

A Levenberg-Marquardt method with approximate projections

R. Behling · A. Fischer · M. Herrich · A. Iusem ·
Y. Ye

Received: 30 November 2012 / Published online: 8 June 2013
© Springer Science+Business Media New York 2013

Abstract The projected Levenberg-Marquardt method for the solution of a system of equations with convex constraints is known to converge locally quadratically to a possibly nonisolated solution if a certain error bound condition holds. This condition turns out to be quite strong since it implies that the solution sets of the constrained and of the unconstrained system are locally the same.

Under a pair of more reasonable error bound conditions this paper proves R-linear convergence of a Levenberg-Marquardt method with approximate projections. In this way, computationally expensive projections can be avoided. The new method is also applicable if there are nonsmooth constraints having subgradients. Moreover, the pro-

Honoring Masao Fukushima at the occasion of his 65th birthday for his extraordinary contribution to the theory and methods of Continuous Optimization.

R. Behling (✉) · Y. Ye
Department of Management Science and Engineering, Stanford University, Stanford,
CA 94305-4121, USA
e-mail: rogerbehling@gmail.com

Y. Ye
e-mail: yinyu-ye@stanford.edu

A. Fischer · M. Herrich
Institute of Numerical Mathematics, Department of Mathematics, Technische Universität Dresden,
01062 Dresden, Germany

A. Fischer
e-mail: Andreas.Fischer@tu-dresden.de

M. Herrich
e-mail: Markus.Herrich@tu-dresden.de

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

jected Levenberg-Marquardt method is a special case of the new method and shares its R-linear convergence.

Keywords Constrained equation · Levenberg-Marquardt method · Approximate projection · Non-isolated solution · Error bound

1 Introduction

Let us consider the constrained system of equations

$$H(x) = 0, \quad x \in \Omega, \quad (1)$$

where $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is continuously differentiable with a locally Lipschitz continuous Jacobian $J_H : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$. The set $\Omega \subseteq \mathbb{R}^n$ is closed and convex. Throughout the paper, $O := \{x \in \mathbb{R}^n \mid H(x) = 0\}$ denotes the set of zeros of H . Moreover, the solution set $X := O \cap \Omega$ of problem (1) is assumed to be nonempty. For later use let x^* denote an arbitrary but fixed element of X .

In order to describe the aim and developments of this paper in more detail we need to recall some basic facts and recent work on Levenberg-Marquardt methods. The basic idea of regularizing the Gauss-Newton method goes back to Levenberg [18] and Marquardt [19]. For the unconstrained case ($\Omega = \mathbb{R}^n$), in each step of a Levenberg-Marquardt method, the subproblem

$$\min_{x \in \mathbb{R}^n} \varphi_0(x, s) \quad (2)$$

has to be solved for some $s \in \mathbb{R}^n$ (current iterate), where $\varphi_0 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is given by

$$\varphi_0(x, s) := \frac{1}{2} \|H(s) + J_H(s)(x - s)\|^2 + \frac{1}{2} \alpha(s) \|x - s\|^2$$

with a regularization parameter $\alpha(s) > 0$. Here and throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. Note that (2) has a strongly convex objective, is equivalent to a system of linear equations, and possesses a unique solution.

In 2001 Yamashita and Fukushima [21] opened a new direction of research for Levenberg-Marquardt methods. In the unconstrained case they showed that the regularization parameter can be chosen in such a way that superlinear convergence to nonisolated solutions is attainable. The main ingredient for obtaining this result is the following error bound condition on a ball

$$\mathcal{B}(x^*, \delta) := \{x \in \mathbb{R}^n \mid \|x - x^*\| \leq \delta\}.$$

By $\text{dist}[x, W] := \inf\{\|x - w\| \mid w \in W\}$ the distance of $x \in \mathbb{R}^n$ to a nonempty set $W \subset \mathbb{R}^n$ is denoted.

Condition 1 *There exist $\omega > 0$ and $\delta > 0$ such that*

$$\omega \text{dist}[x, O] \leq \|H(x)\| \quad \text{for all } x \in \mathcal{B}(x^*, \delta).$$

This condition is significantly weaker than the classical nonsingularity assumption on $J_H(x^*)$ (for $m = n$). In particular, Condition 1 does not imply that x^* is an isolated solution. If Condition 1 holds then the system $H(x) = 0$ is called *calm at x^** . For other and related notions see [12, 20]. If H is affine (see [16]) or if the rank of $J_H(x^*)$ is full, then $H(x) = 0$ is calm at x^* . However, Condition 1 encompasses also situations in which $J_H(x^*)$ is not full rank. For example, for the calm system $H(x_1, x_2) := (x_2, x_2 \exp(x_1))^T = 0$, the rank of J_H is 1 on the solution set and 2 otherwise. For a detailed analysis of consequences of Condition 1 see [1, 3].

In the last decade several works continued the ideas of [21]. Among them are papers on a more robust regularization [11, 12], on inexact Levenberg-Marquardt methods [4, 9, 10, 14], on applications [7, 13], and on constrained Levenberg-Marquardt methods [2, 17, 22]. Instead of (2), the latter methods employ more complex subproblems. In particular,

$$\min_{x \in \Omega} \varphi_0(x, s) \tag{3}$$

was suggested by Kanzow, Yamashita, and Fukushima [17]. They proved local quadratic convergence by means of the following error bound condition.

Condition 2 *There exist $\omega > 0$ and $\delta > 0$ such that*

$$\omega \operatorname{dist}[x, X] \leq \|H(x)\| \quad \text{for all } x \in \mathcal{B}(x^*, \delta) \cap \Omega.$$

This condition has the advantage that it is restricted to points in Ω only. It turned out that this error bound condition (together with further assumptions) can be quite useful for dealing with problem (1) if H is nonsmooth; see [6]. Since solving the constrained subproblem (3) can be computationally challenging or, at least, is more expensive than solving the unconstrained minimization problem (2), a projected Levenberg-Marquardt method is suggested in [17] (see also [8, 13]). Per step, the latter method requires the solution of the unconstrained minimization problem (2) plus a projection onto Ω . If the projection is cheap then the computational costs per step compare favorably to the costs of solving (3). However, in order to prove local quadratic convergence, the following quite strong error bound condition is needed.

Condition 3 *There exist $\omega > 0$ and $\delta > 0$ such that*

$$\omega \operatorname{dist}[x, X] \leq \|H(x)\| \quad \text{for all } x \in \mathcal{B}(x^*, \delta).$$

Obviously, Condition 3 is much stronger than Condition 2. In particular, Condition 3 implies that

$$X \cap \mathcal{B}(x^*, \delta) = O \cap \mathcal{B}(x^*, \delta)$$

holds for some sufficiently small $\delta > 0$, i.e., the set Ω can be disregarded in a neighborhood of x^* .

The aim of this paper is to design a Levenberg-Marquardt type method that does not need such a strong error bound condition but still employs the simple subproblems (2). Instead, we will use Conditions 1 and 2 as a pair. The price we have to pay

is that only local R-linear convergence can be shown, but still to a possibly nonisolated solution of problem (1). To replace the exact projections onto the set Ω the new method allows *approximate projections* without losing the linear rate. Often, such approximate projections can be done in a computationally cheap way and can be applied even if the set Ω is described by nonsmooth inequalities. Besides approximate projections the new method is able to cope with a certain level of inexactness in solving the subproblem (2).

The paper is organized as follows. In Sect. 2 we formally introduce the Levenberg-Marquardt algorithm with approximate projections. After presenting the assumptions, Sect. 3 is devoted to the convergence analysis of the new algorithm. The paper is completed by final remarks in Sect. 4. They include discussions on the sharpness of the inexactness level and of the convergence rate, as well as a numerical example.

