Computational Optimization and Applications (2024) 89:33–67
https://doi.org/10.1007/s10589-024-00589-1

The Levenberg–Marquardt method: an overview of modern convergence theories and more

Andreas Fischer[1] · **Alexey F. Izmailov**[2,3] · **Mikhail V. Solodov**[4]

Received: 27 November 2023 / Accepted: 31 May 2024 / Published online: 11 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

The Levenberg–Marquardt method is a fundamental regularization technique for the Newton method applied to nonlinear equations, possibly constrained, and possibly with singular or even nonisolated solutions. We review the literature on the subject, in particular relating to each other various convergence frameworks and results. In this process, the analysis is performed from a unified perspective, and some new results are obtained as well. We discuss smooth and piecewise smooth equations, inexact solution of subproblems, and globalization techniques. Attention is also paid to the LP-Newton method, because of its relations to the Levenberg–Marquardt method.

Keywords Nonlinear equation · Constrained equation · Piecewise smooth equation · Nonisolated solution · Local error bound · Gauss–Newton method · Levenberg–Marquardt method · LP-Newton method · Singular solution

Mathematics Subject Classification 47J05 · 65K15

✉ Alexey F. Izmailov
izmaf@cs.msu.ru

Andreas Fischer
Andreas.Fischer@tu-dresden.de

Mikhail V. Solodov
solodov@impa.br

1 Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany

2 OR Department, VMK Faculty, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory, Moscow, Russia 119991

3 Derzhavin Tambov State University (TSU), Internationalnaya 33, Tambov, Russia 392000

4 IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil

1 Introduction and description of the methods

We consider the constrained equation

$$\Phi(u) = 0, \quad u \in P, \tag{1.1}$$

where $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ is a given mapping, and $P \subset \mathbb{R}^p$ is a given nonempty closed set. Let U stand for the solution set of (1.1).

The case of $P = \mathbb{R}^p$ and $p = q$ corresponds to the classical setting of unconstrained nonlinear equations. The fundamental paradigm for solving such problems is the Newton method; see, e.g., [45]. In its basic form, the Newton method requires for convergence that a given solution is nonsingular (in particular, isolated). The Levenberg–Marquardt (LM) method for unconstrained nonlinear equations [52, 56] (see also [59, Section 10.3]) is a classical regularization technique for handling cases when a solution can be singular (and possibly nonisolated, and possibly $p \neq q$).

Before introducing the LM method, it is natural to consider first the (constrained) Gauss–Newton (GN) method. For the current iterate $u \in P$, the GN method defines the next iterate as $u + v$, where v minimizes the (squared) residual of the linearized equation from (1.1) over $P - u$, i.e., v is a solution of the optimization problem

$$\text{minimize } \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 \ \text{ subject to } \ u + v \in P. \tag{1.2}$$

Due to the Frank–Wolfe Theorem [39], this subproblem always has a solution when P is polyhedral, but a solution need not be unique.

Partially because of the potential lack of uniqueness of solutions in the subproblem (1.2), it is natural to regularize it. This yields the *constrained* LM method [8, 50] for (1.1), defining the next iterate as $u + v$, where v solves the optimization problem

$$\text{minimize } \frac{1}{2}\|\Phi(u) + G(u)v\|^2 + \frac{1}{2}\sigma(u)\|v\|^2 \text{ subject to } u + v \in P, \tag{1.3}$$

with a function $\sigma : P \to \mathbb{R}_+$ defining the values of the regularization parameter, and with $G(u)$ being $\Phi'(u)$, or its suitable substitute when it does not exist or is not available. If $\sigma(u) > 0$, and if Euclidean norms are used, the objective function of (1.3) is strongly convex quadratic, and hence, this subproblem has a solution, which is necessarily unique if P is convex. Moreover, if P is a polyhedral set, then (1.3) is a strongly convex quadratic programming problem.

For convergence properties of the classical (unconstrained) LM method in the case of smooth Φ and isolated solutions, as well as related references, see [19, Theorem 10.2.6]. The cases of nonsmooth (semismooth) Φ with isolated solutions is considered in [25]; see also [48, 49].

The purpose of this work is to discuss, and in some aspects to clarify, the more recent theories developed under weaker assumptions allowing, in particular, for nonisolated solutions. The key conditions are of the local Lipschitzian error bound type, originating in the context of LM methods from [63]. Another attention is on the combination of

nonisolated solutions with nonsmoothness, which is an especially challenging situation [22]. We emphasize that our focus is on the asymptotic convergence properties. For some complexity results on LM methods, we refer the reader to [13, 57]. Yet some other issues that we do not consider in this paper are the following: more general regularization terms (allowing other norms and their exponents) in (1.3), as in [3]; the case when (1.1) has no solutions and so the LM method is intended to minimize the residual to some nonzero optimal value [12, 61]; the case of Hölderian (rather than Lipschitzian) error bounds [1] and Hölderian continuity of the derivative, and the corresponding choices of the regularization and inexactness parameters in solving the subproblems.

As applications of LM methods to specific instances of (constrained) equations, we mention the following: [25, 48–50] for complementarity problems; [34] for complementarity systems; [34, 46] for Karush–Kuhn–Tucker systems; [24, 34] for generalized Nash equilibrium problems; [36] for multicriteria optimization; [14, 61] for (constrained) least squares; [47, 62] for bilevel optimization.

Before proceeding, we define some notation. The Euclidean inner product for $u,\ v \in \mathbb{R}^p$ is denoted by $\langle u,\ v\rangle$. Generally, $\|\cdot\|$ stands for some (arbitrary) norm. When necessary, we specify that it is Euclidean. We denote $\|u\|_\infty = \max_{i=1,\ldots,p} |u_i|$. For a set $U \subset \mathbb{R}^p$ and a point $u \in \mathbb{R}^p$, $\mathrm{dist}(x,\ U) = \inf_{v\in U} \|v-u\|$, and $B(u,\ \delta) = \{v \in \mathbb{R}^p \mid \|v-u\| \le \delta\}$ is the closed ball of radius $\delta \ge 0$ centered at u. By $\mathcal{I}$ we denote the identity matrix, with dimension always clear from the context.

For a closed convex set $U \subset \mathbb{R}^p$, by $\pi_U(x)$ we mean the unique metric projection of $u \in \mathbb{R}^p$ onto U, i.e., the solution of $\min_{v\in U} \|v-u\|$ where the norm is Euclidean. By $N_U(u)$ we denote the normal cone to U at u, i.e., $N_U(u) = \{v \in \mathbb{R}^p \mid \langle v,\ \widetilde{u}-u\rangle \le 0,\ \forall \widetilde{u} \in U\}$ if $u \in U$, and $N_U(u) = \emptyset$ otherwise.

A function $F : \mathbb{R}^p \to \mathbb{R}^q$ is Hölder-continuous of order $\beta \in (0,\ 1]$ on a set $U \subset \mathbb{R}^p$ if $\|F(u)-F(v)\| \le L\|u-v\|^\beta$ for some $L > 0$ and all $u,\ v \in U$. It is Lipschitz-continuous if $\beta = 1$.

For a sequence $\{u^k\} \subset \mathbb{R}^p$ convergent to some $\bar{u} \in \mathbb{R}^p$, we say that convergence is of Q-order $\theta > 1$ if there exists $c > 0$ such that $\|u^{k+1}-\bar{u}\| \le c\|u^k-\bar{u}\|^\theta$ for all k large enough. Such rate of convergence is superlinear, and it is at least quadratic if $\theta \ge 2$. We say that $\{u^k\}$ converges to $\bar{u}$ with R-order θ if there exist $c > 0$ and a sequence $\{t_k\} \subset \mathbb{R}_+$ converging to 0 with Q-order θ, such that $\|u^{k+1}-\bar{u}\| \le ct_k$ for all k large enough.

One alternative approach to adapting the idea of the classical LM method to constrained equations with convex constraint sets was apparently first considered in [50]. This approach can make sense when computing the metric projection onto P is easier than solving (1.3). Then one can define the next iterate as $\pi_P(u+v)$, where v is obtained by solving the unconstrained optimization problem obtained by removing the constraint in (1.3). The iterative process constructed this way is called *projected* LM method.

In what follows, we also pay some attention to an algorithm which is somehow related to the LM method (1.3), but is different. It was introduced in [23] under the name LP-Newton (LPN) method. Given the current iterate u, its subproblem for defining the next iterate $u+v$ solves for $(v,\ \gamma)$

$$\begin{aligned} &\text{minimize } \gamma \text{ subject to } \|\Phi(u) + G(u)v\| \le \gamma \|\Phi(u)\|^2, \\ &\qquad\qquad \|v\| \le \gamma \|\Phi(u)\|, \\ &\qquad\qquad u + v \in P, \ \gamma \ge 0. \end{aligned} \tag{1.4}$$

The nonnegativity constraint on γ is redundant when $u \notin U$, and thus can be dropped in practical computations (recall that U is the solution set of (1.1)). But in theoretical considerations it is convenient to have this constraint, because it ensures solvability of the subproblem (1.4) even when $u \in U$. Observe that if P is a polyhedral set, and if the ∞-norm is used in the left-hand sides of inequality constraints in (1.4), the latter turns into a linear programming problem, thus giving the name to the method. On the other hand, if the Euclidean norms are used, and if P is just convex, the LPN method turns out to be closely related to the LM method, being an instance of the latter with a special choice of the regularization parameter. Indeed, the first-order necessary optimality condition for the LM subproblem (1.3) has the form

$$(G(u))^\top (\Phi(u) + G(u)v) + \sigma(u)v + N_P(u + v) \ni 0 \tag{1.5}$$

(see, e.g., [15, Proposition 2.1.2], [59, Theorem 12.9]). Similarly, assuming that $\Phi(u) \ne 0$, the first-order necessary optimality conditions for the LPN subproblem with squared left-hand and right-hand sides of inequality constraints consist of the relations

$$2\mu_1 (G(u))^\top (\Phi(u) + G(u)v) + 2\mu_2 v + N_P(u + v) \ni 0, \tag{1.6}$$

$$1 - 2\mu_1 \gamma \|\Phi(u)\|^4 - 2\mu_2 \gamma \|\Phi(u)\|^2 = 0, \tag{1.7}$$

complemented by the inequality constraints in (1.4), by nonnegativity requirements on dual variables μ_1 and μ_2, and by the complementarity slackness conditions. Equation (1.7) implies that μ_1 and μ_2 cannot be both zero, and in particular, at least one of the inequality constraints in (1.4) must be active at its solution. If $\mu_1 > 0$, then (1.6) coincides with (1.5) with $\sigma(u) = \mu_2/\mu_1$. If $\mu_1 = 0$, then (1.6) implies that $-v \in N_P(u + v)$, and since $u \in P$, the latter further implies that $0 \ge \langle -v, \ u - (u + v)\rangle = \|v\|^2$. Therefore, in this case, it necessarily holds that $v = 0$, and since the second inequality constraint in (1.4) must be active, it also holds that $\gamma = 0$ (recall that $\Phi(u) \ne 0$). But then the first inequality constraint in (1.4) cannot hold, unless $\Phi(u) = 0$.

A deeper relation between the LM and LPN methods will be demonstrated below by the fact that, under appropriate assumptions, both methods fit the local convergence frameworks to be described in Sect. 2.

The rest of this paper is organized as follows. Section 2 reviews convergence frameworks and results relevant for the context: [22, 23, 34, 38], and relations between them. Section 3 focuses on the smooth case, where in particular some clarifications and new results are obtained. In Sect. 4 we discuss the piecewise smooth case, and in Sect. 5 approximate solution of subproblems. Section 6 is devoted to some globalization techniques of the local algorithms in question. We conclude with Sect. 7, where some open questions are summarized.

2 Local convergence frameworks

We start with a discussion of the abstract local convergence framework recently proposed in [38]. For this, it is convenient to consider a problem with a single scalar equation:

$$\varphi(u) = 0, \quad u \in P, \tag{2.1}$$

where $\varphi : \mathbb{R}^p \to \mathbb{R}_+$ is a given scalar-valued function, and $P \subset \mathbb{R}^p$ is a given nonempty closed set. Evidently, the constrained equation (1.1) can be stated in the form (2.1) with, say,

$$\varphi(u) = \|\Phi(u)\|^\nu, \tag{2.2}$$

with any $\nu > 0$. To that end, let for now U stand for the solution set of (2.1).

Let $\Psi : P \to P$ be a given mapping, and consider an abstract iterative process updating the current iterate $u \in P$ to the new one of the form $\Psi(u)$. The following is [38, Theorem 2.1].

Theorem 2.1 *Let $\varphi : \mathbb{R}^p \to \mathbb{R}_+$ be a continuous function, $P \subset \mathbb{R}^p$ be a nonempty closed set, and $\bar{u} \in U$. Moreover, let $\Psi : P \to P$ be a mapping such that, with some $\tau > 1$,*

$$\|\Psi(u) - u\| = O(\varphi(u)) \tag{2.3}$$

and

$$\varphi(\Psi(u)) = O((\varphi(u))^\tau) \tag{2.4}$$

as $u \in P$ tends to $\bar{u}$.

Then, for every $\delta > 0$, and every $u^0 \in P$ close enough to $\bar{u}$, the sequence $\{u^k\}$ defined by $u^{k+1} = \Psi(u^k)$ for all k is contained in $B(\bar{u}, \delta)$ and converges to some $u^* \in U$ with the R-order τ.

If, in addition, there exists $\beta \in (0, 1]$ such that

$$\varphi(u) = O((\mathrm{dist}(u, U))^\beta) \tag{2.5}$$

as $u \in P$ tends to $\bar{u}$, then, for each integer s, there exists $c_s > 0$ such that

$$\|u^{k+s} - u^*\| \le c_s \|u^k - u^*\|^{\beta\tau^s}$$

holds for all k large enough. In particular, if $\beta\tau^s > 1$, the rate of convergence of the sequence $\{u^k\}$ is s-step superlinear with the Q-order $\beta\tau^s$.

