the function is not identically zero in any neighborhood of x^*, Corollary 1 implies that Problem (1) cannot be calm at x^*. Nevertheless, note that in this example the systems $J_H(x)H(x) = 0$ and $H(x) = 0$ are equivalent around x^*. This means that the equivalence between these two systems of equations does not imply calmness.

We now show that calm problems are not that far away from full rank problems. This is formally described by the next lemma, where we will rewrite Problem (1) as an equivalent full rank system of equations, also calm.

Theorem 2 *Assume that Problem* (1) *is calm at* x^*. *Then, there exists a continuously differentiable mapping* $\bar{H} : \mathbb{R}^n \to \mathbb{R}^r$, *with* $r := \operatorname{rank}(J_H(x^*))$, *so that the problem*

$$\bar{H}(x) = 0$$

is calm at x^* *and locally equivalent to* $H(x) = 0$. *Moreover, there exists* $\delta_2 > 0$ *so that* $\operatorname{rank}(J_{\bar{H}}(x)) = r$, *for all* $x \in \mathcal{B}(x^*, \delta_2)$.

Proof Lemma 1 implies that for all $x \in \mathcal{B}(x^*, \bar{\delta})$ we have that

$$\begin{aligned} \operatorname{dist}[x, X^*] &\leq \bar{\omega} \|J_H(x)H(x)\| \\ &\leq \bar{\omega} \| \left(J_H(x) - J_H(x^*)\right) H(x)\| + \bar{\omega}\|J_H(x^*)H(x)\| \\ &\leq \bar{\omega}\|J_H(x) - J_H(x^*)\|\|H(x)\| + \bar{\omega}\|J_H(x^*)H(x)\|. \end{aligned} \tag{11}$$

Then, using the local Lipschitz continuity of H and the continuity of J_H, we have

$$\|J_H(x) - J_H(x^*)\|\|H(x)\| \leq \frac{2}{\bar{\omega}} \operatorname{dist}[x, X^*],$$

in $\mathcal{B}(x^*, \hat{\delta})$, for $\hat{\delta} > 0$ sufficiently small. This inequality combined with (11) implies that

$$\operatorname{dist}[x, X^*] \leq 2\bar{\omega}\|J_H(x^*)H(x)\|,$$

for all $x \in \mathcal{B}(x^*, \hat{\delta})$. On the other hand, defining

$$\bar{H}(x) := \left(\sigma_1(x^*) \left(U_*^\top H(x)\right)_1, ..., \sigma_r(x^*) \left(U_*^\top H(x)\right)_r\right),$$

with $\sigma_1(x^*), ..., \sigma_r(x^*)$ and U_* as in Theorem 1, we get

$$\|J_H(x^*)H(x)\| = \|V_* \Sigma_*^\top U_*^\top H(x)\| = \|\bar{H}(x)\|.$$

Obviously, $\bar{H} : \mathbb{R}^n \to \mathbb{R}^r$ is continuously differentiable and rank $\left(J_{\bar{H}(x)}\right) \leq r$ for all x. Furthermore, $\bar{H}(x) = 0$ is calm at x^* and equivalent to $H(x) = 0$ in a neighborhood of x^*.

In order to complete the proof we just need to show that $\operatorname{rank}(J_{\bar{H}}(x^*)) = r$, since the rank cannot diminish locally. It can be easily checked that it suffices to prove that

$$\operatorname{Kernel}\left(J_{\bar{H}}(x^*)^\top\right) \subset \operatorname{Kernel}\left(J_H(x^*)^\top\right). \tag{12}$$

So, let us prove this inclusion. We know that there exist $\delta_3 > 0$ and $\hat{\omega}$ so that

$$\operatorname{dist}[x, X^*] \leq \hat{\omega}\|\bar{H}(x)\|,$$

for all $x \in \mathcal{B}(x^*, \delta_3)$. Take $u \in \text{Kernel}\,(J_{\bar{H}}(x^*))$ with $\|u\| = 1$. Then, for every $t > 0$ sufficiently small we have that

$$\begin{aligned} o(t) &= \|\bar{H}(x^* + tu) - \bar{H}(x^*) - tJ_{\bar{H}}(x^*)^\top u\| \\ &= \|\bar{H}(x^* + tu)\| \\ &\geq \frac{1}{\hat{\omega}}\text{dist}[x^* + tu, X^*] \\ &\geq \frac{1}{L\hat{\omega}}\|H(x^* + tu)\|, \end{aligned} \tag{13}$$