2 Levenberg-Marquardt algorithm with approximate projections

In this section we present a Levenberg-Marquardt method with approximate projections and a possibly inexact solution of the subproblems (2). The next two subsections collect the necessary ingredients for defining the algorithm in Sect. 2.3.

2.1 Inexact subproblems

The Levenberg-Marquardt subproblems we are going to use enable an inexact solution and read as

$$\min_{x \in \mathbb{R}^n} \varphi(x, s), \tag{4}$$

where $\varphi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by

$$\varphi(x, s) := \frac{1}{2} \|H(s) + J_H(s)(x - s)\|^2 + \frac{1}{2} \alpha(s) \|x - s\|^2 - \pi(s)^\top (x - s) \tag{5}$$

with $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$\alpha(s) := \begin{cases} \|H(s)\| & \text{if } s \in \mathbb{R}^n \setminus O, \\ 1 & \text{if } s \in O. \end{cases} \tag{6}$$

Moreover, $\pi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes a function that, for some $\varkappa, \eta > 0$, satisfies

$$\|\pi(s)\| \leq \varkappa \|H(s)\|^{2+\eta} \tag{7}$$

for all $s \in \mathbb{R}^n$. The function π is not explicitly needed within the algorithm but is used to describe and analyze the possible level of inexactness. With the above definition of subproblem (4) we easily see that, for any $s \in \mathbb{R}^n$, this subproblem has a unique solution which we denote by $x(s)$. If $s \in O$, the definitions of the functions α and π imply $x(s) = s$. For any $s \in \mathbb{R}^n$, the solution $x(s)$ can be characterized by the necessary and sufficient optimality condition $\nabla_x \varphi(x, s) = 0$, i.e., by a linear system

of n equations with n variables. With the definition of the objective φ_0 in the exact Levenberg-Marquardt subproblem (2), we obtain

$$\nabla_x \varphi_0(x(s), s) = \pi(s)$$

Thus, $\pi(s)$ can be regarded as the residual of the optimality condition

$$\nabla_x \varphi_0(x, s) = 0$$

for subproblem (2) at $x = x(s)$.

2.2 Approximate projections

Projected Levenberg-Marquardt methods employ a projection step that maps an iterate $s \in \mathbb{R}^n$ onto its nearest point in the set Ω . In order to relax this projection step, we introduce the notion of an approximate projector.

Definition 1 Let $B \subseteq \mathbb{R}^n$ be a nonempty set. A function $\tilde{P}_\Omega : B \rightarrow \mathbb{R}^n$ is called an *approximate projector* defined on B and associated to Ω if the following properties are satisfied:

1. There exists a number $\gamma \in [0, 1)$ such that

$$\text{dist}[\tilde{P}_\Omega(x), \Omega] \leq \gamma \text{dist}[x, \Omega] \quad \text{for all } x \in B. \tag{8}$$

2. For any $x \in B$, it holds that

$$(x - \tilde{P}_\Omega(x))^\top (y - \tilde{P}_\Omega(x)) \leq 0 \quad \text{for all } y \in \Omega. \tag{9}$$

The vector $\tilde{P}_\Omega(x)$ is called an *approximate projection* of $x \in B$.

Property 2 in the previous definition means that, for a point $x \in B \setminus \Omega$, the approximate projection $\tilde{P}_\Omega(x)$ is the orthogonal projection of x onto some hyperplane h_x separating x from Ω .

Of course, the orthogonal projector onto Ω , $P_\Omega : \mathbb{R}^n \rightarrow \Omega$, with $P_\Omega(x)$ being the unique solution of the minimization problem

$$\min_{w \in \Omega} \|x - w\|,$$

is an approximate projector defined on \mathbb{R}^n associated to Ω . Indeed, for $\tilde{P}_\Omega = P_\Omega$, Property 1 in Definition 1 is satisfied with $\gamma = 0$ and Property 2 is well-known for the orthogonal projector. For some sets Ω , calculating the orthogonal projection of a given point $x \in \mathbb{R}^n$ is not very hard, for instance if Ω is a halfspace, or if it is described by box constraints. However, the projection task can become computationally expensive if Ω is a more general convex set. Fortunately, determining an approximate projection is possible with only little effort in most cases. For example, let us consider the important case in which Ω is described by general convex inequalities, i.e., $\Omega := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\}$, where $g_1, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$

are convex but not necessarily differentiable functions. Then, it is well-known that $g : \mathbb{R}^n \rightarrow \mathbb{R}$ defined by $g(x) := \max\{g_1(x), \dots, g_m(x)\}$ is convex and Ω can be written as $\Omega := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$. Therefore, without loss of generality, only the case $m = 1$ needs to be considered. The following proposition is based on [15] and provides a special approximate projector by means of subgradients of g , i.e., of elements of the subdifferential $\partial g(x)$ of g at x . Moreover, an explicit formula for computing an approximate projection is given. The computational expense of this approximate projection is low, provided that the computation of a subgradient of g is not too expensive.

Proposition 1 *Suppose that $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function and that there exists some $\hat{x} \in \mathbb{R}^n$ with $g(\hat{x}) < 0$ (Slater condition). Let $v : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denote a mapping with $v(x) \in \partial g(x)$ for any $x \in \mathbb{R}^n$. Moreover, let the mapping $\tilde{p} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be defined pointwise, with $\tilde{p}(x)$ being the unique solution of the problem*

$$\min_p \|p - x\| \quad \text{s.t.} \quad g(x) + v(x)^\top (p - x) \leq 0. \tag{10}$$

Then, the following assertions hold:

- (i) For any nonempty compact set $B \subset \mathbb{R}^n$, the mapping \tilde{p} is an approximate projector defined on B associated to the set $\Omega := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$.
- (ii) For any $x \in \mathbb{R}^n$, the solution $\tilde{p}(x)$ of (10) is given by

$$\tilde{p}(x) = \begin{cases} x - \frac{g(x)}{\|v(x)\|^2} v(x) & \text{if } g(x) > 0, \\ x & \text{if } g(x) \leq 0. \end{cases} \tag{11}$$

Proof

(i) The existence of $\gamma \in [0, 1)$ such that $\text{dist}[\tilde{p}(x), \Omega] \leq \gamma \text{dist}[x, \Omega]$ for all $x \in B$ follows from [15, Lemma 4], i.e., (8) holds for $\tilde{P}_\Omega := \tilde{p}$. Furthermore, for $x \in \mathbb{R}^n \setminus \Omega$, the solution $\tilde{p}(x)$ of (10) is the orthogonal projection of x onto the hyperplane

$$h_x := \{w \in \mathbb{R}^n \mid g(x) + v(x)^\top (w - x) = 0\}.$$

Since $v(x)$ is a subgradient of g in x , this hyperplane separates x from the convex set Ω and therefore (9) holds. Thus, the mapping \tilde{p} is an approximate projector defined on B associated to Ω .

- (ii) This assertion can be easily verified. □

2.3 The algorithm

Let \tilde{P}_Ω denote an approximate projector defined on B associated to Ω .

Since subproblem (4) has always a unique solution, Algorithm 1 is well-defined as long as the approximate projection $\tilde{P}_\Omega(x(x^k))$ is defined. The latter is the case if $x(x^k)$ belongs to B , see Definition 1. The restriction to B is only a formal aspect. For example let us consider the mapping \tilde{p} defined in Proposition 1. Due to assertion (ii) of this proposition, \tilde{p} is well-defined on \mathbb{R}^n and satisfies Property 2 of Definition 1. Property 1 of this definition holds at least on compact sets B . The latter, however, is sufficient for our local convergence analysis. Note that the iterates of a sequence $\{x^k\}$ generated by Algorithm 1 do not necessarily belong to Ω .