We remark that [38, Theorem 2.1] employs the Hölder-continuity assumption for φ in some neighborhood of $\bar{u}$. This can be easily relaxed by the weaker requirement (2.5) above.

We now get back to the problem setting (1.1), and recall the local convergence framework developed in [22] on the basis of the constructions for the LPN method in [23], and with some modifications discussed in [34]. Define the set-valued mapping $\mathcal{F}$ from $\mathbb{R}^p \times \mathbb{R}_+$ to the subsets of $\mathbb{R}^p$ by

$$\mathcal{F}(u,\ \gamma) = \{v \in P - u \mid \|\Phi(u) + G(u)v\| \le \gamma\|\Phi(u)\|^2,\ \|v\| \le \gamma\|\Phi(u)\|\}. \tag{2.6}$$

Observe that the constraints defining $\mathcal{F}(u,\ \gamma)$ exactly correspond to the two inequality constraints of the LPN subproblem (1.4). For the current iterate u, the iterative framework in question generates the next iterate as $u + v$, with any

$$v \in \mathcal{F}(u,\ \Gamma),$$

where $\Gamma > 0$ is a pre-fixed parameter.

To ensure local quadratic convergence of this iterative scheme, we make use of the following assumptions.

Assumption 1 Lipschitzian growth of the equation residual:

$$\|\Phi(u)\| = O(\mathrm{dist}(u,\ U))$$

as $u \in P$ tends to $\bar{u}$.

Assumption 2 Existence of approximate solutions of subproblems: There exists $\Gamma > 0$ such that

$$\mathcal{F}(u,\ \Gamma) \neq \emptyset$$

for all $u \in P$ close enough to $\bar{u}$, where $\mathcal{F}(u,\ \Gamma)$ is defined according to (2.6).

Assumption 3 Quality of approximation: For any $v \in P - u$ satisfying

$$\|\Phi(u) + G(u)v\| \le t^2, \quad \|v\| \le t, \tag{2.7}$$

it holds that

$$\|\Phi(u + v)\| = O(t^2)$$

as $u \in P \setminus U$ tends to $\bar{u}$ and $t \to 0+$.

The following theorem essentially recovers [22, Theorem 1], but here it is derived as a particular case of Theorem 2.1 above, under Assumptions 1–3. Observe that there is an additional error bound condition in [22], as well as in [23, Theorem 1] and [34, Theorem 1]. We point out that this additional condition resulted from the way of reasoning adopted to prove [23, Theorem 1]. Here, it is avoided by a subtler argument used in [38] to prove Theorem 2.1 above. In particular, the proof of Theorem 2.1 in [38]

does not involve distances to the solution set at all. Note, however, that appropriate error bound conditions still play an important role in the convergence theory for ensuring that Assumption 2 holds for specific mappings G; see the subsequent Sects. 3 and 4.

Theorem 2.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a continuous mapping, $P \subset \mathbb{R}^p$ a nonempty closed set, and $\bar{u} \in U$. For a fixed mapping $G : \mathbb{R}^p \to \mathbb{R}^{q\times p}$, let Assumptions 1–3 be satisfied.*

Then, for every $\delta > 0$, every $\Gamma > 0$ large enough, and every $u^0 \in P$ close enough to $\bar{u}$, there exists a sequence $\{u^k\}$ such that $u^{k+1} - u^k \in \mathcal{F}(u^k, \Gamma)$ for all k, with $\mathcal{F}(u^k, \Gamma)$ defined according to (2.6); any such sequence is contained in $B(\bar{u}, \delta)$, converges to some $u^* \in U$, and the rate of convergence is quadratic.

Assumption 1 is readily satisfied if Φ is Lipschitz-continuous near $\bar{u}$. Moreover, if Assumption 1 holds, then (2.5) holds with $\varphi(u) = \|\Phi(u)\|$ and $\beta = 1$ as $u \in P$ tends to $\bar{u}$. Observe again that with this choice of φ, problems (1.1) and (2.1) have the same solution set U.

The fulfilment of Assumptions 2 and3 depends on the choice of G. Assumption 2 just means that for $\Gamma > 0$ large enough, and for a current iterate $u \in P$ close enough to $\bar{u}$, the iteration process considered in Theorem 2.2 successfully generates the next iterate. Therefore, under this assumption, one can always consider that this iterative process can be characterized by some mapping $\Psi : P \to P$ in Theorem 2.1. Moreover, (2.6) directly implies condition (2.3) as $u \in P$ tends to $\bar{u}$.

Finally, if Assumption 3 holds, applying it with $t = \max\{\sqrt{\Gamma}, \Gamma\}\|\Phi(u)\|$ yields (2.4) as $u \in P$ tends to $\bar{u}$, for Ψ specified above, and with $\tau = 2$. This completes the observations demonstrating that Theorem 2.2 follows from Theorem 2.1.

To conclude this section, in addition to the iterative frameworks dealt with in Theorems 2.1 and 2.2 above, we would like to mention two other convergence frameworks with some relations to the issues considered here. One is given in [37], and it will be used in the sequel as Proposition 3.2. It helps to easily show a Q-order of convergence of an abstract sequence. The second framework we would like to mention is that of [32], which allows to investigate convergence of a large class of Newton-type methods for generalized equations.

3 Local convergence in the smooth case

We proceed with local convergence analysis of the constrained LM algorithm under the assumption that Φ is differentiable in some neighborhood of a solution $\bar{u}$ of (1.1), and with Φ' being Lipschitz-continuous in this neighborhood. Accordingly, throughout this section, we set $G(u) = \Phi'(u)$ for $u \in \mathbb{R}^p$ close enough to $\bar{u}$. In the first part of this section, we shall verify Assumptions 1–3, in order for Theorem 2.2 to be applicable. Thereafter, further topics will be discussed, including the influence of the regularization parameter.

In the setting of this section, Assumption 1 is automatically satisfied since Φ is Lipschitz-continuous near $\bar{u}$. To see that Assumption 3 is fulfilled, let us take any

$v \in \mathbb{R}^p$ satisfying (2.7). Then the Mean-Value Theorem [45, Theorem A.10] yields

$$\begin{aligned}\|\Phi(u+v)\| &\le \|\Phi(u)+\Phi'(u)v\| + \|\Phi(u+v)-\Phi(u)-\Phi'(u)v\| \\ &\le t^2 + \sup_{\tau\in[0,\,1]} \|\Phi'(u+\tau v)-\Phi'(u)\|\|v\| \\ &\le t^2 + O(\|v\|^2) \\ &= O(t^2)\end{aligned}$$

as $u \to \bar{u}$ and $t \to 0+$. Thus, Assumption 3 is satisfied.

It remains to verify Assumption 2, and this should be done by demonstrating that the LM step belongs to $\mathcal{F}(u,\,\Gamma)$ with fixed $\Gamma > 0$ large enough, for $u \in P$ close enough to $\bar{u}$. The latter, together with the already verified Assumptions 1 and 3, will also mean that the constrained LM method fits the iterative framework of Theorem 2.2.

If $u \in U$, then according to (2.6), it holds that $0 \in \mathcal{F}(u,\,\Gamma)$ for any choice of Γ, and moreover, under the reasonable convention that the constrained LM method always picks up specifically $v = 0$ at any solution $u \in U$, we formally conclude that in this case the LM step belongs to $\mathcal{F}(u,\,\Gamma)$. Now we need to demonstrate the same for $u \in P \setminus U$.

Assume that the regularization parameter satisfies the requirements

$$\|\Phi(u)\|^2 = O(\sigma(u)), \quad \sigma(u) = O(\|\Phi(u)\|^2) \tag{3.1}$$

as $u \in P$ tends to $\bar{u}$, where the first estimate implies, in particular, that $\sigma(u) > 0$ for all $u \in P \setminus U$ close enough to $\bar{u}$. We also need to assume that

$$\operatorname{dist}(u,\,U) = O(\|\Phi(u)\|) \tag{3.2}$$

holds as $u \in P$ tends to $\bar{u}$. The latter property is known as the constrained error bound. Then, from the first estimate in (3.1), it also follows that

$$\frac{(\operatorname{dist}(u,\,U))^4}{\sigma(u)} = (\operatorname{dist}(u,\,U))^2 O\left(\frac{\|\Phi(u)\|^2}{\sigma(u)}\right) = O((\operatorname{dist}(u,\,U))^2) \tag{3.3}$$

as $u \in P \setminus U$ tends to $\bar{u}$.

Since $\sigma(u) > 0$ for $u \in P \setminus U$ near $\bar{u}$, the LM subproblem (1.3) necessarily has a solution v. We shall show that there exists $\Gamma > 0$ such that $v \in \mathcal{F}(u,\,\Gamma)$ for all $u \in P \setminus U$ close enough to $\bar{u}$. The argument essentially follows the one in [50, Lemma 2.3].

Let $\widehat{u}$ stand for a metric projection of u onto the solution set U:

$$\|u - \widehat{u}\| = \operatorname{dist}(u,\,U). \tag{3.4}$$

Since v is a global solution of (1.3), it holds that

$$\|\Phi(u)+\Phi'(u)v\|^2+\sigma(u)\|v\|^2 \le \|\Phi(u)+\Phi'(u)(\widehat{u}-u)\|^2+\sigma(u)\|\widehat{u}-u\|^2. \tag{3.5}$$

Evidently, $\widehat{u} \to \bar{u}$ as $u \to \bar{u}$, and from (3.4), (3.5), applying again the Mean-Value Theorem [45, Theorem A.10], we derive that

$$\begin{aligned}
\|v\|^2 &\leq \frac{1}{\sigma(u)} \left(\|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 + \sigma(u)\|\widehat{u} - u\|^2 \right) \\
&= \frac{1}{\sigma(u)} \|\Phi(u) - \Phi(\widehat{u}) - \Phi'(u)(u - \widehat{u})\|^2 + \|u - \widehat{u}\|^2 \\
&\leq \frac{1}{\sigma(u)} \sup_{\tau \in [0,\,1]} \|\Phi'(\tau u + (1-\tau)\widehat{u}) - \Phi'(u)\|^2 \|u - \widehat{u}\|^2 + \|u - \widehat{u}\|^2 \\
&= \|u - \widehat{u}\|^2 + O\left(\frac{\|u - \widehat{u}\|^4}{\sigma(u)} \right) \\
&= O((\mathrm{dist}(u,\, U))^2), \qquad (3.6)
\end{aligned}$$

where the last estimate is by (3.3). Hence,

$$\|v\| = O(\mathrm{dist}(u,\, U)) \qquad (3.7)$$

as $u \in P \setminus U$ tends to $\bar{u}$.

Furthermore, by a similar reasoning as in (3.6), from (3.4), (3.5) we also have that

$$\begin{aligned}
\|\Phi(u) + \Phi'(u)v\|^2 &\leq \|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 + \sigma(u)\|\widehat{u} - u\|^2 \\
&= \sigma(u)\|u - \widehat{u}\|^2 + O(\|u - \widehat{u}\|^4) \\
&= O(\|\Phi(u)\|^2 (\mathrm{dist}(u,\, U))^2) + O((\mathrm{dist}(u,\, U))^4),
\end{aligned}$$

where the last relation is by the second estimate in (3.1), and hence, taking into account that Assumption 1 is satisfied, we conclude that

$$\|\Phi(u) + \Phi'(u)v\| = O((\mathrm{dist}(u,\, U))^2) \qquad (3.8)$$

as $u \in P \setminus U$ tends to $\bar{u}$.

The needed inclusion $v \in \mathcal{F}(u,\, \Gamma)$ now follows by combining (3.7) and (3.8) with (3.2) as $u \in P$ tends to $\bar{u}$.

The observations above together with Theorem 2.2 yield the following result, essentially recovering [50, Theorem 2.11].

Theorem 3.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ be a nonempty closed set, and $\bar{u} \in U$. Assume that Φ is differentiable on P, with Φ' being Lipschitz-continuous near $\bar{u}$. Let the error bound (3.2) hold, and let the regularization function $\sigma : P \to \mathbb{R}_+$ satisfy (3.1) as $u \in P$ tends to $\bar{u}$.*

Then, for every $u^0 \in P$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is a solution of (1.3) with $u = u^k$ and $G(u) = \Phi'(u^k)$ (with the convention that $u^{k+1} = u^k$ if $u^k \in U$). For every $\delta > 0$, if $u^0 \in P$ is close enough to $\bar{u}$, then every such sequence is contained in $B(\bar{u},\, \delta)$ and converges to some $u^* \in U$, and the rate of convergence is quadratic.

The smoothness requirements in Theorem 3.1 can be relaxed to piecewise smoothness at least, and this will be done in Sect. 4. Observe that the analysis above justifying Theorem 3.1 relies directly and solely on the property of the step to be a global solution of the LM subproblem (1.3). However, when Φ is smooth and P is convex, some additional powerful tools were developed in [8], leading to sharper results. In particular, [8] allows for a wider range of choices for the regularization parameter than those in (3.1). These tools consider the LM step also as a stationary point of (1.3). To that end, from now on and up to the end of this section, let the norms used in the LM subproblem (1.3) be Euclidean. However, before proceeding with stationarity conditions, the following observation is in order.

Assume now that

$$\|\Phi(u)\|^\theta = O(\sigma(u)), \quad \sigma(u) = O(\|\Phi(u)\|^\theta) \tag{3.9}$$

as $u \in P$ tends to $\bar{u}$, with some fixed $\theta \in (0, 2]$. From the first estimate and from (3.2), as $u \in P$ tends to $\bar{u}$, it then follows that (3.3) is still valid, further implying (3.7) as $u \in P \setminus U$ tends to $\bar{u}$, where $v = v(u)$ is now the unique solution of (1.3).