where the last inequality follows from the local Lipschitz continuity of H, with $L > 0$ as in Theorem 1. On the other hand, from Taylor's formula we know that

$$\lim_{t \to 0} \frac{1}{t}\|H(x^* + tu) - H(x^*) - tJ_H(x^*)^\top u\| = 0.$$

This, together with (13), leads to

$$J_H(x^*)^\top u = 0,$$

which implies the inclusion (12). Therefore, there exists $\delta_2 > 0$ so that

$$\text{rank}\,(J_{\bar{H}}(x)) = r,$$

for all $x \in \mathcal{B}(x^*, \delta_2)$. □

The next corollary characterizes the geometry of the solution set of a calm problem.

Corollary 1 *Assume that Problem* (1) *is calm at* x^**. Then* X^* *is locally, a differentiable manifold of codimension* $r := \text{rank}(J_H(x^*))$.

Proof Given a continuously differentiable system $H(x) = 0$, such that the rank of $J_H(x)$ is constant on a neighborhood of a zero $\tilde{x}$ of H, it is well known that the set of solutions $\{x \in \mathbb{R}^n : H(x) = 0\}$ is locally a differentiable manifold (see, e.g., Proposition 12 in [16], p. 65). In view of Lemma 2 we conclude that this result applies to the set of zeroes of $\bar{H}$. The statement follows then from the local equivalence of $H(x) = 0$ and $\bar{H}(x) = 0$ at x^*, also proved in Theorem 2. □

Due to this corollary one can easily see that sets like $X^* := \{x \in \mathbb{R}^3 | x_2 = 0 \text{ or } x_1^2 + (x_2 - 1)^2 + x_3^2 = 1\}$ (the union of a sphere and a hyperplane with nonempty interception) cannot represent the solution set of a calm problem, though X^* is the solution set of the differentiable system $H(x) = 0$ with

$$H(x_1, x_2, x_3) = x_2\left(x_1^2 + (x_2 - 1)^2 + x_3^2 - 1\right).$$

Another direct but interesting consequence of Corollary 1 is given next.

Corollary 2 *Assume that Problem* (1) *is calm at* x^**. Then, there cannot exist a sequence of isolated solutions converging to* x^**.*

Proof Such a solution set cannot be a differentiable manifold. □

The example we will present now shows how delicate Corollary 2 is.

Example 4 Let $H : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$H(x_1, x_2) := \begin{cases} \left(x_2, x_2 - x_1^2 \sin(\frac{1}{x_1})\right), & \text{if } x_1 \neq 0; \\ (x_2, x_2), & \text{if } x_1 = 0. \end{cases}$$

In this example we have exactly one non-isolated solution, namely $x^* := (0, 0)$, and x^* is the limit of the isolated solutions

$$x^k := \left(\frac{1}{2k\pi}, 0\right),$$

with k integer and $|k| \to \infty$. One can also observe that Problem (1) is calm at x^* and that

$$J_H(x^k) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

for all $k > 0$. Nevertheless, the Jacobian of H at x^* is given by

$$J_H(x^*) = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}.$$

This change of rank does not contradict Corollary 2, since the Jacobian of H is not continuous at x^*. In this case, modifying a little bit the example in order to have continuity of the Jacobian and calmness is an impossible task. In fact, these two properties conflict with each other in the following sense. If one replaces x_1^2 by something smoother, like x_1^β, with $\beta > 2$, one gets continuity of the Jacobian but loses calmness.

Before closing the section we discuss one last example.