Algorithm 1 Levenberg-Marquardt algorithm with approximate projections (LMAP)

S1 Initialization	Choose $x^0 \in \mathbb{R}^n$. Set $k := 0$.
S2 Levenberg-Marquardt step	Compute $x(x^k)$ as solution of (4).
S3 Approximate projection	Set $x^{k+1} := \tilde{P}_\Omega(x(x^k))$.
S4 Update	Set $k := k + 1$ and go to S2.

3 Convergence analysis

In the first subsection we state the assumptions that will be used later on in Sect. 3.2 for analyzing the local convergence properties of Algorithm 1.

3.1 Assumptions

As in Sect. 1, let $x^* \in X$ denote an arbitrary but fixed solution of (1). Moreover, let $\delta \in (0, 1]$ denote the arbitrary but fixed radius of the ball $\mathcal{B}(x^*, \delta)$.

Assumption 1 The mapping $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable on $\mathcal{B}(x^*, \delta)$ and its Jacobian $J_H : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times n}$ is Lipschitz continuous on $\mathcal{B}(x^*, \delta)$.

To avoid Condition 3 we now present the pair of reasonable error bound conditions.

Assumption 2 There exists $\omega > 0$ such that

$$\omega \text{dist}[x, O] \leq \|H(x)\| \quad \text{for all } x \in \mathcal{B}(x^*, \delta) \tag{12}$$

and

$$\omega \text{dist}[x, X] \leq \|H(x)\| \quad \text{for all } x \in \mathcal{B}(x^*, \delta) \cap \Omega. \tag{13}$$

Roughly speaking, the error bound (12) (together with the Lipschitz continuity of H around x^*) implies that $\|H(x)\|$ grows locally like $\text{dist}[x, O]$, while (13) does not allow Ω to be in some sense tangent to O .

The next examples show that the error bounds (12) and (13) do not imply each other.

Example 1 Let $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ and Ω be given by $H(x_1, x_2) := x_2$ and

$$\Omega := \{(x_1, x_2)^\top \in \mathbb{R}^2 \mid g(x_1, x_2) \leq 0\} \quad \text{with } g(x_1, x_2) := \begin{cases} x_1^2 - x_2 & \text{if } x_1 \leq 0, \\ -x_2 & \text{if } x_1 > 0. \end{cases}$$

It is easy to see that the error bound (12) holds for any solution $x^* \in X = \{(x_1, 0)^\top \mid x_1 \geq 0\}$ and that (13) is violated for $x^* = 0$.

Example 2 Let $H : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ and Ω be given by

$$H(x_1, x_2, x_3) := (x_3, x_1 x_2)^\top \quad \text{and} \quad \Omega := \{(x_1, x_2, x_3)^\top \in \mathbb{R}^3 \mid x_3 \geq |x_2|\}.$$

In this example, (13) is satisfied while (12) is violated for $x^* = 0$.

3.2 Local analysis

In order to prove our main convergence theorem, we need some auxiliary results. Some of them are quite technical, whereas the first lemma is elementary.

Lemma 1 *Let Assumption 1 be satisfied. Then, there exists $L > 0$ such that, for all $x, s \in \mathcal{B}(x^*, \delta)$, the following inequalities hold:*

$$\|H(x) - H(s)\| \leq L\|x - s\|, \tag{14}$$

$$\|H(x) - H(s) - J_H(s)(x - s)\| \leq L\|x - s\|^2. \tag{15}$$

The next lemma lists some properties of an inexact unconstrained Levenberg-Marquardt step. Recall that, for a given point $s \in \mathbb{R}^n$, we denote by $x(s)$ the unique solution of subproblem (4).

Lemma 2 *Let Assumptions 1 and 2 be satisfied. Then, there exist $\delta_1 \in (0, \delta]$ and $L_1 > 0$ so that the inequalities*

$$\|x(s) - s\| \leq L_1 \operatorname{dist}[s, O], \tag{16}$$

$$\operatorname{dist}[x(s), O] \leq L_1 \operatorname{dist}[s, O]^2, \tag{17}$$

and

$$\|x(s) - x^*\| \leq \delta \tag{18}$$

hold for all $s \in \mathcal{B}(x^*, \delta_1)$. Moreover,

$$\lim_{\operatorname{dist}[s, O] \rightarrow 0} \frac{\|x(s) - s\|}{\operatorname{dist}[s, O]} = 1. \tag{19}$$

Proof The existence of $\delta_1 \in (0, \delta]$ and $L_1 > 0$ so that (16) and (17) hold is a well-known fact, see for example [2, 14]. Furthermore, using (16), we obtain (18) by

$$\|x(s) - x^*\| \leq \|x(s) - s\| + \|s - x^*\| \leq L_1 \operatorname{dist}[s, O] + \|s - x^*\| \leq (L_1 + 1)\delta_1 \leq \delta$$

for $\delta_1 > 0$ sufficiently small. So, only (19) needs a proof. To this end, let $s \in \mathcal{B}(x^*, \delta_1) \setminus O$ be arbitrarily chosen. The set O is closed by continuity of H , and thus there exists $s_O \in O$ with $\|s_O - s\| = \operatorname{dist}[s, O]$. Hence, we get

$$\begin{aligned} \|H(s)\| \|x(s) - s\|^2 &\leq 2\varphi(x(s), s) + 2\pi(s)^\top (x(s) - s) \\ &\leq 2\varphi(s_O, s) + 2\pi(s)^\top (x(s) - s) \\ &= \|H(s_O) - H(s) - J_H(s)(s_O - s)\|^2 + \|H(s)\| \|s_O - s\|^2 \\ &\quad - 2\pi(s)^\top (s_O - s) + 2\pi(s)^\top (x(s) - s) \end{aligned}$$

$$\begin{aligned}
 &\leq L^2 \|s_O - s\|^4 + \|H(s)\| \|s_O - s\|^2 + \\
 &\quad + 2\kappa \|H(s)\|^{2+\eta} (\|s_O - s\| + L_1 \operatorname{dist}[s, O]) \\
 &\leq \frac{L^2}{\omega} \|H(s)\| \operatorname{dist}[s, O]^3 + \|H(s)\| \operatorname{dist}[s, O]^2 \\
 &\quad + 2\kappa L^{1+\eta} (1 + L_1) \|H(s)\| \operatorname{dist}[s, O]^{2+\eta}, \tag{20}
 \end{aligned}$$

using (5) in the first inequality and the first equality, (15), (7) and (16) in the third inequality, and finally (12) and (14) in the fourth one. Dividing (20) by $\|H(s)\|$ and extracting the square root, we obtain on the one hand

$$\|x(s) - s\| \leq \sqrt{\frac{L^2}{\omega} \operatorname{dist}[s, O] + 1 + 2\kappa L^{1+\eta} (1 + L_1) \operatorname{dist}[s, O]^\eta \operatorname{dist}[s, O]}. \tag{21}$$

On the other hand, the triangle inequality and (17) yield

$$\operatorname{dist}[s, O] \leq \|s - x(s)\| + \operatorname{dist}[x(s), O] \leq \|x(s) - s\| + L_1 \operatorname{dist}[s, O]^2,$$

which leads to

$$\operatorname{dist}[s, O] - L_1 \operatorname{dist}[s, O]^2 \leq \|x(s) - s\|. \tag{22}$$