Any $u \in U$ is a (global) solution of the optimization problem

$$\text{minimize } \frac{1}{2}\|\Phi(u)\|^2 \text{ subject to } u \in P, \tag{3.10}$$

and the objective function of this problem is differentiable at u provided Φ is, with the gradient being $(\Phi'(u))^\top \Phi(u)$. Therefore, any such u must satisfy the first-order necessary optimality condition

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni 0. \tag{3.11}$$

The key role in this analysis is played by the following result, demonstrating that the constrained error bound implies the upper-Lipschitzian property of the solution set of the generalized equation (3.11) subject to the right-hand side perturbations.

Proposition 3.1 *([8, Lemma 1]) Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ a nonempty closed convex set, and $\bar{u} \in U$. Assume that Φ is differentiable near $\bar{u}$ with Φ' being Lipschitz-continuous. Let the error bound condition (3.2) hold as $u \in P$ tends to $\bar{u}$.*

Then, for any solution u of the perturbed generalized equation

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni \omega, \tag{3.12}$$

close enough to $\bar{u}$, it holds that

$$\text{dist}(u, U) = O(\|\omega\|)$$

as $\omega \to 0$.

Recall that the first-order necessary optimality condition for the LM subproblem (1.3) has the form (1.5). For any given $u \in \mathbb{R}^p$, (3.11) is equivalent to saying that

$$(\Phi'(u+v))^\top \Phi(u+v) + N_P(u+v) \ni 0 \tag{3.13}$$

holds with $v = 0$. Then (1.5) can be regarded as a perturbation of the generalized equation (3.13). Specifically, if we set

$$\begin{aligned}\omega(u,\, v) &= (\Phi'(u+v))^\top \Phi(u+v) - (\Phi'(u))^\top (\Phi(u) + \Phi'(u)v) - \sigma(u)v \\ &= \left((\Phi'(u+v))^\top - (\Phi'(u))^\top\right) \Phi(u+v) \\ &\quad + (\Phi'(u))^\top (\Phi(u+v) - \Phi(u) - \Phi'(u)v) - \sigma(u)v,\end{aligned} \tag{3.14}$$

then (1.5) can be written in the form

$$(\Phi'(u+v))^\top \Phi(u+v) + N_P(u+v) \ni \omega(u,\, v).$$

From Lipschitz-continuity of Φ' near $\bar{u}$, and from (3.7) with $v = v(u)$, employing again the Mean-Value Theorem [45, Theorem A.10], one can readily derive the estimates

$$\begin{aligned}\|\omega(u,\, v(u))\| &= O((\mathrm{dist}(u,\, U))^{\theta+1}) + O((\mathrm{dist}(u,\, U))^2) \\ &= O((\mathrm{dist}(u,\, U))^{\min\{\theta+1,\, 2\}})\end{aligned} \tag{3.15}$$

as $u \in P \setminus U$ tends to $\bar{u}$.

Summarizing the considerations above, $u + v(u)$ is a solution of the generalized equation (3.12) with $\omega = \omega(u,\, v(u))$, and it holds that $u + v(u) \to \bar{u}$ and $\omega(u,\, v(u)) \to 0$ as $u \in P \setminus U$ tends to $\bar{u}$. Therefore, since we assume the constrained error bound condition to hold, Proposition 3.1 allows to conclude that

$$\begin{aligned}\mathrm{dist}(u + v(u),\, U) &= O(\|\omega(u,\, v(u))\|) \\ &= O((\mathrm{dist}(u,\, U))^{\min\{\theta+1,\, 2\}}) \\ &= o(\mathrm{dist}(u,\, U)) \\ &\le \frac{1}{2}\,\mathrm{dist}(u,\, U)\end{aligned} \tag{3.16}$$

for $u \in P \setminus U$ close enough to $\bar{u}$, where the second estimate is by (3.15), and the third one is by the requirement $\theta > 0$.

In order to complete the proof of the announced sharper result on the local superlinear convergence rate of the LM method when the regularization parameter is chosen according to (3.9) (see Theorem 3.2 below), we employ the following auxiliary result not related to any specific algorithm but rather giving sufficient conditions for convergence of a generic sequence, and with a certain Q-order of convergence.

Proposition 3.2 *([37, Lemma 2.9]) Let $\{u^k\} \subset \mathbb{R}^p$ and $\{r_k\} \subset \mathbb{R}_+$ be given sequences such that, for some $\rho \in [0,\ 1)$ and $R > 0$, $u^k \in B(u^0,\ Rr_0/(1-\rho))$ implies that*

$$\|u^{k+1} - u^k\| \leq Rr_k$$

and

$$r_{k+1} \leq \rho r_k$$

for every k.

Then $r_k \to 0$, and $\{u^k\}$ converges to some $u^* \in \mathbb{R}^p$.

If, in addition, there exist $C > 0$ and $\tau > 1$ such that

$$r_{k+1} \leq Cr_k^\tau \tag{3.17}$$

and

$$\|u^k - u^*\| \geq r_k \tag{3.18}$$

hold for all k, then $\{u^k\}$ converges to u^* with the Q-order τ.

Consider now any $u^0 \in P$ and the corresponding uniquely defined iterative sequence $\{u^k\}$ generated by the LM algorithm. For each k, we set $r_k = \operatorname{dist}(u^k,\ U)$. Furthermore, let $\rho = 1/2$, $\tau = \min\{\theta + 1,\ 2\}$, and let a constant $R > 0$ be fixed. If $u^k \in B(u^0,\ Rr_0/(1-\rho))$, then

$$\|u^k - \bar{u}\| \leq \|u^k - u^0\| + \|u^0 - \bar{u}\| \leq 2R \operatorname{dist}(u^0,\ U) + \|u^0 - \bar{u}\|,$$

where the right-hand side becomes arbitrarily small provided u^0 is close enough to $\bar{u}$, for any pre-fixed value of $R > 0$. Then (3.7) with $v = v(u^k)$ allows to take $R > 0$ large enough, so that

$$\|v(u^k)\| \leq R \operatorname{dist}(u^k,\ U)$$

provided u^0 is close enough to $\bar{u}$, and hence,

$$\|u^{k+1} - u^k\| = \|v(u^k)\| \leq Rr_k.$$

Moreover, employing (3.16), we also have that for such u^0 it holds that

$$r_{k+1} = \operatorname{dist}(u^{k+1},\ U) = \operatorname{dist}(u^k + v(u^k),\ U) \leq \frac{1}{2} \operatorname{dist}(u^k,\ U) = \rho r_k.$$

Therefore, the first part of the claim of Proposition 3.2 yields convergence of $\{\operatorname{dist}(u^k,\ U)\}$ to 0, and of $\{u^k\}$ to some $u^* \in \mathbb{R}^p$ that in this case necessarily belongs to U.

Furthermore, by the intermediate estimates in (3.16), one can take $C > 0$ large enough so that

$$\operatorname{dist}(u^k + v(u^k),\ U) \le C(\operatorname{dist}(u^k,\ U))^{\min\{\theta+1,\ 2\}}$$

provided u^0 is close enough to $\bar{u}$. Hence, (3.17) holds, while (3.18) is automatic. Therefore, the second part of the claim of Proposition 3.2 yields convergence with the Q-order $\min\{\theta + 1,\ 2\}$.

We have thus proven a particular case of [8, Theorem 1], stated as Theorem 3.2 below. For unconstrained equations, a similar result was obtained in [65] by somewhat different tools, involving singular-value decompositions (cf. [31]). In [65], the approach was further extended to the case of nonnegativity constraints on some variables, but with a weaker convergence result when compared to the theorem below. Finally, [27, Section 2] contains convergence results that are somewhat weaker than those in Theorem 1 of [8]; in particular, the same convergence rate estimate is obtained only by assuming that convergence is to a solution in the interior of P.

Theorem 3.2 *Under the assumptions of Proposition 3.1, let a function $\sigma : P \to \mathbb{R}_+$ satisfy (3.9) as $u \in P$ tends to $\bar{u}$, with some fixed $\theta \in (0,\ 2]$.*

Then, for every $u^0 \in P$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the solution of (1.3) with the Euclidean norms, with $u = u^k$ and $G(u) = \Phi'(u^k)$, and with the additional convention that $u^{k+1} = u^k$ if $u^k \in U$. For any $\delta > 0$, if $u^0 \in P$ is close enough to $\bar{u}$, then this sequence is contained in $B(\bar{u},\ \delta)$ and converges to some $u^* \in U$, and the rate of convergence is superlinear with the Q-order $\min\{\theta + 1,\ 2\}$.

The next example is concerned with the restriction $\theta \in (0,\ 2]$ in Theorem 3.2, and demonstrates that at least the requirement $\theta < 4$ cannot be avoided. To the best of our knowledge, the question regarding the values $\theta \in (2,\ 4)$ remains open. That said, [6, Example 4.2] claims the lack of quadratic convergence for $\theta > 3$, but this claim seems to be based on some numerical observations only. On the other hand, [30] additionally obtains superlinear convergence to 0 of the sequence $\{\operatorname{dist}(u^k,\ U)\}$ for $\theta \in (2,\ 3)$. A similar result and some of its extensions can also be found in the recent work [64].

Example 3.1 Let $p = q = 2$, $P = \mathbb{R}^2$, $\Phi(u) = (2u_1(1 + u_2),\ u_1^2)$. Observe that (1.1) with this data is the Lagrange optimality system for the equality-constrained optimization problem

$$\text{minimize } u_1^2 \text{ subject to } u_1^2 = 0,$$

with u_2 being the dual variable. This is a model problem used in [43] and other works on critical solutions (it also appears in DEGEN test collection [18] under the identifier 20101). We have $U = \{0\} \times \mathbb{R}$, and the unconstrained local Lipschitzian error bound condition holds at all solutions except for the solution $(0,\ -1)$.

The LM subproblem (1.5) with $G(u) = \Phi'(u)$ takes the form

$$4u_1(1 + u_2)^2 + 2u_1^3 + 4(u_1^2 + (1 + u_2)^2)v_1 + 4u_1(1 + u_2)v_2 + \sigma(u)v_1 = 0,$$

$$4u_1^2(1+u_2)+4u_1(1+u_2)v_1+4u_1^2v_2+\sigma(u)v_2=0,$$

and we are interested in those u that are close to some solution distinct from $(0,\ -1)$, say, to $\bar{u}=(0,\ 0)$. Then it makes sense to consider $u=(u_1,\ 0)$ with $u_1\neq 0$ but close to 0. Solving the iteration system above, we then obtain

$$\begin{aligned} v_1 &= -u_1-u_1v_2-\frac{\sigma(u)}{4u_1}v_2,\\ &\left(u_1^2+\frac{\sigma(u)}{2}+\frac{\sigma(u)}{4u_1^2}+\frac{(\sigma(u))^2}{16u_1^2}\right)v_2=-\frac{u_1^2}{2}-\frac{\sigma(u)}{4}. \end{aligned} \tag{3.19}$$

Let $\sigma(u)=\|\Phi(u)\|^\theta$ with $\theta\geq 4$. Then $\sigma(u)=(2u_1)^\theta+o(u_1^\theta)=O(u_1^4)$ as $u_1\to 0$, and from (3.19) we have

$$\left(u_1^2+(2u_1)^{\theta-2}+o(u_1^2)\right)v_2=-\frac{u_1^2}{2}+o(u_1^2),$$

implying that v_2 is separated from 0 by some negative constant independent of u_1 close enough to 0. This demonstrates that the claim of Theorem 3.2 is not valid for $\theta\geq 4$, as the next iterate will not stay in $B(\bar{u},\ \delta)$ with any pre-fixed small $\delta>0$, no matter how close u_1 is to 0.

Running the LM method with $\theta\geq 4$ on this example, the iterative sequences are observed to converge to the unique critical solution $(0,\ -1)$ (rather than to any solution close to a starting point), and at a linear (rather than quadratic) rate with the asymptotic common ratio $1/2$. This is the typical behavior of the pure Newton method when converging to a critical solution; see [41], and also [43] and references therein.

We now briefly discuss local convergence results for projected LM methods under the smoothness requirements of this section, and with P being convex. Local quadratic convergence was established in [50] for the regularization parameter function $\sigma(\cdot)$ satisfying (3.9) with $\theta=2$, and under the assumption that (3.2) holds as $u\to\bar{u}$, where u need not be restricted to P. Beyond the unconstrained case, this assumption is very restrictive, in particular, because it subsumes that $\Phi^{-1}(0)\subset U$ near $\bar{u}$. The same conclusion under the same restrictive assumption was derived in [37], however for a larger regularization parameter satisfying (3.9) with $\theta=1$.

The restrictive error bound assumption in [50] and [37] was relaxed in [10], where it was replaced by the combination of (3.2) as $u\in P$ tends to $\bar{u}$ and the unconstrained error bound

$$\operatorname{dist}(u,\ \Phi^{-1}(0))=O(\|\Phi(u)\|)$$

as $u\to\bar{u}$. It is well-known that the regularity condition $\operatorname{rank}\Phi'(\bar{u})=q$ is sufficient for the unconstrained error bound to hold. Related necessary conditions can be found in [7]. However, under the specified assumptions, only local R-linear convergence was established in [10] for the regularization parameter satisfying (3.9) with $\theta=1$. This result was sharpened in [9], where it was shown that R-linear convergence holds under

the constrained error bound only, i.e., without additionally assuming the unconstrained error bound to hold.

The work [27] provides superlinear convergence of the projected LM method under the same assumptions as in [50] and [37], but for all $\theta \in (0, 2]$, and involving an interiority assumption for obtaining the exact convergence rate estimate in some cases.