Example 5 Consider $H : \mathbb{R}^3 \to \mathbb{R}^2$ so that

$$H(x_1, x_2, x_3) := \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1^2 + x_3^2 - 1 \end{bmatrix},$$

The solution set X^* associated to (1) is the intersection of two perpendicular cylinders. The Jacobian of H is given by

$$J_H(x_1, x_2, x_3)^\top = \begin{bmatrix} 2x_1 & 2x_2 & 0 \\ 2x_1 & 0 & 2x_3 \end{bmatrix}.$$

The rank of the Jacobian is 2 at any solution except at $x := (1, 0, 0)$ and $x := (-1, 0, 0)$, where it is 1. Therefore, Problem (1) is not calm at these two solutions. But what makes this example illustrative is the fact that the equivalence of $J_H(x)H(x) = 0$ and $H(x) = 0$ is destroyed at $(1, 0, 0)$ and $(-1, 0, 0)$. Indeed, the solution set of $J_H(x)H(x) = 0$ is a surface while the solution set of $H(x) = 0$ is the union of two perpendicular ellipses that intercept each other at $(1, 0, 0)$ and $(-1, 0, 0)$.

4 Our theorem and iterative algorithms for solving systems of nonlinear equations

Although Theorem 1 seems to be just a result of Analysis, and refers specifically to the geometry of solution sets of calm problems, it echoes in practical algorithms for solving Problem (1). Apparently, our results suggest that one should not be that preoccupied with the magnitude of the regularization parameter in unconstrained Levenberg-Marquardt methods. In order to explain this better let us describe a Levenberg-Marquardt iteration.

Interpret $s \in \mathbb{R}^n$ as the current iterate. Then, the Levenberg-Marquardt method demands solution of the following subproblem:

$$\min_{d \in \mathbb{R}^n} \|H(s) + \nabla J_H(s)^\top d\|^2 + \alpha(s)\|d\|^2, \tag{14}$$

where $\alpha(s) > 0$ is a regularization parameter. If we set this parameter equal to 0 and consider the minimum norm solution of (14), we recover the classical Gauss-Newton method. It is known that for calm problems, the local convergence rate of Gauss-Newton methods is superlinear (or quadratic) if the rank of the Jacobian is constant in a whole neighborhood of a solution. In other words, under constant rank, the Levenberg-Marquardt regularization is needless. Of course one can easily construct functions where the Levenberg-Marquardt parameter has to be precisely chosen in order to maintain fast local convergence. In Example 1, for instance, quadratic convergence of the Levenberg-Marquardt method is only achieved if $\alpha(s)$ is chosen so that it remains proportional to $\|H(s)\|^\beta$, with $\beta \in [1, 3]$. Nevertheless, in view of Lemma 2, such problems are kind of artificial. In fact, the numerical results in [7] have shown that the Levenberg-Marquardt parameter could be chosen with significant freedom without changing the accuracy of the method. The constant rank on the solution set of calm problems might also be the reason for the efficiency of the conjugate gradient method in solving the Levenberg-Marquardt subproblems, stated in the same reference.

Acknowledgements The authors would like to thank the two anonymous referees for their valuable comments, Prof. Alexey Izmailov for coming up with the function in Example 4 and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support.

References

1. 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).
2. Demmel, J.W.: Applied Numerical Linear Algebra. SIAM, Philadelphia (1997).
3. Fan, J., Pan, J.: Inexact Levenberg-Marquardt method for nonlinear equations. Discret. Contin. Dyn. Systems – Ser. B **4**, 1223–1232 (2004).
4. Fan, J., Yuan, Y.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Comput. **74**, 23–39 (2005).

5. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002).
6. Fischer, A., Shukla, P.K.: A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. **36**, 643–646 (2008).
7. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg-Marquardt methods. Optim. **59**, 273–287 (2010).
8. Hoffman, A.J.: On approximate solutions of systems of linear inequalities. J. Research of the National Bureau of Standards **49**, 263–265 (1952).
9. Izmailov A.F., Solodov M.V.: Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. Math. Program. **95**, 631-650 (2003).
10. 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).
11. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. **2**, 164–168 (1944).
12. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. **11**, 431–441 (1963).
13. Pang J.S.: Error bounds in mathematical programming. Math. Program. **79**, 299–332 (1997).
14. Robinson, S.M.: Generalized equations and their solutions, Part II: Applications to nonlinear programming. Math. Program. **19**, 200–221 (1982).
15. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin, (1998).
16. Spivak M.: A Comprehensive Introduction to Differential Geometry, Volume 1. Publish or Perish, Berkeley (1979).
17. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. Comput. **15** (Suppl.), 239–249 (2001).
18. Zhang, J.-L.: On the convergence properties of the Levenberg-Marquardt method. Optim. **52**, 739–756 (2003).