Now, dividing the inequalities (21) and (22) by $\operatorname{dist}[s, O]$, and taking the limit with $\operatorname{dist}[s, O] \rightarrow 0$, (19) follows. \square

In order to formulate the following results, we introduce the set-valued map $P_X : \mathbb{R}^n \rightrightarrows X$, defined as:

$$P_X(s) := \{x \in X \mid \|x - s\| = \operatorname{dist}[s, X]\}.$$

Lemma 3 *Let Assumptions 1 and 2 be satisfied. Then, for any $c \in (0, 1)$, there exists $\delta(c) > 0$ such that, for any $s \in \mathcal{B}(x^*, \delta(c))$,*

$$c \|x(s) - s_X\| \leq \|s - s_X\| \quad \text{for all } s_X \in P_X(s).$$

Proof Let us assume the contrary. Then there exist $c_1 \in (0, 1)$ and sequences $\{s^k\} \subset \mathbb{R}^n$ and $\{s_X^k\}$ with $s_X^k \in P_X(s^k)$ for all $k \in \mathbb{N}$ such that $\{s^k\}$ converges to x^* and

$$\|s^k - s_X^k\| < c_1 \|x(s^k) - s_X^k\| \quad \text{for all } k \in \mathbb{N}. \tag{23}$$

Without loss of generality we assume that, for some $\kappa \in (0, \delta_1]$,

$$\|s^k - x^*\| \leq \kappa \quad \text{for all } k \in \mathbb{N}. \tag{24}$$

Later on, κ will be restricted further. Obviously, (23) implies

$$\|x(s^k) - s_X^k\| > 0 \quad \text{for all } k \in \mathbb{N}. \tag{25}$$

For later use, we define

$$c_2 := \frac{1+c_1}{2} \quad \text{and} \quad c_3 := \sqrt{1 - \left(\frac{1-c_2}{c_2}\right)^2}.$$

Since $c_1 \in (0, 1)$, we easily obtain

$$\max\left\{c_1, \frac{1}{2}\right\} < c_2 < 1 < \frac{c_2}{c_1} \quad \text{and} \quad 0 < c_3 < 1 < \frac{1+c_3}{2c_3}. \quad (26)$$

Therefore, (19) in Lemma 2 implies, for $\kappa > 0$ sufficiently small,

$$\|s^k - x(s^k)\| \leq \min\left\{\frac{c_2}{c_1}, \frac{1+c_3}{2c_3}\right\} \text{dist}[s^k, O] \quad \text{for all } k \in \mathbb{N}. \quad (27)$$

For the sake of simplicity, we now fix $k \in \mathbb{N}$ and omit this index. Then, from (27) and (23), we obtain

$$\|s - x(s)\| \leq \frac{c_2}{c_1} \text{dist}[s, O] \leq \frac{c_2}{c_1} \text{dist}[s, X] < \frac{c_2}{c_1} c_1 \|x(s) - s_X\| = c_2 \|x(s) - s_X\|.$$

Thus, by (23) and the fact that $c_1 < c_2$, established in (26), we have

$$\|s - s_X\| < c_2 \|x(s) - s_X\| \quad \text{and} \quad \|s - x(s)\| < c_2 \|x(s) - s_X\|. \quad (28)$$

Now, taking into account (25),

$$\mathcal{L}(s) := \{x(s) + \lambda(s_X - x(s)) \mid \lambda \in \mathbb{R}\},$$

defines a straight line. With $z(s)$ denoting the orthogonal projection of s onto this line, we have that

$$(s - z(s))^\top (x(s) - s_X) = 0 \quad (29)$$

is valid. Since $(s, z(s), s_X)$ and $(s, x(s), z(s))$, respectively, are vertices of a right triangle with the right angle at $z(s)$,

$$\|z(s) - s_X\| \leq \|s - s_X\| \quad \text{and} \quad \|z(s) - x(s)\| \leq \|s - x(s)\| \quad (30)$$

follow. We conclude from (30) and (28) that the inequalities

$$\|z(s) - s_X\| < c_2 \|x(s) - s_X\| \quad \text{and} \quad \|z(s) - x(s)\| < c_2 \|x(s) - s_X\| \quad (31)$$

are satisfied. With $\lambda(s) \in \mathbb{R}$ given by

$$z(s) = x(s) + \lambda(s)(s_X - x(s)), \quad (32)$$

(31) implies

$$\lambda(s) \in (1 - c_2, c_2) \subset (0, 1), \quad (33)$$

i.e., $z(s)$ lies on the line $\mathcal{L}(s)$ between $x(s)$ and s_X . Therefore,

$$\|x(s) - z(s)\| > (1 - c_2)\|x(s) - s_X\|$$

follows and we get, using (28),

$$\|x(s) - z(s)\| > \frac{1 - c_2}{c_2}\|s - x(s)\|. \tag{34}$$

Moreover, since $z(s)$ belongs to the line $\mathcal{L}(s)$ we have, as in (29), that

$$(s - z(s))^\top (x(s) - z(s)) = 0$$

holds. This and (34) yield

$$\|s - z(s)\|^2 = \|s - x(s)\|^2 - \|x(s) - z(s)\|^2 < \|s - x(s)\|^2 - \left(\frac{1 - c_2}{c_2}\right)^2 \|s - x(s)\|^2.$$

With the above definition of c_3 , we have

$$\|s - z(s)\| < c_3\|s - x(s)\|. \tag{35}$$

From (27), we get

$$\|s - x(s)\| \leq \frac{1 + c_3}{2c_3} \text{dist}[s, O],$$

which implies, in view of (35),

$$\|s - z(s)\| < c_3 \frac{1 + c_3}{2c_3} \text{dist}[s, O] = \frac{c_3 + 1}{2} \text{dist}[s, O]. \tag{36}$$

The triangle inequality and (36) lead to

$$\text{dist}[s, O] \leq \|s - z(s)\| + \text{dist}[z(s), O] < \frac{c_3 + 1}{2} \text{dist}[s, O] + \text{dist}[z(s), O]. \tag{37}$$

Since (36) implies $\text{dist}[s, O] > 0$, dividing (37) by $\text{dist}[s, O]$, we obtain

$$1 < \frac{c_3 + 1}{2} + \frac{\text{dist}[z(s), O]}{\text{dist}[s, O]}. \tag{38}$$

Due to $(c_3 + 1)/2 < 1$ we would get the desired contradiction if the second term on the right hand side of (38) is sufficiently close to 0. We proceed to prove this fact. By (24), we have

$$\|s_X - x^*\| \leq \|s_X - s\| + \|s - x^*\| \leq 2\kappa.$$

Moreover, since $s \in \mathcal{B}(x^*, \delta_1)$, (16) in Lemma 2 yields

$$\|x(s) - x^*\| \leq \|x(s) - s\| + \|s - x^*\| \leq L_1 \text{dist}[s, O] + \kappa \leq (L_1 + 1)\kappa.$$

It follows from (36) and the fact that $c_3 < 1$, resulting from (26), that

$$\|z(s) - x^*\| \leq \|z(s) - s\| + \|s - x^*\| < \frac{c_3 + 1}{2} \text{dist}[s, O] + \kappa < 2\kappa.$$