For unconstrained equations (i.e., when $P = \mathbb{R}^p$), the particular cases of the results discussed in this section can be found in [17, 28, 31, 32, 37, 63, 65].

We now turn our attention to local convergence of the LPN method under the smoothness requirements of this section, and assume that $G(u) = \Phi'(u)$ for $u \in \mathbb{R}^p$ close enough to $\bar{u}$. Since Assumptions 1 and 3 are not algorithm-related, and are shown above to be satisfied in the current setting, it remains to verify that Assumption 2 holds under the constrained error bound (3.2) as $u \in P$ tends to $\bar{u}$ by demonstrating that the LPN step v from the current iterate u belongs to $\mathcal{F}(u, \Gamma)$ with fixed $\Gamma > 0$ large enough, for $u \in P$ close enough to $\bar{u}$. Considering the relation between the set-valued mapping $\mathcal{F}$ defined in (2.6) and the constraints in (1.4), it is sufficient to show that the optimal value $\gamma(u)$ of (1.4) is bounded from above for $u \in P$ near $\bar{u}$.

To do that, observe first that $\gamma(u) = 0$ for any $u \in U$. Furthermore, for $u \in P \setminus U$, define $\widehat{u} \in U$ as above (i.e., satisfying (3.4)), and then, as in (3.6), we obtain that

$$\|\Phi(u) + \Phi'(u)(\widehat{u} - u)\| = O(\|\widehat{u} - u\|^2) = O((\operatorname{dist}(u, U))^2) = O(\|\Phi(u)\|^2),$$

and

$$\|\widehat{u} - u\| = \operatorname{dist}(u, U) = O(\|\Phi(u)\|)$$

as $u \in P$ tends to $\bar{u}$, where the last estimates in these two chains of relations follow from (3.2). This implies the existence of $\Gamma > 0$ such that $(\widehat{u} - u, \Gamma)$ is feasible in (1.4) for $u \in P$ near $\bar{u}$, and hence, $\gamma(u) \leq \Gamma$ for such u.

Applying now Theorem 2.2 we obtain the following counterpart of [23, Theorem 1]. In the latter, the assumption that $\gamma(u)$ is bounded for $u \in P$ near $\bar{u}$ (actually equivalent to Assumption 2, as pointed out in [34]) was imposed directly, together with Assumption 3, and thus avoiding the need to use the specific G.

Theorem 3.3 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ a nonempty closed set, and $\bar{u} \in U$. Assume that Φ is differentiable near $\bar{u}$ with Φ' being Lipschitz-continuous. Let (3.2) hold as $u \in P$ tends to $\bar{u}$.*

Then, for every $u^0 \in P$, there exists a sequence $\{u^k\}$ such that for every k, the pair $(u^{k+1} - u^k, \gamma_k)$ with some scalar γ_k is a solution of (1.4) with $u = u^k$ and $G(u) = \Phi'(u^k)$. For any $\delta > 0$, if $u^0 \in P$ is close enough to $\bar{u}$, then any such sequence is contained in $B(\bar{u}, \delta)$ and converges to some $u^* \in U$, and the rate of convergence is quadratic.

Local convergence of some quasi-Newton versions of the LPN method was investigated in [54, 55].

One observation is in order regarding the unconstrained case. The first-order necessary optimality condition (1.5) for the unconstrained ($P = \mathbb{R}^p$) LM subproblem (1.3)

with Euclidean norms takes the form of the linear equation

$$((G(u))^\top G(u) + \sigma(u)\mathcal{I})v = -(G(u))^\top \Phi(u), \tag{3.20}$$

and this condition is also sufficient for optimality in (1.3), due to the convexity of the objective function of the latter. At the same time, the LPN subproblem

$$\begin{aligned} \text{minimize } \gamma \text{ subject to } & \|\Phi(u) + G(u)v\|_\infty \le \gamma \|\Phi(u)\|^2, \\ & \|v\|_\infty \le \gamma \|\Phi(u)\|, \\ & \gamma \ge 0, \end{aligned}$$

is still a linear programming problem, and solving it is in general more costly than solving the linear system (3.20).

The convergence rate estimate in Theorem 3.2 is sharp even in the unconstrained case, and even when the solution $\bar{u}$ in question is isolated. This is demonstrated by the following simple example, also showing that the error bound condition (3.2) cannot be dropped in Theorems 3.1 and 3.2.

Example 3.2 Let $p = q = 1$, $P = \mathbb{R}$, $\Phi(u) = au + bu^2$, with a and b being scalar parameters. For any value of these parameters, $\bar{u} = 0$ is a solution of (1.1), and since $\Phi'(0) = a$, the unconstrained local Lipschitzian error bound condition holds at this solution provided $a \ne 0$.

The LM subproblem (1.5) with $G(u) = \Phi'(u) = a + 2bu$ reduces to the equation

$$(a + 2bu)(au + bu^2 + (a + 2bu)v) + \sigma(u)v = 0,$$

and if $\sigma(u) > 0$, its unique solution is

$$v(u) = -\frac{(a + 2bu)(a + bu)u}{(a + 2bu)^2 + \sigma(u)}.$$

Then

$$\begin{aligned} u + v(u) &= \frac{((a + 2bu)^2 + \sigma(u) - (a + 2bu)(a + bu))u}{(a + 2bu)^2 + \sigma(u)} \\ &= \frac{abu^2 + \sigma(u)u + 2b^2u^3}{(a + 2bu)^2 + \sigma(u)}. \end{aligned}$$

Take $\sigma(u) = |\Phi(u)|^\theta = |au + bu^2|^\theta$ with some $\theta > 0$. Thus, for $u \ne 0$, we have

$$u + v(u) = \frac{|au + bu^2|^\theta u + abu^2 + 2b^2u^3}{(a + 2bu)^2 + |au + bu^2|^\theta}.$$

If $a \ne 0$ and $b \ne 0$, since the denominator above tends to a^2 as $u \to 0$, the iterative sequence generated this way converges to 0 superlinearly, with the Q-order being precisely $\min\{\theta + 1,\ 2\}$.

On the other hand, if $a = 0$, but $b \neq 0$, then, say, for $u > 0$,

$$u + v(u) = \frac{|b|^\theta u^{2\theta+1} + 2b^2u^3}{|b|^\theta u^{2\theta} + 4b^2u^2} = \frac{|b|^\theta u^{2\theta} + 2b^2u^2}{|b|^\theta u^{2\theta} + 4b^2u^2}u.$$

If $\theta > 1$, this yields linear convergence with asymptotic common ratio $1/2$. If $\theta = 1$, the rate is still linear but with asymptotic common ratio $(|b| + 2b^2)/(|b| + 4b^2)$. Finally, if $0 < \theta < 1$, the convergence rate is sublinear. In particular, none of these cases agrees with the assertions of Theorems 3.1 and 3.2, and the reason is that the unconstrained local Lipschitzian error bound does not hold at $\bar{u} = 0$ when $a = 0$. Observe that the convergence rate estimate for $\theta > 1$ agrees with the one obtained in [43] (under the stronger restriction $\theta \geq 3/2$) for the case of singular solutions of unconstrained equations satisfying some 2-regularity condition, which holds in this example when $a = 0$.

Finally, consider the case when $a \neq 0$, but $b = 0$. Then

$$u + v(u) = \frac{|au|^\theta u}{a^2 + |au|^\theta},$$

and convergence is again superlinear, but with the Q-order being $\theta + 1$. Therefore, the actual convergence rate in this case is actually the higher the larger is θ, becoming quadratic for $\theta \geq 1$, cubic for $\theta \geq 2$, etc.

In other words, the "worst case" superlinear convergence rate estimate with Q-order $\min\{\theta + 1,\ 2\}$ is sharp under the assumptions of Theorem 3.3, but for particular problem instances, larger values of θ may give faster convergence.

To conclude this section, we provide some clarification to the observation above that Theorem 2.2 was derived in [22] under the additional error bound condition. In fact, under the smoothness assumptions of this section, and in the unconstrained case, not only Assumption 2 is implied by the error bound condition (3.2) as $u \to \bar{u}$ (a fact demonstrated above in this section), but the converse implication is true as well. We show this next.

We can assume, without loss of generality, that $\bar{u} = 0$. Let Π be the orthogonal projector onto $(\operatorname{im} \Phi'(0))^\perp$. We make use of the uniquely defined decomposition of every $u \in \mathbb{R}^p$ into the sum $u = u_1 + u_2$, with $u_1 \in (\ker \Phi'(0))^\perp$ and $u_2 \in \ker \Phi'(0)$. Following the Lyapunov–Schmidt procedure (see, e.g., [40, Chapter VII]), one can replace the equation in (1.1) by the equivalent system

$$(\mathcal{I} - \Pi)\Phi(u_1 + u_2) = 0, \quad \Pi\Phi(u_1 + u_2) = 0, \tag{3.21}$$

and consider the first equation in (3.21) with respect to u_1, with u_2 treated as a parameter. By the classical Implicit Function Theorem, for every u_2 close enough to 0, this equation has the unique near 0 solution $u_1(u_2) \in (\ker \Phi'(0))^\perp$, the function $u_1(\cdot)$ inherits the smoothness of Φ, and necessarily $u_1(0) = 0$, and $u_1'(0) = 0$. Then one substitutes $u_1 = u_1(u_2)$ into the second equation, and obtains an equation with respect to u_2 only.

Let $\{u_2^k\} \subset \ker \Phi'(0)$ be any sequence convergent to 0, and such that for every k it holds that $\Phi(u^k) \neq 0$, where we set $u^k = u_1(u_2^k) + u_2^k$. Suppose further that for every k there exists $v^k \in \mathbb{R}^p$ such that

$$\|\Phi(u^k) + \Phi'(u^k)v^k\| = O(\|\Phi(u^k)\|^2), \quad \|v^k\| = O(\|\Phi(u^k)\|)$$

as $k \to \infty$. Since $\Phi(u^k) = \Pi\Phi(u^k)$, from the first estimate we then have

$$\left\| \frac{\Phi(u^k)}{\|\Phi(u^k)\|} + \Pi\Phi'(u^k)\frac{v^k}{\|\Phi(u^k)\|} \right\| = O(\|\Phi(u^k)\|),$$

where the norm of the first term in the left-hand side equals 1, while the sequence $\{v^k/\|\Phi(u^k)\|\}$ is bounded. Moreover, $\{\Pi\Phi'(u^k)\} \to 0$, and hence the left-hand side tends to 1, while the right-hand side tends to 0. This yields a contradiction.

Therefore, Assumption 2 may only hold if $\Phi(u_1(u_2)+u_2) = \Pi\Phi(u_1(u_2)+u_2) = 0$ for all $u_2 \in \ker \Phi'(0)$ close enough to 0. But this property implies that the solution set $\Phi^{-1}(0)$ coincides near 0 with $S = \{u \in \mathbb{R}^p \mid u_1 = u_1(u_2)\}$. Indeed, by this property, S near 0 is contained in $\Phi^{-1}(0)$. On the other hand, if we suppose that there exists $u \in \Phi^{-1}(0)$, arbitrarily close to 0, and such that $u_1 \neq u_1(u_2)$, we have that this u_1 solves the first equation in (3.21) for the given u_2, yielding a contradiction with the fact that $u_1(u_2)$ is the unique near 0 solution of that equation for u_2 close enough to 0.

Observe, finally, that since $u_1'(0) = 0$, the set S is a smooth manifold near 0 (hence Clarke-regular at 0) with the tangent subspace at 0 equal to $\ker \Phi'(0)$. Indeed, it holds that $S = F^{-1}(u)$, with $F : \mathbb{R}^p \to (\ker \Phi'(0))^\perp$, $F(u) = u_1 - u_1(u_2)$, and

$$F'(0)v = v_1 - u_1'(0)v_2 = v_1, \quad v \in \mathbb{R}^p,$$

implying that $\operatorname{im} F'(0) = (\ker \Phi'(0))^\perp$, $\ker F'(0) = \{v \in \mathbb{R}^p \mid v_1 = 0\} = \ker \Phi'(0)$, yielding the needed conclusion about S. According to [44, Definition 1], this conclusion further means that 0 is a noncritical solution of the equation in (1.1), which is equivalent to saying that the local Lipschitzian error bound holds at 0 (see [44, Theorem 1]).

For constrained equations, the question remains open whether the constrained error bound is implied by Assumption 2.

4 Local convergence in the piecewise smooth case

Our goal in this section is to relax the smoothness assumptions of Sect. 3. Specifically, we consider the constrained equation (1.1) with a mapping Φ being piecewise smooth near a solution $\bar{u}$. This means that there exists a finite collection of selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^q$ which are continuously differentiable near $\bar{u}$, such that

$$\Phi(u) \in \{\Phi^1(u), \ldots, \Phi^s(u)\} \quad \forall\, u \in \mathbb{R}^p,$$

and Φ is continuous near $\bar{u}$. In particular, taking $s = 1$ brings us back to the smooth case considered in Sect. 3.

For every $u \in \mathbb{R}^p$, we define the set

$$\mathcal{A}(u) = \{j \in \{1, \ldots, s\} \mid \Phi(u) = \Phi^j(u)\} \tag{4.1}$$

of indices of all selection mappings active at u. Let G be any mapping such that

$$G(u) \in \{(\Phi^j)'(u) \mid j \in \mathcal{A}(u)\} \tag{4.2}$$

for $u \in \mathbb{R}^p$ close enough to $\bar{u}$. In the discussion below, we do not assume any control of the choice of the specific value of $G(u)$ when there is more than one active selection mapping (i.e., $\mathcal{A}(u)$ is not a singleton). This is consistent with the black-box paradigm in nonsmooth optimization; see, e.g., [16, Part II].