Therefore, for $\kappa \in (0, \delta_1]$ sufficiently small, we have

$$s, s_X, x(s), z(s) \in \mathcal{B}(x^*, \delta_1). \tag{39}$$

Now, (16) in Lemma 2 and (28) imply that

$$\|x(s) - s_X\| \leq \|x(s) - s\| + \|s - s_X\| < L_1 \text{dist}[s, O] + c_2 \|x(s) - s_X\|$$

and, using $c_2 < 1$ from (26), we obtain

$$\|x(s) - s_X\| < \frac{L_1}{1 - c_2} \text{dist}[s, O]. \tag{40}$$

Taking into account (39), we get

$$\begin{aligned} \omega \text{dist}[z(s), O] &\leq \|H(z(s))\| \\ &= \|H(z(s)) - H(s_X)\| \\ &\leq \|J_H(s_X)(z(s) - s_X)\| + L \|z(s) - s_X\|^2 \\ &= (1 - \lambda(s)) \|J_H(s_X)(x(s) - s_X)\| \\ &\quad + (1 - \lambda(s))^2 L \|x(s) - s_X\|^2 \\ &\leq \|H(x(s)) - H(s_X)\| + 2L \|x(s) - s_X\|^2 \\ &\leq L \text{dist}[x(s), O] + 2L \|x(s) - s_X\|^2 \\ &< \left(LL_1 + \frac{2LL_1^2}{(1 - c_2)^2} \right) \text{dist}[s, O]^2, \end{aligned} \tag{41}$$

using the error bound (12) in Assumption 2 in the first inequality, (15) in Lemma 1 in the second one, (32) and (33) in the second equality, (15) again in the third inequality, (14) in the fourth one, and (17) in Lemma 2, together with (40), in the fifth one. Dividing (41) by $\omega \text{dist}[s, O]$, we have

$$\frac{\text{dist}[z(s), O]}{\text{dist}[s, O]} < \left(\frac{LL_1}{\omega} + \frac{2LL_1^2}{\omega(1 - c_2)^2} \right) \text{dist}[s, O].$$

Hence, without loss of generality, we can assume that κ in (24) is sufficiently small so that the right hand side of the previous inequality becomes strictly less than $(1 - c_3)/2$ and, thus, (38) yields a contradiction. \square

The following lemma provides the key result for proving local linear convergence of Algorithm 1 since it shows that locally one step of this algorithm provides a linear decrease w.r.t. the distance to the solution set X .

Lemma 4 *Let Assumptions 1 and 2 be satisfied and suppose that \tilde{P}_Ω is an approximate projector defined on $B \supseteq \mathcal{B}(x^*, 1)$. Then, there exist $\delta_2 \in (0, \delta_1]$ and $\tau \in [0, 1)$ such that*

$$x(s) \in B \quad \text{and} \quad \text{dist}[\tilde{P}_\Omega(x(s)), X] \leq \tau \text{dist}[s, X]$$

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

Proof Due to (18), we have $x(s) \in \mathcal{B}(x^*, 1) \subseteq B$ for all $s \in \mathcal{B}(x^*, \delta_1)$. This implies that $\tilde{P}_\Omega(x(s))$ is well-defined. By the definitions of $x(s)$ and \tilde{P}_Ω , we get

$$x(s) = s \quad \text{and} \quad \tilde{P}_\Omega(x(s)) = \tilde{P}_\Omega(s) = s \quad \text{for all } s \in X.$$

Therefore, the asserted inequality holds for all $s \in X$ for any $\tau \geq 0$. In order to prove this inequality for $s \notin X$ let us now proceed by contradiction, assuming that there exists a sequence $\{s^k\} \subset \mathbb{R}^n \setminus X$, converging to x^* , such that

$$\liminf_{k \rightarrow \infty} \frac{\text{dist}[\tilde{P}_\Omega(x(s^k)), X]}{\text{dist}[s^k, X]} \geq 1. \tag{42}$$

Without loss of generality we assume that, for some $\delta_2 \in (0, \delta_1]$,

$$\|s^k - x^*\| \leq \delta_2 \quad \text{for all } k \in \mathbb{N}. \tag{43}$$

Later, δ_2 will be further restricted. In order to obtain the desired contradiction, the remainder of the proof will be divided in two parts.

(a) In this part we will show that there exists $\rho \in [0, 1)$ such that

$$\|s_X^k - \tilde{P}_\Omega(x(s^k))\| \leq \rho \|x(s^k) - s_X^k\| \quad \text{for all } k \in \mathbb{N}, \tag{44}$$

where s_X^k denotes an arbitrary element of $P_X(s^k)$. For simplicity, we now fix $k \in \mathbb{N}$ and omit the index k . In order to show (44), we first prove that

$$x(s) \notin \Omega \quad \text{if } \delta_2 \in (0, \delta_1] \text{ is sufficiently small.} \tag{45}$$

Let us assume that (45) does not hold. It follows from (43) and (18), that $s \in \mathcal{B}(x^*, \delta_1)$ and $x(s) \in \mathcal{B}(x^*, \delta)$. Therefore, we can apply the error bound (13), (14), and (17) and obtain, for $x(s) \in \Omega$,

$$\begin{aligned} \omega \text{dist}[\tilde{P}_\Omega(x(s)), X] &\leq \|H(x(s))\| \\ &\leq L \text{dist}[x(s), O] \\ &\leq LL_1 \text{dist}[s, O]^2 \\ &\leq LL_1 \text{dist}[s, X]^2, \end{aligned}$$

which contradicts (42) for $\delta_2 > 0$ sufficiently small. Thus, (45) holds.

Now, we continue with the proof of (44). From the definition of the approximate projector \tilde{P}_Ω we have

$$(x(s) - \tilde{P}_\Omega(x(s)))^\top (s_X - \tilde{P}_\Omega(x(s))) \leq 0,$$

i.e., the angle at the vertex $\tilde{P}_\Omega(x(s))$ of the triangle $(x(s), s_X, \tilde{P}_\Omega(x(s)))$ is obtuse. Therefore, it follows from the cosine rule that

$$\|x(s) - \tilde{P}_\Omega(x(s))\| \leq \|x(s) - s_X\|, \quad \|s_X - \tilde{P}_\Omega(x(s))\| \leq \|x(s) - s_X\|, \quad (46)$$

and

$$\|x(s) - \tilde{P}_\Omega(x(s))\|^2 + \|s_X - \tilde{P}_\Omega(x(s))\|^2 \leq \|x(s) - s_X\|^2 \quad (47)$$

hold. Since, by (45), $x(s)$ does not belong to Ω for $\delta_2 > 0$ small enough, we have

$$\|x(s) - s_X\| > 0. \quad (48)$$

Thus, dividing (47) by $\|x(s) - s_X\|^2$ yields

$$\frac{\|x(s) - \tilde{P}_\Omega(x(s))\|^2}{\|x(s) - s_X\|^2} + \frac{\|s_X - \tilde{P}_\Omega(x(s))\|^2}{\|x(s) - s_X\|^2} \leq 1. \quad (49)$$

From the definition of \tilde{P}_Ω , we have, for some $\gamma \in (0, 1)$,

$$\begin{aligned} \text{dist}[\tilde{P}_\Omega(x(s)), \Omega] &\leq \gamma \text{dist}[x(s), \Omega] \\ &\leq \gamma (\|x(s) - \tilde{P}_\Omega(x(s))\| + \text{dist}[\tilde{P}_\Omega(x(s)), \Omega]). \end{aligned}$$

Thus, defining $w(s) := P_\Omega(\tilde{P}_\Omega(x(s)))$, it holds that

$$\text{dist}[\tilde{P}_\Omega(x(s)), \Omega] = \|\tilde{P}_\Omega(x(s)) - w(s)\| \leq \frac{\gamma}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\|. \quad (50)$$