The algorithm with the subproblem (1.3) employing G defined by (4.1) and (4.2), is naturally referred to as the constrained *piecewise* LM method. Observe that with this definition of G, for any $u \in \mathbb{R}^p$ close enough to $\bar{u}$, the subproblem (1.3) can be written in the form

$$\text{minimize } \frac{1}{2}\|\Phi^j(u) + (\Phi^j)'(u)v\|^2 + \frac{1}{2}\sigma(u)\|v\|^2 \text{ subject to } u + v \in P, \tag{4.3}$$

with *some* $j \in \mathcal{A}(u)$, and this can be seen as the subproblem of the LM method applied to a *smooth* constrained equation

$$\Phi^j(u) = 0, \quad u \in P. \tag{4.4}$$

Needless to say, the index j can vary from one iteration to another.

We are now going to establish local quadratic convergence of the constrained piecewise LM method by means of Theorem 2.2. To that end, we need to provide conditions ensuring the fulfilment of Assumptions 1–3, and to demonstrate that the method fits the framework of that theorem. This discussion essentially follows the exposition in [34], which in turn relies on the results in [22, 23]. To begin with, since a mapping that is piecewise smooth near some point is necessarily Lipschitz-continuous near that point [42, Theorem 2.1], Assumption 1 is automatically satisfied in our current setting.

From now on in this section, we assume that $(\Phi^j)'$ is Lipschitz-continuous near $\bar{u}$ for all $j \in \mathcal{A}(\bar{u})$. In order to verify Assumption 3, we observe first that, by continuity of Φ and its smooth selections at $\bar{u}$, from (4.1) it follows that $\mathcal{A}(u) \subset \mathcal{A}(\bar{u})$ for all $u \in \mathbb{R}^p$ close enough to $\bar{u}$. For any such u, picking up $j \in \mathcal{A}(u)$ such that $G(u) = (\Phi^j)'(u)$ (see (4.2)), and employing again the Mean-Value Theorem [45, Theorem A.10], for any $v \in \mathbb{R}^p$ satisfying (2.7), we obtain that

$$\begin{aligned}
\|\Phi^j(u+v)\| &\le \|\Phi^j(u) + (\Phi^j)'(u)v\| + \|\Phi^j(u+v) - \Phi^j(u) - (\Phi^j)'(u)v\| \\
&= \|\Phi(u) + G(u)v\| + O(\|v\|^2) \\
&\le t^2 + O(\|v\|^2) \\
&= O(t^2)
\end{aligned} \tag{4.5}$$

as $u \to \bar{u}$ and $t \to 0+$.

At this point we need to introduce an additional requirement on the piecewise smooth structure of Φ: let

$$\|\Phi(u)\| = O(\|\Phi^j(u)\|) \quad \forall\, j \in \mathcal{A}(\bar{u}) \tag{4.6}$$

as $u \in P$ tends to $\bar{u}$. Combined with (4.5), this readily yields the fulfilment of Assumption 3.

In order to get a better understanding of the nature of condition (4.6), we next recall the *P-property* at $\bar{u}$, introduced in [34, p. 434]. It consists of saying that near $\bar{u}$ the constraint set P excludes all zeroes of smooth selections active at $\bar{u}$, which are not zeroes of Φ. Formally it means that $U_j \subset U$ near $\bar{u}$ for all $j \in \mathcal{A}(\bar{u})$, where U_j stands for the solution set of (4.4). Obviously, the P-property at $\bar{u}$ is implied by (4.6) as $u \in P$ tends to $\bar{u}$. As demonstrated in [34, Example 2], the converse implication does not hold in general, but it holds under the additional requirement that

$$\operatorname{dist}(u,\, U_j) = O(\|\Phi^j(u)\|) \quad \forall\, j \in \mathcal{A}(\bar{u}) \tag{4.7}$$

as $u \in P$ tends to $\bar{u}$. This is the constrained error bound for each active selection. For any $j \in \mathcal{A}(\bar{u})$ and $u \in P$, let $\widehat{u}^j$ stand for a metric projection of u onto U_j:

$$\|u - \widehat{u}^j\| = \operatorname{dist}(u,\, U_j). \tag{4.8}$$

From (4.7) we then obtain that

$$\|u - \widehat{u}^j\| = O(\|\Phi^j(u)\|)$$

as $u \in P$ tends to $\bar{u}$. Evidently, $\widehat{u}^j \to \bar{u}$ as $u \to \bar{u}$, and hence, the P-property at $\bar{u}$ implies that $\widehat{u}^j \in U$ provided u is close enough to $\bar{u}$. Therefore, by the Lipschitz-continuity of Φ near $\bar{u}$, and by (4.8), we obtain the estimate

$$\|\Phi(u)\| = \|\Phi(u) - \Phi(\widehat{u}^j)\| = O(\|u - \widehat{u}^j\|) = O(\|\Phi^j(u)\|),$$

i.e., (4.6) holds as $u \in P$ tends to $\bar{u}$.

As a side observation, employing again the inclusion $\mathcal{A}(u) \subset \mathcal{A}(\bar{u})$ for $u \in \mathbb{R}^p$ close enough to $\bar{u}$, one can easily see that the P-property at $\bar{u}$, and the condition (4.7) as $u \in P$ tends to $\bar{u}$, imply the constrained error bound (3.2) as $u \in P$ tends to $\bar{u}$.

We finally need to verify Assumption 2 by showing that $v \in \mathcal{F}(u,\, \Gamma)$ holds for a step v of the constrained piecewise LM method, provided $\Gamma > 0$ is fixed large enough and $u \in P$ is close enough to $\bar{u}$.

Let the regularization parameter satisfy (3.1) as $u \in P$ tends to $\bar{u}$. For $u \in \mathbb{R}^p$ close to $\bar{u}$, and for $j \in \mathcal{A}(u) \subset \mathcal{A}(\bar{u})$ such that $G(u) = (\Phi^j)'(u)$ (see (4.2)), we have that $\Phi^j(u) = \Phi(u)$ (see (4.1)). Then (3.1) implies the estimates

$$\|\Phi^j(u)\|^2 = O(\sigma(u)), \quad \sigma(u) = O(\|\Phi^j(u)\|^2) \tag{4.9}$$

as $u \in P$ tends to $\bar{u}$.

As discussed above, an iteration of the constrained piecewise LM method can be interpreted as an iteration of the usual constrained LM method for the smooth constrained equation (4.4), with the subproblem (4.3). Using (4.9), we can now apply the corresponding argument in Sect. 3 with Φ substituted by Φ^j, to derive the estimates

$$\|v\| = O(\mathrm{dist}(u, U_j))$$

and

$$\|\Phi^j(u) + (\Phi^j)'(u)v\| = O((\mathrm{dist}(u, U_j))^2)$$

as $u \in P \setminus U$ tends to $\bar{u}$. To obtain the needed conclusion, it remains to recall again (4.7) and the equalities $\Phi^j(u) = \Phi(u)$ and $G(u) = (\Phi^j)'(u)$.

Putting together all the ingredients above, we arrive to the following local convergence result for the constrained piecewise LM method, corresponding to [34, Theorems 1, 2].

Theorem 4.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ a nonempty closed set, and $\bar{u} \in U$. Assume that Φ is piecewise smooth near $\bar{u}$, and the derivatives of its smooth selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^q$ are Lipschitz-continuous near $\bar{u}$. Let the P-property at $\bar{u}$ and condition (4.7) be satisfied as $u \in P$ tends to $\bar{u}$. Moreover, let $G : \mathbb{R}^p \to \mathbb{R}^{q \times p}$ be a fixed mapping satisfying (4.2), and assume that the function $\sigma : P \to \mathbb{R}_+$ satisfies (3.1) as $u \in P$ tends to $\bar{u}$.*

Then, for every $u^0 \in P$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is a solution of (1.3) with $u = u^k$, and with the additional convention that $u^{k+1} = u^k$ if $u^k \in U$. For any $\delta > 0$, if $u^0 \in P$ is close enough to $\bar{u}$, then any such sequence is contained in $B(\bar{u}, \delta)$ and converges to some $u^* \in U$, and the rate of convergence is quadratic.

Although Theorem 2.2 does not impose any explicit differentiability condition on the mapping Φ, it is difficult to apply this theorem for cases beyond piecewise smoothness of Φ when there are solutions that are nonisolated and at which Φ is not differentiable. If Φ is a nonsmooth (but not piecewise smooth) reformulation of a complementarity system, i.e., when the nondifferentiability comes from the complementarity function, we know about just two approaches [11] and [38] to get superlinear convergence by means of LM methods. Both approaches, employing the Fischer–Burmeister complementarity function, allow significantly larger steps (if compared to Theorem 2.2). Whereas the approach in [11] is based on longer steps for the multipliers of a Karush-Kuhn-Tucker system associated to a variational inequality, the article [38] exploits the freedom in the very recent framework described in Theorem 2.1 and allows significantly larger steps without a restriction to a special class of variables. Based on this and on a new error bound condition allowing again certain problems with nonisolated solutions, an LM method for complementarity systems is shown to exhibit superlinear convergence.

We now turn our attention to the *piecewise* LPN method, that is, the algorithm with the subproblem (1.4) employing G defined by (4.1), (4.2). Similarly to the constrained

piecewise LM method, the subproblem (1.4) can then be written in the form

$$\begin{aligned} \text{minimize } \gamma \text{ subject to } & \|\Phi^j(u) + (\Phi^j)'(u)v\| \le \gamma \|\Phi^j(u)\|^2, \\ & \|v\| \le \gamma \|\Phi^j(u)\|, \\ & u + v \in P,\ \gamma \ge 0, \end{aligned} \tag{4.10}$$

with some $j \in \mathcal{A}(u)$, and this is the subproblem of the LPN method applied to a smooth constrained equation (4.4).

In order to derive local quadratic convergence of the piecewise LPN method from Theorem 2.2, we first recall yet again that Assumptions 1 and 3 are established above, under the appropriate requirements. Assumption 2 is established through the interpretation (4.10) of the piecewise LPN subproblem, similarly to how this is done above for the piecewise LM method, but employing the related argument for the LPN method before Theorem 3.3. The result obtained this way corresponds to [34, Theorems 1, 2].

Theorem 4.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ a nonempty closed set, and $\bar{u} \in U$. Assume that Φ is piecewise smooth near $\bar{u}$, and the derivatives of its smooth selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^q$ are Lipschitz-continuous near $\bar{u}$. Let the P-property at $\bar{u}$ and condition (4.7) be satisfied as $u \in P$ tends to $\bar{u}$. Moreover, let $G : \mathbb{R}^p \to \mathbb{R}^{q\times p}$ be a fixed mapping satisfying (4.2).*

Then, for every $u^0 \in P$, there exists a sequence $\{u^k\}$ such that for every k, the pair $(u^{k+1} - u^k,\ \gamma_k)$ with some scalar γ_k is a solution of (1.4) with $u = u^k$. For any $\delta > 0$, if $u^0 \in P$ is close enough to $\bar{u}$, then any such sequence is contained in $B(\bar{u},\ \delta)$ and converges to some $u^* \in U$, and the rate of convergence is quadratic.

Let us finally mention that the P-property, required in Theorems 4.1 and 4.2, can be easily guaranteed for reformulations of complementarity systems by means of the "min" (natural residual) complementarity function, see [23, 34].

5 Effect of inexactness

Solving the constrained LM subproblems exactly can be computationally too costly, or even impossible. The goal of this section is to characterize the "level" of controllable inexactness that does not interfere with local convergence and rate of convergence properties of the LM method established in Sect. 3. To that end, we shall restrict ourselves in this section to the smoothness requirements of Sect. 3, i.e., Φ is differentiable near a solution $\bar{u}$ of (1.1) and Φ' is Lipschitz-continuous near $\bar{u}$. That said, we note that Algorithm 5.1 and Theorem 5.1 below, coming from [22], can actually be extended to the piecewise smooth setting of Sect. 4, and even to more general kinds of nonsmoothness. For those extensions, as well as for the original source of Algorithm 5.1 and Theorem 5.1, we refer the reader to [22].

Algorithm 5.1 Choose the parameters $\gamma > 0$ and $\Theta > 1$. Choose $u^0 \in P$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.

2. Set $\sigma(u^k) = \|\Phi(u^k)\|^2$. If there exists $v \in P - u^k$ such that

$$\|\Phi(u^k) + \Phi'(u^k)v\|^2 + \sigma(u^k)\|v\|^2 \le \gamma^2\|\Phi(u^k)\|^4, \tag{5.1}$$

define v^k as any such v, and go to Step 4.

3. Compute v^k as a solution of (1.3) with $u = u^k$, $G(u) = \Phi'(u^k)$, and set

$$\gamma = \Theta\frac{\sqrt{\|\Phi(u^k) + \Phi'(u^k)v^k\|^2 + \sigma(u^k)\|v^k\|^2}}{\|\Phi(u^k)\|^2}. \tag{5.2}$$

4. Set $u^{k+1} = u^k + v^k$, increase k by 1 and go to Step 1.

The idea of this algorithm can be explained as follows: the process of solving the LM subproblem (1.3) is terminated once the value of the objective function of this subproblem becomes small enough; specifically, once the condition (5.1) at Step 2 is achieved. If this never happens for a given k, then the subproblem would need to be solved exactly at Step 3. But as will be demonstrated in Theorem 5.1 below, Step 2 always takes effect from some point on at least (thus skipping solving the subproblem exactly at Step 3).