In view of Lemma 3, we may assume that

$$\frac{1}{2} \|x(s) - s_X\| \leq \text{dist}[s, X] \quad (51)$$

is valid for $\delta_2 > 0$ sufficiently small. From (51), (42), and (50) we conclude that, for $\delta_2 > 0$ sufficiently small,

$$\begin{aligned} \frac{1}{4} \|x(s) - s_X\| &\leq \frac{1}{2} \text{dist}[s, X] \\ &\leq \text{dist}[\tilde{P}_\Omega(x(s)), X] \\ &\leq \text{dist}[w(s), X] + \|\tilde{P}_\Omega(x(s)) - w(s)\| \\ &\leq \text{dist}[w(s), X] + \frac{\gamma}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\|. \end{aligned} \quad (52)$$

The next estimate provides an upper bound for $\text{dist}[w(s), O]$. For $\delta_2 > 0$ sufficiently small, the triangle inequality, (17), (50), (42), and (46) yield

$$\begin{aligned} \text{dist}[w(s), O] &\leq \text{dist}[x(s), O] + \|x(s) - \tilde{P}_\Omega(x(s))\| + \|\tilde{P}_\Omega(x(s)) - w(s)\| \\ &\leq L_1 \text{dist}[s, O]^2 + \|x(s) - \tilde{P}_\Omega(x(s))\| + \frac{\gamma}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\| \\ &\leq L_1 \text{dist}[s, X]^2 + \frac{1}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\| \\ &\leq 2L_1 \text{dist}[\tilde{P}_\Omega(x(s)), X]^2 + \frac{1}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\| \\ &\leq 2L_1 \|\tilde{P}_\Omega(x(s)) - s_X\|^2 + \frac{1}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\| \\ &\leq 2L_1 \|x(s) - s_X\|^2 + \frac{1}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\|. \end{aligned} \tag{53}$$

Using (46) and (16), we can easily deduce that $s \in \mathcal{B}(x^*, \delta_2)$ implies $w(s) \in \mathcal{B}(x^*, \delta)$ for $\delta_2 > 0$ small enough. Thus, since $w(s) \in \Omega$ by definition, we get from the error bound (13) and from (14)

$$\omega \text{dist}[w(s), X] \leq \|H(w(s))\| \leq L \text{dist}[w(s), O]. \tag{54}$$

Combining (52), (54), and (53), we obtain

$$\begin{aligned} \frac{1}{4}\omega \|x(s) - s_X\| &\leq L \text{dist}[w(s), O] + \frac{\omega\gamma}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\| \\ &\leq 2LL_1 \|x(s) - s_X\|^2 + \frac{L + \omega\gamma}{1-\gamma} \|x(s) - \tilde{P}_\Omega(x(s))\|. \end{aligned} \tag{55}$$

In view of (48), we can divide (55) by $\|x(s) - s_X\|$, yielding

$$0 < \frac{1}{4}\omega \leq 2LL_1 \|x(s) - s_X\| + \frac{L + \omega\gamma}{1-\gamma} \frac{\|x(s) - \tilde{P}_\Omega(x(s))\|}{\|x(s) - s_X\|}. \tag{56}$$

Choosing $\delta_2 \leq (32LL_1)^{-1}\omega$ and taking into account (51), we get

$$2LL_1 \|x(s) - s_X\| \leq 4LL_1 \text{dist}[s, X] \leq 4LL_1\delta_2 \leq \frac{1}{8}\omega.$$

Hence, by (56), it follows that

$$\frac{\|x(s) - \tilde{P}_\Omega(x(s))\|}{\|x(s) - s_X\|} \geq \sigma := \min \left\{ \frac{1}{2}, \frac{\omega(1-\gamma)}{8(L + \omega\gamma)} \right\} \in (0, 1).$$

Therefore, (49) implies (44) with $\rho := \sqrt{1 - \sigma^2}$ completing the proof of (a).

(b) We complete now the proof of the lemma. It follows from (44) that, for $\delta_2 > 0$ sufficiently small,

$$\text{dist}[\tilde{P}_\Omega(x(s^k)), X] \leq \|\tilde{P}_\Omega(x(s^k)) - s_X^k\| \leq \rho \|x(s^k) - s_X^k\|.$$

With $c := 2\rho/(1 + \rho)$, Lemma 3 implies

$$\text{dist}[\tilde{P}_\Omega(x(s^k)), X] \leq \frac{\rho}{c} \|s^k - s_X^k\| = \frac{1 + \rho}{2} \text{dist}[s^k, X]$$

for $\delta_2 > 0$ sufficiently small. Since $(1 + \rho)/2 < 1$, we have a contradiction to (42). \square

We show next that Lemma 4 need not hold if the error bound on Ω , given by (13), is violated.

Example 3 Let $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ and Ω be defined as in Example 1, where we have already noted that the error bound (12) is satisfied for any solution, while the error bound (13) on Ω is violated for $x^* = (0, 0)^\top$. Now, let us consider the point $s = (-t, 0)^\top \in \mathbb{R}^2$ with $t \in (0, 1]$ arbitrary but fixed. Since s belongs to O , it follows that $x(s) = s$. Since g (see Example 1) is continuously differentiable, its only subgradient is its gradient, i.e., $v(s) = \nabla g(s) = (-2t, -1)^\top$. The approximate projection $\tilde{p}(s)$ on $\mathcal{B}(0, 1)$ associated to Ω is given by (11) and, since $g(s) = t^2 > 0$, we obtain

$$\tilde{p}(s) = \left(\frac{2t^3}{4t^2 + 1} - t, \frac{t^2}{4t^2 + 1} \right)^\top.$$

This and $\text{dist}[s, X] = t$ imply

$$\frac{\text{dist}[\tilde{p}(s), X]^2}{\text{dist}[s, X]^2} = \frac{(\frac{2t^3}{4t^2+1} - t)^2 + (\frac{t^2}{4t^2+1})^2}{t^2} \xrightarrow{t \rightarrow 0} 1.$$

Thus, the assertion of Lemma 4 is not satisfied in any neighborhood of $x^* = (0, 0)^\top$.

Now we are in the position to state and prove the main theorem of this paper.

Theorem 1 *Let Assumptions 1 and 2 be satisfied and suppose that Algorithm 1 uses an approximate projector \tilde{P}_Ω defined on $B \supseteq \mathcal{B}(x^*, 1)$. Then, there exists $\epsilon \in (0, \delta_2]$ such that Algorithm 1 generates a well-defined sequence $\{x^k\}$ for any starting point $x^0 \in \mathcal{B}(x^*, \epsilon)$. This sequence converges R-linearly to a solution $\hat{x} \in X$ of problem (1).*

Proof Let us first choose ϵ according to

$$0 < \epsilon \leq \left(1 + \frac{2L_1 + 1}{1 - \tau} \right)^{-1} \delta_2 \tag{57}$$

with $\tau \in (0, 1)$ from Lemma 4 and $L_1 > 0$ from Lemma 2. We first show by induction that Algorithm 1 generates a well-defined sequence $\{x^k\} \subset \mathcal{B}(x^*, \delta_2)$ if $x^0 \in \mathcal{B}(x^*, \epsilon)$.