According to the analysis preceding Theorem 3.1, under the hypotheses therein, Assumptions 1–3 are necessarily satisfied. Observe that the subproblem (1.3) in Step 3 of Algorithm 5.1 is always solvable. Hence, for every k, the algorithm successfully defines a step v^k. Moreover, the mentioned analysis also shows that if this v^k is indeed produced using Step 3, it belongs to $\mathcal{F}(u^k, \Gamma^*)$ with some fixed $\Gamma^* > 0$, for all $u^k \in P$ close enough to $\bar{u}$. Therefore, in order to apply Theorem 2.2 to Algorithm 5.1, it is sufficient to show that there exists $\Gamma \ge \Gamma^*$ such that for each k, if v^k is produced using Step 2, it belongs to $\mathcal{F}(u^k, \Gamma)$. This can be done by induction, assuming that u^0 is close enough to $\bar{u}$.

The key ingredients of the proof are as follows. If v^k is taken satisfying (5.1), then

$$\begin{aligned}\max\{\|\Phi(u^k) + \Phi'(u^k)v^k\|^2, \sigma(u^k)\|v^k\|^2\} &\le \|\Phi(u^k) + \Phi'(u^k)v^k\|^2 + \sigma(u^k)\|v^k\|^2\\ &\le \gamma^2\|\Phi(u^k)\|^4,\end{aligned}$$

and according to (2.6) and the definition of $\sigma(u^k)$ at Step 2, this implies that $v^k \in \mathcal{F}(u^k, \gamma)$, and γ stays unchanged. On the other hand, if v^k is a constrained LM direction defined at Step 3, then according to the inclusion $v^k \in \mathcal{F}(u^k, \Gamma^*)$, (2.6), and (5.2), the new value of γ satisfies $\gamma \le \sqrt{2}\Theta\Gamma^*$. Therefore, as long as u^k stays close enough to $\bar{u}$, the values of γ remain bounded above by $\Gamma = \max\{\gamma_0, \sqrt{2}\Theta\Gamma^*\}$, with γ_0 being the value of this parameter taken at initialization of the algorithm. It remains to remove the requirement that u^k stays close enough to $\bar{u}$, and this is what is done by induction, employing also the claim in Theorem 2.2 that $\{u^k\}$ is contained in $B(\bar{u}, \delta)$. This completes the proof of the fact that $v^k \in \mathcal{F}(u^k, \Gamma)$ for all k.

For the sake of formal consistency, we shall adopt the convention that when Algorithm 5.1 is terminated at Step 1 for some k, it is considered to still generate an infinite sequence with all the subsequent elements equal to u^k. The same convention will be used for the algorithms discussed in Sect. 6.

Theorem 5.1 *Under the assumptions of Theorem 3.1, Algorithm 5.1 successfully generates a sequence* $\{u^k\}$. *For any* $\delta > 0$, *if* u^0 *is close enough to* $\bar{u}$, *then any such sequence is contained in* $B(\bar{u}, \delta)$ *and converges to some* $u^* \in U$, *and the rate of convergence is quadratic. Moreover, starting from some* k, *Step 3 of Algorithm 5.1 is always avoided.*

The last claim of this theorem is obtained by the following reasoning. The fraction in the right-hand side of (5.2) is greater than the current value of γ, since otherwise, the exact constrained LM direction would satisfy (5.1). Since $\Theta > 1$, this implies that the sequence of values of γ is nondecreasing, and since this sequence was demonstrated to be bounded above (by the specified Γ), γ can be changed a finite number of times only. This means that, from some iteration on, Step 3 of the algorithm is never activated.

We proceed with the analysis from [8], for the case when the regularization parameter satisfies the requirements in (3.9) with some $\theta \in (0, 2]$ (i.e., not necessarily for $\theta = 2$). Assume now that P is convex, and let the norms used in (1.3) be Euclidean. Recall that in this case, the first-order necessary optimality condition for (1.3) has the form (1.5). We consider the version of the inexact constrained LM method, with inexactness measured by the violation of (1.5). Specifically, the process of solving the subproblem (1.3) with $G(u) = \Phi'(u)$ is terminated once

$$(\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) + \sigma(u)v + N_P(u + v) \ni w \tag{5.3}$$

is satisfied with some $w \in \mathbb{R}^p$ smaller than the given tolerance. Even more specifically, we assume that the "size" of inexactness conforms with the exponent θ in (3.9) by the requirement

$$\|w\| = O(\|\Phi(u)\|^{\theta+1}) \tag{5.4}$$

as $u \in P$ tends to $\bar{u}$.

One essential trick in [8] is to observe that (5.3) is a necessary and sufficient optimality condition for the following convex optimization problem, which is a perturbation of (1.3):

$$\text{minimize } \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{2}\sigma(u)\|v\|^2 - \langle w,\ v\rangle \text{ subject to } u + v \in P. \tag{5.5}$$

Then we follow the argument used above to prove (3.7), but taking into account that now the objective function of (5.5) has an extra term involving w.

Specifically, for a metric projection $\widehat{u}$ of $u \in P \setminus U$ onto U, and for the unique global solution v of (5.5), similarly to (3.5) we obtain that

$$\|\Phi(u) + \Phi'(u)v\|^2 + \sigma(u)\|v\|^2 - 2\langle w,\ v\rangle \le \|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 + \sigma(u)\|\widehat{u} - u\|^2 - 2\langle w,\ \widehat{u} - u\rangle.$$

Then, similarly to (3.6) (and in particular, making use of (3.2) and (3.4)), we derive the estimate

$$\begin{aligned}\|v\|^2 &\le \frac{1}{\sigma(u)}\Big(\|\Phi(u)+\Phi'(u)(\widehat{u}-u)\|^2\\ &\quad +\sigma(u)\|\widehat{u}-u\|^2+2\langle w,\,v\rangle-2\langle w,\,\widehat{u}-u\rangle\Big)\\ &\le \frac{2\|w\|}{\sigma(u)}(\|v\|+\operatorname{dist}(u,\,U))+O((\operatorname{dist}(u,\,U))^2)\\ &=O(\|\Phi(u)\|(\|v\|+\operatorname{dist}(u,\,U)))+O((\operatorname{dist}(u,\,U))^2)\\ &=O(\operatorname{dist}(u,\,U)\|v\|)+O((\operatorname{dist}(u,\,U))^2) \end{aligned} \tag{5.6}$$

as $u\in P\setminus U$ tends to $\bar{u}$, where the next-to-the-last estimate is by (5.4) and the first relation in (3.9), and the last one is by Assumption 1 that holds under the current smoothness requirements. Evidently, (5.6) implies (3.7) as $u\in P\setminus U$ tends to $\bar{u}$.

The remaining part of the argument in [8] consists of the reasoning used above to prove Theorem 3.2, but with $\omega(u,\,v)$ defined in (3.14) replaced by $\omega(u,\,v)+w$. Taking again into account (5.4), we conclude that this modification does not affect (3.15) and the subsequent analysis. This yields the full version of [8, Theorem 1].

Theorem 5.2 *Under the assumptions of Proposition 3.1, let the function $\sigma:P\to\mathbb{R}_+$ satisfy (3.9) with some fixed $\theta\in(0,\,2]$, and the function $\psi:P\to\mathbb{R}_+$ satisfy $\psi(u)=O(\|\Phi(u)\|^{\theta+1})$ as $u\in P$ tends to $\bar{u}$.*

Then, for every $u^0\in P$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1}-u^k$ is the solution of (5.3) with $u=u^k$, with some $w\in\mathbb{R}^p$ satisfying $\|w\|\le\psi(u^k)$, and with the additional convention that $u^{k+1}=u^k$ if $u^k\in U$. For any $\delta>0$, if $u^0\in P$ is close enough to $\bar{u}$, any such sequence is contained in $B(\bar{u},\,\delta)$ and converges to some $u^*\in U$, and the rate of convergence is superlinear with the Q-order $\min\{\theta+1,\,2\}$.

Sharpness of the requirement (5.4) on allowed inexactness, used in Theorem 5.2, is demonstrated by an example in [8, Section 5]. For instance, if $\theta=2$, quadratic convergence is preserved if inexactness is of order $O(\|\Phi(u)\|^3)$, while if $\theta=1$, the allowed order for inexactness preserving quadratic convergence is $O(\|\Phi(u)\|^2)$.

One general observation is in order. We consider here a parametric family of algorithms, parameterized by the exponent θ. The local convergence and rate of convergence result in Theorem 3.2 demonstrates, in particular, that for the range $[1,\,2]$ of values of this parameter, there is local convergence with the quadratic rate of convergence estimate. By itself, this does not yield any preferences for any particular value within this range: all these values do the job in the specified sense. Inexactness result in Theorem 5.2 is in favor of taking smaller values of θ though. Yet again, these are only the "worst case" estimates, of course.

For the unconstrained case, results on inexactness related to Theorem 5.2 can be found in [17, 28, 30, 37]. In particular, [17] establishes quadratic convergence for $\theta=2$ if inexactness is of order $O(\|\Phi(u)\|^4)$, while in [28] the same is proven for $\theta=1$ if inexactness is of order $O(\|\Phi(u)\|^3)$. Therefore, these works employ more

restrictive assumptions on the allowed level of inexactness than those coming from Theorem 5.2. At the same time, in [37], $\theta = 1$ and allowed inexactness is of order $O(\|\Phi(u)\|^2)$, thus agreeing with the claim of Theorem 5.2. Moreover, [30] recovers results from [37] and as already mentioned in Sect. 3, it further obtains superlinear convergence of the sequence $\{\mathrm{dist}(u^k, U)\}$ to 0 for $\theta \in (2, 3)$, in the inexact case as well.

We next discuss the relation between the two kinds of inexactness considered in Theorems 5.1 and 5.2. Since in Theorem 5.1 the values of γ stay bounded, inexactness allowed by (5.1) means that the corresponding value of the objective function of problem (1.3) must be of order $O(\|\Phi(u)\|^4)$. According to the analysis leading to Theorem 3.1, the optimal value of (1.3) (corresponding to the exact LM step) is of the same order $O(\|\Phi(u)\|^4)$. Therefore, that theorem allows to take any feasible direction keeping the order $O(\|\Phi(u)\|^4)$ for the value of the objective function of the constrained LM subproblem (1.3). Using this observation, it can be easily shown that inexactness allowed by Theorem 5.2 for $\theta = 2$ satisfies the requirements on inexactness in Theorem 5.1. (That said, we emphasize again that Theorem 5.2 allows to take smaller values of θ.)

The converse is not true, as demonstrated by the following simple example, which is Example 3.2 with $a = 1$, $b = 0$. Let $p = q = 1$, $P = \mathbb{R}$, $\Phi(u) = u$, and let $\sigma(u) = u^2$. Inexactness allowed by Theorem 5.1 is characterized by the relation

$$(u + v)^2 + u^2 v^2 = O(u^4)$$

as $u \to 0$. Obviously, any $v = -u + O(u^2)$ satisfies this requirement. At the same time, inexactness allowed by Theorem 5.2 is characterized by the equality

$$u + v + u^2 v = w$$

with some $w = O(|u|^3)$. Substituting v here by $-u + O(u^2)$ yields $w = -u^3 + O(u^2)$, and this quantity does not need to be $O(|u|^3)$ as $u \to 0$.

To conclude this discussion, we mention that a projected LM method with inexact projections but R-linear convergence was considered in [10], and with superlinear convergence under the very restrictive unconstrained error bound condition in [60].

Counterparts of Algorithm 5.1 and Theorem 5.1 for the LPN method can be found in [22].

6 Globalization of convergence

In this section, we assume that P is convex and that Φ is differentiable on P. Furthermore, let all the norms be Euclidean.

One natural approach to globalization of convergence of the constrained LM method is some backtracking linesearch for the merit function φ defined as the squared residual of the equation in (1.1), i.e., by (2.2) with $\nu = 2$. We start with the following algorithm that is fully hybrid in the sense of [45, Section 5.3]: it combines the constrained LM

method as a local phase with the projected gradient method as a global phase. Being considered with $\theta = 2$, this algorithm corresponds to [50, Algorithm 2.12].

Algorithm 6.1 Choose the parameters $\theta > 0$, $\rho \in (0, 1)$, $\widehat{\alpha} > 0$, $\varepsilon \in (0, 1)$, and $\kappa \in (0, 1)$. Choose $u^0 \in P$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Set $\sigma(u^k) = \|\Phi(u^k)\|^\theta$, and compute v^k as the solution of (1.3) with $u = u^k$, $G(u) = \Phi'(u^k)$.
3. If

$$\|\Phi(u^k + v^k)\| \leq \rho \|\Phi(u^k)\|, \tag{6.1}$$

   set $u^{k+1} = u^k + v^k$, increase k by 1, and go to Step 1.
4. Set $\alpha = \widehat{\alpha}$. If the inequality

$$\varphi(\pi_P(u^k - \alpha\varphi'(u^k))) \leq \varphi(u^k) + \varepsilon\langle\varphi'(u^k),\ \pi_P(u^k - \alpha\varphi'(u^k)) - u^k\rangle \tag{6.2}$$

   is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\kappa\alpha$, check the inequality (6.2) again, etc., until (6.2) is satisfied.
5. Set $u^{k+1} = \pi_P(u^k - \alpha_k\varphi'(u^k))$, increase k by 1 and go to Step 1.

Global convergence properties of Algorithm 6.1 are evidently inherited from those of the projected gradient method in [15, Proposition 2.3.3]. Step 3 aims at eventual switching to full-step LM iterations, thus accelerating convergence, at least from some point on. Convergence and rate of convergence properties of Algorithm 6.1 are given by the following theorem that corresponds to [50, Theorem 2.13] if $\theta = 2$ is used.