To this end we assume that, for some $k \in \mathbb{N}$, x^0, x^1, \dots, x^k were generated by Algorithm 1 and

$$\|x^\nu - x^*\| \leq \delta_2 \quad \text{for } \nu = 0, 1, \dots, k. \tag{58}$$

Using the triangle inequality, we obtain

$$\|x^{k+1} - x^*\| \leq \|x^0 - x^*\| + \sum_{\nu=0}^k \|x^{\nu+1} - x^\nu\|. \tag{59}$$

In order to find an upper bound for the second term in the right hand side of (59) note first that $x^{\nu+1} = \tilde{P}_\Omega(x(x^\nu))$ is well-defined for $\nu = 0, 1, \dots, k - 1$ since x^0, \dots, x^k were successfully generated by Algorithm 1. Moreover, by (58) for $\nu = k$, Lemma 4 with $s := x^k$ implies $x(x^k) \in B$. Thus, $x^{k+1} = \tilde{P}_\Omega(x(x^k))$ is also well-defined and we obtain, for $\nu = 0, \dots, k$,

$$\|x^{\nu+1} - x^\nu\| \leq \|\tilde{P}_\Omega(x(x^\nu)) - x(x^\nu)\| + \|x(x^\nu) - x^\nu\|.$$

Moreover, Property 1 of an approximate projector (see Definition 1) and (16) of Lemma 2 applied to $s := x^\nu$ yield for $\nu = 0, \dots, k$

$$\begin{aligned} \|x^{\nu+1} - x^\nu\| &\leq \gamma \operatorname{dist}[x(x^\nu), \Omega] + \|x(x^\nu) - x^\nu\| \\ &\leq 2\|x(x^\nu) - x^\nu\| + \operatorname{dist}[x^\nu, \Omega] \\ &\leq 2L_1 \operatorname{dist}[x^\nu, O] + \operatorname{dist}[x^\nu, \Omega] \\ &\leq (2L_1 + 1) \operatorname{dist}[x^\nu, X]. \end{aligned} \tag{60}$$

Therefore, (59) leads to

$$\|x^{k+1} - x^*\| \leq \|x^0 - x^*\| + (2L_1 + 1) \sum_{\nu=0}^k \operatorname{dist}[x^\nu, X].$$

In view of (58), Lemma 4 gives

$$\begin{aligned} \|x^{k+1} - x^*\| &\leq \|x^0 - x^*\| + (2L_1 + 1) \operatorname{dist}[x^0, X] \sum_{\nu=0}^k \tau^\nu \\ &\leq \left(1 + \frac{2L_1 + 1}{1 - \tau}\right) \|x^0 - x^*\|. \end{aligned}$$

Since $x^0 \in \mathcal{B}(x^*, \epsilon)$, we have from (57) that x^{k+1} belongs to $\mathcal{B}(x^*, \delta_2)$. Hence, it follows by induction that, for any starting point $x^0 \in \mathcal{B}(x^*, \epsilon)$, Algorithm 1 generates a well-defined infinite sequence $\{x^k\} \subset \mathcal{B}(x^*, \delta_2)$.

Next, it will be shown that the sequence $\{x^k\}$ is a Cauchy-sequence. Take $i, k \in \mathbb{N}$ with $k < i$. Then, as in (60), we obtain

$$\|x^i - x^k\| \leq \sum_{\nu=k}^{i-1} \|x^{\nu+1} - x^\nu\| \leq (2L_1 + 1) \sum_{\nu=k}^{i-1} \operatorname{dist}[x^\nu, X]$$

and Lemma 4 yields

$$\|x^i - x^k\| \leq \tau^k (2L_1 + 1) \operatorname{dist}[x^0, X] \sum_{v=0}^{i-k-1} \tau^v \leq \tau^k \frac{2L_1 + 1}{1 - \tau} \epsilon. \quad (61)$$

Hence, for $k \rightarrow \infty$ the right hand side of (61) tends to 0. Thus, $\{x^k\}$ is a Cauchy-sequence and converges to some $\hat{x} \in \mathcal{B}(x^*, \delta_2)$. From Lemma 4, we have $\operatorname{dist}[x^{k+1}, X] \leq \tau \operatorname{dist}[x^k, X]$ for all $k \in \mathbb{N}$ and, thus, $\lim_{k \rightarrow \infty} \operatorname{dist}[x^k, X] = 0$ holds. Since X is a closed set we conclude that \hat{x} belongs to X . Finally, let us consider (61) and take the limit for $i \rightarrow \infty$. This yields

$$\|\hat{x} - x^k\| \leq \tau^k \frac{2L_1 + 1}{\tau - 1} \quad \text{for all } k \in \mathbb{N}.$$

Hence, if $x^0 \in \mathcal{B}(x^*, \epsilon)$ then $\{x^k\}$ converges R -linearly to $\hat{x} \in X$. \square

4 Final remarks

We have shown that the Levenberg-Marquardt algorithm with approximate projections (LMAP) is locally R -linearly convergent under a pair of reasonable error bound conditions. A special case of this algorithm is the projected Levenberg-Marquardt method [17]. Note that in the existing literature local convergence is known only under the very restrictive Condition 3 which we could avoid. Although the Algorithm 1 has no superlinear convergence in general, the approach presented in this paper has some advantages. One is that it allows computationally cheap approximate projections instead of exact ones. This might be helpful when dealing with more complex convex sets Ω , in particular if Ω is described by nonlinear or even nonsmooth functions. A second advantage lies in the fact that the unconstrained Levenberg-Marquardt subproblems of Algorithm 1 are linear systems of equations, in contrast to constrained Levenberg-Marquardt methods (see [2, 17]), where constrained convex programs are used (even if the latter guarantee quadratic convergence). A more hidden benefit of using unconstrained Levenberg-Marquardt subproblems instead of constrained ones is that one can easily control the inexactness of their solution. This is of importance to ensure a certain convergence rate of the whole method. For example, if the linear system of equations of an unconstrained Levenberg-Marquardt subproblem is solved iteratively, then the residual of the corresponding approximate solution is $\pi(s)$, and so the norm of this residual may be directly used to stop the iterative solver, see [14]. In contrast to this, an appropriate stopping of an iterative solver for constrained Levenberg-Marquardt subproblems seems more difficult.

In the next subsection we show that, in general, the results on the level of inexactness and on the convergence rate cannot be improved. Finally, the last subsection provides numerical results for a test example.

4.1 Sharpness of results

The relation of the regularization parameter $\alpha(s)$ and the inexactness $\|\pi(s)\|$ within Levenberg-Marquardt subproblems and their influence on the convergence rate were

investigated in [4, 9, 14] and in [2], where the most general results can be found. In particular, if the regularization parameter $\alpha(s)$ in subproblem (4) is $\|H(s)\|$, then a level of inexactness of order $\mathcal{O}(\|H(s)\|^2)$, i.e., $\|\pi(s)\| \sim \|H(s)\|^2$, still maintains the local quadratic convergence of unconstrained or constrained exact Levenberg-Marquardt methods.

In Algorithm 1 we allow only a slightly lower level of inexactness, namely $\|\pi(s)\| \sim \|H(s)\|^{2+\eta}$ with some $\eta > 0$. Interestingly, the following example shows that, for $\eta = 0$, the algorithm may fail to converge at all.

Example 4 Let $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $\Omega \subset \mathbb{R}^2$ be given by

$$H(x_1, x_2) := x_2 \quad \text{and} \quad \Omega := \{x \in \mathbb{R}^2 \mid x_1 + x_2 \geq 0\}.$$

Assumptions 1 and 2 are satisfied at any solution of problem (1). In particular, the error bounds given by (12) and (13) hold since H is affine and Ω a polyhedron. For some $t \in (0, 1]$ let us define

$$s := (-t, t)^\top \in \Omega \quad \text{and} \quad \pi(s) := \left(\frac{-t^2}{t+1}, 0 \right)^\top.$$

Obviously, we have $\|\pi(s)\| \leq t^2 = \|H(s)\|^2$. For the solution of the Levenberg-Marquardt subproblem (4) we get

$$x(s) = \left(-t - \frac{t}{t+1}, t - \frac{t}{t+1} \right)^\top.$$

It is easy to verify that $P_\Omega(x(s)) = (-t, t)^\top = s$, where P_Ω is the orthogonal projector on Ω . Thus, the Algorithm 1 is trapped at $s = (-t, t)^\top$. No matter how close a starting point is to the solution $x^* := (0, 0)^\top$, allowing an inexactness of order $\mathcal{O}(\|H(s)\|^2)$, may make Algorithm 1 generate a nonconvergent sequence.