Theorem 6.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, and $P \subset \mathbb{R}^p$ be a nonempty closed convex set. Assume that Φ is continuously differentiable on P.*

Then Algorithm 6.1 uniquely defines the sequence $\{u^k\}$, and any accumulation point $\bar{u}$ of this sequence is a stationary point of the optimization problem (3.10), i.e., (3.11) holds with $u = \bar{u}$. Moreover, if Algorithm 6.1 is run with $\theta \in (0, 2]$, if an accumulation point $\bar{u}$ satisfies $\Phi(\bar{u}) = 0$, if the derivative of Φ is Lipschitz-continuous near $\bar{u}$, and if (3.2) holds as $u \in P$ tends to $\bar{u}$, then the entire sequence $\{u^k\}$ converges to $\bar{u}$, and the rate of convergence is superlinear with the Q-order $\min\{\theta + 1, 2\}$.

To show that the last assertion of the theorem above follows from Theorem 3.2, one only needs to demonstrate that the test (6.1) is always passed when u^k is close enough to $\bar{u}$. But this readily follows from the intermediate estimates in (3.16), from the Lipschitz-continuity of Φ near $\bar{u}$, and from the constrained error bound condition (3.2) as $u \in P$ tends to $\bar{u}$:

$$\|\Phi(u^k + v^k)\| = O(\operatorname{dist}(u^k + v^k, U)) = o(\operatorname{dist}(u^k, U)) = o(\|\Phi(u^k)\|) \leq \rho\|\Phi(u^k)\|$$

for $u^k \in P$ close enough to $\bar{u}$.

Close counterparts of Algorithm 6.1 and Theorem 6.1 can be found in [27], yet again involving the extraneous interiority assumption.

Hybrid globalization strategies like the one in Algorithm 6.1 are theoretically attractive because of their universality; see [45, Section 5.3]. For instance, [27, 50] also discuss hybrid globalization of the projected (rather than the constrained) LM method, similar to that in Algorithm 6.1. At the same time, practical features of hybrid globalizations are always questionable, in particular, because there are no general reasons to expect that most (or at least many) iterations of Algorithm 6.1 would be LM steps. Instead, it might happen that most would be the projected gradient steps, switching to the LM steps only at some very late iterations. To that end, we next present the "pure" linesearch globalization of the constrained LM method, developed in [36], and free from any hybrid ingredients.

Algorithm 6.2 Choose the parameters $\theta > 0$, $\varepsilon \in (0, 1)$, and $\kappa \in (0, 1)$. Choose $u^0 \in P$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Set $\sigma(u^k) = \|\Phi(u^k)\|^\theta$, and compute v^k as the solution of (1.3) with $u = u^k$, $G(u) = \Phi'(u^k)$.
3. Set $\alpha = 1$. If the inequality
$$\varphi(u^k + \alpha v^k) \le \varphi(u^k) - \varepsilon\sigma(u^k)\alpha\|v^k\|^2 \tag{6.3}$$
is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\kappa\alpha$, check the inequality (6.3) again, etc., until (6.3) is satisfied.
4. Set $u^{k+1} = u^k + \alpha_k v^k$, increase k by 1 and go to Step 1.

The following theorem corresponds to a combination of the results in [36, Theorems 5, 7], where Algorithm 6.2 is stated with $\theta = 2$, but as demonstrated below, this restriction can be avoided.

Theorem 6.2 *Theorem 6.1 remains true with Algorithm 6.1 therein substituted by Algorithm 6.2.*

For completeness, we now provide a full proof of Theorem 6.2, even though demonstrating that accumulation points of the generated sequence are stationary points for the optimization problem (3.10) essentially follows the lines of that in [36, Theorem 5].

Algorithm 6.2 may stop at some iteration k only if $u^k \in P$ satisfies $\Phi(u^k) = 0$, in which case u^k is necessarily a stationary point of the optimization problem (3.10).

On the other hand, if for a given $u^k \in P$ it holds that $\Phi(u^k) \ne 0$, then $\sigma(u^k) > 0$, the value of the objective function of the constrained LM subproblem (1.3) with $u = u^k \in P$ at its unique solution v^k is not larger than its value at 0 (since 0 is also a feasible point of this subproblem). Hence,

$$\begin{aligned}
\varphi(u^k) + \langle\varphi'(u^k),\ v^k\rangle & \\
+\|\Phi'(u^k)v^k\|^2 + \sigma(u^k)\|v^k\|^2 &= \|\Phi(u^k)\|^2 + 2\langle\Phi(u^k),\ \Phi'(u^k)v^k\rangle \\
&\quad +\|\Phi'(u^k)v^k\|^2 + \sigma(u^k)\|v^k\|^2 \\
&= \|\Phi(u^k) + \Phi'(u^k)v^k\|^2 + \sigma(u^k)\|v^k\|^2 \\
&\le \|\Phi(u^k)\|^2 \\
&= \varphi(u^k).
\end{aligned} \tag{6.4}$$

Therefore,

$$\langle \varphi'(u^k),\ v^k\rangle \le -\sigma(u^k)\|v^k\|^2, \tag{6.5}$$

and the definition of differentiability applied to φ at u^k then yields that (6.3) is satisfied by all $\alpha > 0$ small enough. This means that the loop at Step 3 of Algorithm 6.2 will be terminated after a finite number of backtracking steps, and the next iterate u^{k+1} will be successfully generated. Observe also that since both u^k and $u^k + v^k$ belong to P, P is convex, and α_k is taken from $(0,\ 1]$, it always holds that $u^{k+1} \in P$. This demonstrates that Algorithm 6.2 is well defined, and uniquely defines $\{u^k\}$.

Suppose now that the algorithm generates an infinite sequence $\{u^k\} \subset P$. For any accumulation point $\bar{u}$ of this sequence, consider an infinite subsequence $\{u^{k_i}\}$ convergent to $\bar{u}$. If we assume that $\bar{u}$ is not a stationary point of the optimization problem (3.10), i.e., (3.11) does not hold with $u = \bar{u}$, then there exists $\widetilde{u} \in P$ such that $\langle \varphi'(\bar{u}),\ \widetilde{u} - \bar{u}\rangle < 0$. Therefore, there are $\bar{\sigma} > 0$ and $\gamma > 0$ such that

$$\sigma(u^{k_i}) \ge \bar{\sigma}, \quad \langle \varphi'(u^{k_i}),\ \widetilde{v}^{k_i}\rangle \le -\gamma \tag{6.6}$$

for all i large enough, where $\sigma(u^{k_i})$ is defined at Step 2 of Algorithm 6.2, while $\widetilde{v}^{k_i} = \widetilde{u} - u^{k_i} \in P - u^{k_i}$. Then, for all $t \in [0,\ 1]$, it holds that $u^{k_i} + t\widetilde{v}^{k_i} \in P$, and hence, according to the definition of v^{k_i},

$$\begin{aligned}\|\Phi(u^{k_i}) + \Phi'(u^{k_i})v^{k_i}\|^2 + \sigma(u^{k_i})\|v^{k_i}\|^2 \le\ & \|\Phi(u^{k_i}) + t\Phi'(u^{k_i})\widetilde{v}^{k_i}\|^2 \\ & + \sigma(u^{k_i})\|t\widetilde{v}^{k_i}\|^2,\end{aligned}$$

or in other terms,

$$\langle \varphi'(u^{k_i}),\ v^{k_i}\rangle + \langle M_{k_i}v^{k_i},\ v^{k_i}\rangle \le t\langle \varphi'(u^{k_i}),\ \widetilde{v}^{k_i}\rangle + t^2\langle M_{k_i}\widetilde{v}^{k_i},\ \widetilde{v}^{k_i}\rangle,$$

where the matrices $M_{k_i} = (\Phi'(u^{k_i}))^\top\Phi'(u^{k_i}) + \sigma(u^{k_i})\mathcal{I}$ are positive semidefinite for all i. Then, employing the second inequality in (6.6), we derive

$$\langle \varphi'(u^{k_i}),\ v^{k_i}\rangle \le -\gamma t + t^2\langle M_{k_i}\widetilde{v}^{k_i},\ \widetilde{v}^{k_i}\rangle$$

for all $t \in [0,\ 1]$, and all i large enough. Taking into account that the sequences $\{M_{k_i}\}$ and $\{\widetilde{v}^{k_i}\}$ are bounded, it then follows that there exists $\bar{\gamma} > 0$ such that

$$\langle \varphi'(u^{k_i}),\ v^{k_i}\rangle \le -\bar{\gamma}$$

for all i large enough. Since the sequence $\{\varphi'(u^{k_i})\}$ is bounded, this also implies that the sequence $\{\|v^{k_i}\|\}$ is separated from zero.

According to (6.3), the sequence $\{\varphi(u^{k_i})\}$ is nonincreasing. Since this sequence is also bounded from below (by zero), it converges, and hence, again employing (6.3), the first inequality in (6.6), and the fact that $\{\|v^{k_i}\|\}$ is separated from zero, we conclude that $\alpha_{k_i} \to 0$ as $i \to \infty$. This means that for all i large enough, on iteration k_i, Step 3

of Algorithm 6.2 did not accept the first trial value of α, and hence it was reduced at least once, i.e., the value α_{k_i}/κ does not satisfy (6.3):

$$\frac{\varphi(u^{k_i}+\alpha_{k_i}v^{k_i}/\kappa)-\varphi(u^{k_i})}{\alpha_{k_i}/\kappa} > -\varepsilon\sigma(u^{k_i})\|v^{k_i}\|^2.$$

By the Mean-Value Theorem (e.g., [45, Theorem A.10(a)], this implies the existence of $t_{k_i} \in [0,\ 1]$ such that

$$\langle\varphi'(u^{k_i}+t_{k_i}\alpha_{k_i}v^{k_i}/\kappa),\ v^{k_i}\rangle > -\varepsilon\sigma(u^{k_i})\|v^{k_i}\|^2. \tag{6.7}$$

By the intermediate relations in (6.4) and the first inequality in (6.6), the sequence $\{v^{k_i}\}$ is bounded. Taking into account that $\{u^{k_i}\} \to \bar{u}$ and $\alpha_{k_i} \to 0$, and again employing the first inequality in (6.6) and the fact that $\{\|v^{k_i}\|\}$ is separated from zero, it then follows that

$$\begin{aligned}\frac{\langle\varphi'(u^{k_i}+t_{k_i}\alpha_{k_i}v^{k_i}/\kappa),\ v^{k_i}\rangle}{\sigma(u^{k_i})\|v^{k_i}\|^2} &= \frac{\langle\varphi'(u^{k_i}),\ v^{k_i}\rangle}{\sigma(u^{k_i})\|v^{k_i}\|^2}\\ &\quad+\frac{\langle\varphi'(u^{k_i}+t_{k_i}\alpha_{k_i}v^{k_i}/\kappa),\ v^{k_i}\rangle-\langle\varphi'(u^{k_i}),\ v^{k_i}\rangle}{\sigma(u^{k_i})\|v^{k_i}\|^2}\end{aligned}$$

with the second term in the right-hand side tending to zero as $i \to \infty$. Then by (6.5) we conclude that for any fixed $\widetilde{\varepsilon} \in (0,\ 1)$ it holds that, for all i large enough,

$$\frac{\langle\varphi'(u^{k_i}+t_{k_i}\alpha_{k_i}v^{k_i}/\kappa),\ v^{k_i}\rangle}{\sigma(u^{k_i})\|v^{k_i}\|^2} \le -\widetilde{\varepsilon},$$

and hence, according to (6.7),

$$-\widetilde{\varepsilon}\sigma(u^{k_i})\|v^{k_i}\|^2 > -\varepsilon\sigma(u^{k_i})\|v^{k_i}\|^2,$$

yielding a contradiction for $\widetilde{\varepsilon} > \varepsilon$.

The remaining part of the claim of Theorem 6.1 will follow from Theorem 3.2, if we show that $\alpha_k = 1$ is accepted at Step 3 of Algorithm 6.2 for $u^k \in P$ close enough to $\bar{u}$. But this again readily follows from the intermediate estimates in (3.16), from the Lipschitz-continuity of Φ near $\bar{u}$, and from the constrained error bound condition (3.2) as $u \in P$ tends to $\bar{u}$:

$$\begin{aligned}\varphi(u^k+v^k)-\varphi(u^k) &= \|\Phi(u^k+v^k)\|^2-\|\Phi(u^k)\|^2\\ &= -\|\Phi(u^k)\|^2+O((\mathrm{dist}(u^k+v^k,\ U))^2)\\ &= -\|\Phi(u^k)\|^2+o((\mathrm{dist}(u^k,\ U))^2)\\ &= -\|\Phi(u^k)\|^2+o(\|\Phi(u^k)\|^2)\\ &\le -\varepsilon\sigma(u^k)\|v^k\|^2\end{aligned}$$

for $u^k \in P$ close enough to $\bar{u}$, where the last inequality follows from the definition of $\sigma(u^k)$ at Step 2 of Algorithm 6.2, and from (3.7), as the latter, accompanied by (3.2), implies that for $\theta > 0$ it holds that

$$\sigma(u^k)\|v^k\|^2 = O(\|\Phi(u^k)\|^\theta(\mathrm{dist}(u^k,\ U))^2) = O(\|\Phi(u^k)\|^{2+\theta}) = o(\|\Phi(u^k)\|^2)$$

as $u^k \in P$ tends to $\bar{u}$.

Algorithms 6.1 and 6.2 can certainly be applied in the unconstrained case (when $P = \mathbb{R}^p$), and related globalizations of the unconstrained LM method were proposed in [31, 63, 65]. Instead of (6.3), these works employ the standard Armijo linesearch test

$$\varphi(u^k + \alpha v^k) \le \varphi(u^k) + \varepsilon\alpha\langle\varphi'(u^k),\ v^k\rangle.$$

According to (6.5), this condition is more (or at least no less) restrictive than (6.3). Apparently as a consequence of this, all those works [31, 63, 65] also check the linear decrease test (6.1) before the linesearch, in order to ensure asymptotic acceptance of the full LM step (as in the hybrid Algorithm 6.1).

For the unconstrained case, some hybrid globalizations of the inexact LM method were presented in [17, 28] and, more recently, in [4]. They also use a sufficient descent direction test (apparently because of inexactness), like (6.5) with $\sigma(u^k)$ replaced by some small positive constant, and switch to gradient descent when this test is not passed. In this sense, they are "as hybrid" as Algorithm 6.1. A pure linesearch globalization of the inexact LM method along the lines of [36] was recently discussed in [64].

In addition to linesearch globalizations, there also exist globalizations of the LM method in the spirit of trust-region algorithms. Probably the first work in this direction for unconstrained equations (with $p = q$) is [26], where the regularization parameter was taken in the form $\sigma(u) = \chi\|\Phi(u)\|$, with the scalar parameter $\chi > 0$ controlled by means of the trust-region technique. This approach was further developed in [2, 29, 57], with some more sophisticated rules to control the regularization parameter. Moreover, [29] employs $\sigma(u) = \chi\|\Phi(u)\|^\theta$, with superlinear convergence for $\theta \in (0,\ 1)$, and quadratic for $\theta \in [1,\ 2]$. Another recent related work is [13], where $\sigma(u) = \chi\|\Phi(u)\|^2$ is used. For the constrained case, some trust-region related techniques can be found in [53].

To conclude, we give some further comments on the rules that can be used for controlling the regularization parameters. A modification of the basic choice $\sigma(u) = \|\Phi(u)\|^\theta$, suggesting to take $\sigma(u) = \|(\Phi'(u))^\top\Phi(u)\|^\theta$, stems from the analysis in [32]: it is shown there that the growth behaviors of $\|\Phi(\cdot)\|$ and of $\|(\Phi'(\cdot))^\top\Phi(\cdot)\|$ are locally equivalent (are of the same order) if the Lipschitzian error bound holds and Φ is continuously differentiable. For globalizations, the choices $\sigma(u) = \min\{1,\ \|\Phi(u)\|^\theta\}$ and $\sigma(u) = \|\Phi(u)\|^\theta/(1 + \|\Phi(u)\|^\theta)$ become relevant, as they prevent the regularization parameter from becoming too large far from solutions. All of those choices can be with a multiplier $\chi > 0$ somehow varying within positive bounds, or even controlled in a special way in trust-region-like globalizations; see

the discussion above. A sophisticated control of the regularization parameter using estimates of the Lipschitz constant of the Jacobian is presented in [51].

As for the LPN method, a globalization strategy for it was developed in [33], employing linesearch for the merit function φ defined according to (2.2) with the ∞-norm, and with the exponent $\nu = 1$. Moreover, it was demonstrated in [33] that this globalization technique is applicable not only when Φ is smooth, but when it is only piecewise smooth as well. Some improvements of global convergence properties for the latter case were developed in [35]. We also mention the trust-region globalization in [5] and hybrid globalization in [20], the latter coupling the potential reduction algorithm from [21, 58] with the LPN method.

7 Open questions

Below, we state some open questions to further complete the picture given in this paper.

- It was shown in the final part of Sect. 3 that in the unconstrained case and for a smooth mapping Φ, the Lipschitzian error bound condition (3.2) follows from Assumption 2. It remains an open question whether this holds true in the constrained case, perhaps under Assumptions 1 and 3, or under appropriate smoothness.
- It would be interesting and important to achieve a full understanding of local convergence and rate of convergence properties for the exponent θ in the regularization parameter of the LM method taking values in $(2,\ 4)$.
- Can local convergence and rate of convergence properties be established in the piecewise smooth case with the exponent of the regularization parameter satisfying (3.9) (i.e., not only for $\theta = 2$)? In light of the presentation above, it seems plausible, because each step of the (constrained) piecewise LM method can be analyzed separately as in the smooth case.
- Can one obtain the same results concerning inexact solution of subproblems for the piecewise smooth case as for smooth mappings?
- The globalization of LM methods is well understood for the smooth case. Is it possible to make significant improvements for the globalization in the case of piecewise smooth problems?

Acknowledgements We thank the anonymous referees for a thorough reading of the original version and for helpful comments.

Funding The work of the first author was supported by the Volkswagen Foundation. The second author was funded by the Russian Science Foundation Grant 23-11-20020 (https://rscf.ru/en/project/23-11-20020/). The third author was supported in part by CNPq Grant 306775/2023-9, by FAPERJ Grant E-26/200.347/2023, and by PRONEX–Optimization.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest of any kind.

References

1. Ahookhosh, M., Fleming, R.M.T., Vuong, P.T.: Finding zeros of Hölder metrically subregular mappings via globally convergent Levenberg-Marquardt methods. Optim. Methods Softw. **37**, 113–149 (2022)
2. Amini, K., Rostami, F., Caristi, F.: An efficient Levenberg-Marquardt method with a new LM parameter for systems of nonlinear equations. Optimization **67**, 637–650 (2018)
3. Ariizumi, S., Yamakawa, Y., Yamashita, N.: Convergence properties of Levenberg-Marquardt methods with generalized regularization terms. Appl. Math. Comput. **463**, 128365 (2024)
4. Bao, J., Yu, C.K.W., Wang, J., Hu, Y., Yao, J.-C.: Modified inexact Levenberg-Marquardt methods for solving nonlinear least squares problems. Comput. Optim. Appl. **74**, 547–582 (2019)
5. Becher, L., Fernández, D., Ramos, A.: A trust-region LP-Newton method for constrained nonsmooth equations under Hölder metric subregularity. Comput. Optim. Appl. **86**, 711–743 (2023)
6. Behling, R.: The method and the trajectory of Levenberg–Marquardt. PhD thesis. IMPA - Instituto Nacional de Matemática Pura e Aplicada, Rio de Janeiro, Brazil (2011). https://impa.br/wp-content/uploads/2017/08/tese_dout_roger_behling.pdf
7. Behling, R., Iusem, A.: The effect of calmness on the solution set of systems of nonlinear equations. Math. Program. **137**, 155–165 (2013)
8. Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. Optim. Lett. **6**, 927–940 (2012)
9. Behling, R., Fischer, A., Haeser, G., Ramos, A., Schönefeld, K.: On the constrained error bound condition and the projected Levenberg-Marquardt method. Optimization **66**, 1397–1411 (2017)
10. Behling, R., Fischer, A., Herrich, M., Iusem, A., Ye, Y.: A Levenberg-Marquardt method with approximate projections. Comput. Optim. Appl. **59**, 5–26 (2014)
11. Behling, R., Fischer, A., Schönefeld, K., Strasdat, N.: A special complementarity function revisited. Optimization **68**, 65–79 (2019)
12. Behling, R., Gonçalves, D.S., Santos, S.A.: Local convergence analysis of the Levenberg-Marquardt framework for nonzero-residue nonlinear least-squares problems under an error bound condition. J. Optim. Theory Appl. **183**, 1099–1122 (2019)
13. Bergou, E.H., Diouane, Y., Kungurtsev, V.: Convergence and complexity analysis of a Levenberg-Marquardt algorithm for inverse problems. J. Optim. Theory Appl. **185**, 927–944 (2020)
14. Bergou, E.H., Diouane, Y., Kungurtsev, V., Royer, C.W.: A nonmonotone matrix-free algorithm for nonlinear equality-constrained least-squares problems. SIAM J. Sci. Comput. **43**, S743–S766 (2021)
15. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena, Belmont (1999)
16. Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: Numerical Optimization: Theoretical and Practical Aspects. Springer-Verlag, Berlin (2006)
17. 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)
18. DEGEN. http://w3.impa.br/~optim/DEGEN_collection.zip
19. Dennis, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall Inc, Englewood Cliffs (1983)
20. Dreves, A., Facchinei, F., Fischer, A., Herrich, M.: A new error bound result for Generalized Nash Equilibrium Problems and its algorithmic application. Comput. Optim. Appl. **59**, 63–84 (2014)
21. Dreves, A., Facchinei, F., Kanzow, C., Sagratella, S.: On the solution of the KKT conditions of generalized Nash equilibrium problems. SIAM J. Optim. **21**, 1082–1108 (2011)
22. Facchinei, F., Fischer, A., Herrich, M.: A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. Math. Methods Oper. Res. **77**, 433–443 (2013)
23. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**, 1–36 (2014)
24. Facchinei, F., Fischer, A., Piccialli, V.: Generalized Nash equilibrium problems and Newton methods. Math. Program. **117**, 163–194 (2009)
25. Facchinei, F., Kanzow, C.: A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems. Math. Program. **76**, 493–512 (1997)

26. Fan, J.-Y.: A modified Levenberg-Marquardt algorithm for singular systems of nonlinear equations. J. Comput. Math. **21**, 625–636 (2003)
27. Fan, J.-Y.: On the Levenberg-Marquardt methods for convex constrained nonlinear equations. J. Ind. Manag. Optim. **9**, 227–241 (2013)
28. Fan, J.-Y., Pan, J.-Y.: Inexact Levenberg-Marquardt method for nonlinear equations. Discrete Contin. Dyn. Syst. Ser. B. **4**, 1223–1232 (2004)
29. Fan, J.-Y., Pan, J.-Y.: Convergence properties of a self-adaptive Levenberg-Marquardt algorithm under local error bound condition. Comput. Optim. Appl. **34**, 47–62 (2006)
30. Fan, J.-Y., Pan, J.-Y.: On the convergence rate of the inexact Levenberg-Marquardt method. J. Ind. Manag. Optim. **7**, 199–210 (2011)
31. Fan, J.-Y., Yuan, Y.-X.: On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. Computing **74**, 23–39 (2005)
32. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
33. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: A globally convergent LP-Newton method. SIAM J. Optim. **26**, 2012–2033 (2016)
34. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. Comput. Optim. Appl. **63**, 425–459 (2016)
35. Fischer, A., Herrich, M., Izmailov, A.F., Scheck, W., Solodov, M.V.: A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points. Comput. Optim. Appl. **69**, 325–349 (2018)
36. Fischer, A., Shukla, P.K.: A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. **36**, 643–646 (2008)
37. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg-Marquardt methods. Optimization **59**, 273–287 (2010)
38. Fischer, A., Strasdat, N.: An extended convergence framework applied to complementarity systems with degenerate and nonisolated solutions. Pure Appl. Funct. Anal. **8**, 1039–1054 (2023)
39. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Naval Res. Logist. **3**, 95–110 (1956)
40. Golubitsky, M., Schaeffer, D.G.: Singularities and Groups in Bifurcation Theory, vol. I. Springer-Verlag, New York (1985)
41. Griewank, A.: Starlike domains of convergence for Newton's method at singularities. Numer. Math. **35**, 95–111 (1980)
42. Hager, W.W.: Lipschitz continuity for constrained processes. SIAM J. Control Optim. **17**, 321–338 (1979)
43. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: Critical solutions of nonlinear equations: local attraction for Newton-type methods. Math. Program. **167**, 355–379 (2018)
44. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: Critical solutions of nonlinear equations: stability issues. Math. Program. **168**, 475–507 (2018)
45. Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering. Springer, Cham (2014)
46. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: A globally convergent Levenberg-Marquardt method for equality-constrained optimization. Comput. Optim. Appl. **72**, 215–239 (2019)
47. Jolaoso, L.O., Mehlitz, P., Zemkoho, A.B.: A fresh look at nonsmooth Levenberg-Marquardt methods with applications to bilevel optimization. Optimization (2024). https://doi.org/10.1080/02331934.2024.2313688
48. Kanzow, C., Petra, S.: On a semismooth least squares formulation of complementarity problems with gap reduction. Optim. Methods Softw. **19**, 507–525 (2004)
49. 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)
50. 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)
51. Karas, E.W., Santos, S.A., Svaiter, B.F.: Algebraic rules for computing the regularization parameter of the Levenberg-Marquardt method. Comput. Optim. Appl. **65**, 723–751 (2016)
52. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Quart. Appl. Math. **2**, 164–168 (1944)

53. 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)
54. Martínez, M.A., Fernández, D.: A quasi-Newton modified LP-Newton method. Optim. Methods Softw. **34**, 634–649 (2019)
55. Martínez, M.A., Fernández, D.: On the local and superlinear convergence of a secant modified linear-programming-Newton method. J. Optim. Theory Appl. **180**, 993–1010 (2019)
56. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. **11**, 431–441 (1963)
57. Marumo, N., Okuno, T., Takeda, A.: Majorization-minimization-based Levenberg-Marquardt method for constrained nonlinear least squares. Comput. Optim. Appl. **84**, 833–874 (2023)
58. Monteiro, R.D.C., Pang, J.-S.: A potential reduction Newton method for constrained equations. SIAM J. Optim. **9**, 729–754 (1999)
59. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer-Verlag, New York (2006)
60. de Oliveira, F.R., de Oliveira, F.R.: A locally convergent inexact projected Levenberg-Marquardt-type algorithm for large-scale constrained nonsmooth equations. J. Comput. Appl. Math. **427**, 115–121 (2023)
61. Riccietti, E.: Levenberg–Marquardt methods for the solution of noisy nonlinear least squares problems. Ph.D. thesis. University of Florence, Italy (2017)
62. Tin, A., Zemkoho, A.B.: Levenberg-Marquardt method and partial exact penalty parameter selection in bilevel optimization. Optim. Eng. **24**, 1343–1385 (2023)
63. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg-Marquardt method. In: Alefeld, G., Chen, X. (eds.) Topics in Numerical Analysis. Computing Supplementa. Springer, Vienna (2001)
64. Yin, J., Jian, J., Ma, G.: A modified inexact Levenberg-Marquardt method with the descent property for solving nonlinear equations. Comput. Optim. Appl. **87**, 289–322 (2024)
65. Zhang, J.-L.: On the convergence properties of the Levenberg-Marquardt method. Optimization **52**, 739–756 (2003)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.