In general, under Assumptions 1 and 2, the convergence rate of Algorithm 1 we can expect even for exact projections is not better than linear. In order to see this, let us again consider the problem in Example 4. For $t \in (0, 1]$ define $s := (-t, t)^\top$ and $\pi(s) := 0$. Then, we obtain

$$x(s) = \left(-t, t - \frac{t}{1+t} \right)^\top \quad \text{and} \quad P_\Omega(x(s)) = \left(-t + \frac{t}{2+2t}, t - \frac{t}{2+2t} \right)^\top.$$

Then,

$$\text{dist}[s, X] = \sqrt{2}t \quad \text{and} \quad \text{dist}[P_\Omega(x(s)), X] = \sqrt{2}t \left(1 - \frac{1}{2+2t} \right)$$

follow. Hence, in general, Algorithm 1 does not generate a superlinearly convergent sequence.

4.2 A test example

Although the focus of this paper is theoretical, we would like to illustrate how the Levenberg-Marquardt algorithm with approximate projections works to provide a feeling of its behavior.

Example 5 Let $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $\Omega \subset \mathbb{R}^2$ be defined as

$$H(x) := x_2 - x_1$$

and

$$\Omega := \{x \in \mathbb{R}^2 \mid g_i(x) \leq 0, i = 1, \dots, 4\},$$

with functions $g_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$g_1(x) := -x_1,$$

$$g_2(x) := -x_2,$$

$$g_3(x) := x_1^2 + x_2^2 - 1$$

and

$$g_4(x) := \begin{cases} \frac{2}{2x_1+1} - x_2 - \frac{3}{2} & \text{if } x_1 \geq 0, \\ -x_2 + \frac{1}{2} & \text{if } x_1 < 0. \end{cases}$$

One can easily check that Ω is a convex set. Moreover, Ω can be rewritten as

$$\Omega = \{x \in \mathbb{R}^2 \mid g(x) \leq 0\}$$

with

$$g(x) := \max\{g_1(x), g_2(x), g_3(x), g_4(x)\}.$$

With H and Ω defined above, the solution set of problem (1) is

$$X = \left\{ (t, t)^\top \in \mathbb{R}^2 \mid \frac{\sqrt{5}-2}{2} \leq t \leq \frac{1}{2}\sqrt{2} \right\}.$$

Furthermore, it can be shown that Assumptions 1 and 2 are satisfied at any solution. Algorithm 1 is run with the approximate projector given by (11) and starting point $x^0 = (-1, 1)^\top$. The results are shown in Table 1.

If the constrained Levenberg-Marquardt method is used to find a solution of this problem, the subproblems are optimization problems with a convex quadratic objective and the inequality constraints $g_i(x) \leq 0$ for $i = 1, \dots, 4$. Thus, solving the subproblems might be computationally a harder task. One option for avoiding this situation consists of introducing slack variables y_1, \dots, y_4 for the inequalities. More

Table 1 Numerical results for Example 5

k	$\text{dist}[x^k, X]$	$\text{dist}[x^k, X]/\text{dist}[x^{k-1}, X]$
0	1.4240e+00	–
1	7.2654e−01	0.51020
2	1.5282e−01	0.21034
3	1.2117e−02	0.07929
4	3.4422e−03	0.28408
5	6.0794e−04	0.17661
6	1.0242e−04	0.16847
7	1.7102e−05	0.16698

precisely, problem (1) could be reformulated as follows

$$\tilde{H}(x, y) := \begin{pmatrix} x_2 - x_1 \\ g_1(x) + y_1 \\ \dots \\ g_4(x) + y_4 \end{pmatrix} = 0 \quad \text{s.t.} \quad (x, y) \in \tilde{\mathcal{D}} := \mathbb{R}^2 \times \mathbb{R}_+^4.$$

If, regardless of the nondifferentiability of g_4 , the constrained Levenberg-Marquardt method is applied to this reformulation, the subproblems become quadratic optimization problems with box constraints. However, the number of variables increases and, with it, the cost of solving a subproblem.

Experiments with large scale examples are going to be part of our future research. In this direction, one can consider a differentiable reformulation of Karush-Kuhn-Tucker systems which arise from generalized Nash equilibrium problems as in [5]. In this context, that paper provides a sufficient condition for the pair of error bounds (12) and (13) to hold. Moreover, a hybrid algorithm is described in [5] whose local phase consists of a method with quadratic convergence behavior. Replacing the latter method by Algorithm 1 would lead to local R-linear convergence only, but to significantly less expensive subproblems.

Acknowledgements The first author would like to thank *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) and *Católica SC*, from Brazil, for the financial support. We also thank the anonymous referees for valuable comments.

References

- Behling, R.: The Method and the Trajectory of Levenberg-Marquardt. PhD thesis, IMPA, Rio de Janeiro, Brazil (2011)
- Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. *Optim. Lett.* **6**, 927–940 (2012)
- Behling, R., Iusem, A.: The effect of calmness on the solution set of nonlinear equations. *Math. Program.* **137**, 155–165 (2013)
- 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)

5. Dreves, A., Facchinei, F., Fischer, A., Herrich, M.: A new error bound result for generalized Nash equilibrium problems and its algorithmic application. Technical report MATH-NM-1-2013, Institute of Numerical Mathematics, Technische Universität Dresden, Dresden, Germany, January (2013). Available online at: http://www.optimization-online.org/DB_HTML/2013/02/3765.html
6. Facchinei, F., Fischer, A., Herrich, M.: A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. *Math. Methods Oper. Res.* (to appear)
7. Facchinei, F., Fischer, A., Piccialli, V.: Generalized Nash equilibrium problems and Newton methods. *Math. Program.* **117**, 163–194 (2009)
8. Fan, J.: On the Levenberg-Marquardt methods for convex constrained nonlinear equations. *J. Ind. Manag. Optim.* **9**, 227–241 (2013)
9. Fan, J., Pan, J.: Inexact Levenberg-Marquardt method for nonlinear equations. *Discrete Contin. Dyn. Syst., Ser. B* **4**, 1223–1232 (2004)
10. Fan, J., Pan, J.: On the convergence rate of the inexact Levenberg-Marquardt method. *J. Ind. Manag. Optim.* **7**, 199–210 (2011)
11. Fan, J., Yuan, Y.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing* **74**, 23–39 (2005)
12. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program.* **94**, 91–124 (2002)
13. Fischer, A., Shukla, P.K.: A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. *Oper. Res. Lett.* **36**, 643–646 (2008)
14. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg-Marquardt methods. *Optimization* **59**, 273–287 (2010)
15. Fukushima, M.: An outer approximation algorithm for solving general convex programs. *Oper. Res.* **31**, 101–113 (1983)
16. Hoffman, A.J.: On approximate solutions of systems of linear inequalities. *J. Res. Natl. Bur. Stand.* **49**, 263–265 (1952)
17. 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)
18. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **2**, 164–168 (1944)
19. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11**, 431–441 (1963)
20. Rockafellar, R.T., Wets, R.J.-B.: *Variational Analysis*. Springer, Berlin (1998)
21. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. *Computing* **15**, 239–249 (2001)
22. Zhang, J.-L.: On the convergence properties of the Levenberg-Marquardt method. *Optimization* **52**, 739–756 (2003)