Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-023-02350-w

Behavior of Newton-Type Methods Near Critical Solutions of Nonlinear Equations with Semismooth Derivatives

Andreas Fischer[1] · **Alexey F. Izmailov**[2,3] · **Mario Jelitte**[1]

Received: 10 May 2023 / Accepted: 20 November 2023

Abstract

Having in mind singular solutions of smooth reformulations of complementarity problems, arising unavoidably when the solution in question violates strict complementarity, we study the behavior of Newton-type methods near singular solutions of nonlinear equations, assuming that the operator of the equation possesses a strongly semismooth derivative, but is not necessarily twice differentiable. These smoothness restrictions give rise to peculiarities of the analysis and results on local linear convergence and asymptotic acceptance of the full step, the issues addressed in this work. Moreover, we consider not only the basic Newton method, but also some stabilized versions of it intended for tackling singular (including nonisolated) solutions. Applications to nonlinear complementarity problems are also dealt with.

Keywords Nonlinear equation · Constrained equation · Strongly semismooth derivative · Singular solution · Critical solution · 2-Regularity · Perturbed Newton method · Acceptance of the full step · Extrapolation · Nonlinear complementarity problem

Mathematics Subject Classification 49J52 · 65J15 · 65K15 · 90C33

Dedicated to the memory of Professor Boris Polyak.

Communicated by Ebrahim Sarabi.

✉ Alexey F. Izmailov
izmaf@cs.msu.ru

Andreas Fischer
andreas.fischer@tu-dresden.de

Mario Jelitte
mario.jelitte@tu-dresden.de

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

2 VMK Faculty, OR Department, Lomonosov Moscow State University, Moscow, Russia

3 Derzhavin Tambov State University, Tambov, Russia

Published online: 18 December 2023

1 Introduction

There exists a relatively rich literature on the behavior of the Newton method near singular solution of smooth nonlinear equations. With no intention to give a comprehensive survey, we mention only the works [25–27] most closely related to our development below, but dealing with equations as smooth as needed (smoothness is not an issue), and with the basic Newton method. In this setting, these works provide natural conditions ensuring linear local convergence of the Newton method from an asymptotically dense starlike domain around a singular solution, and also provide some acceleration techniques bases on the established convergence pattern. Partial extensions of these convergence results to wide classes of methods that can be interpreted as a perturbed Newton method were developed in [30]. Acceleration of convergence and a related issue of asymptotic acceptance of the full Newton step by a linesearch globalization procedure were further investigated in [21, 22], while [20] contains some extensions of these results to the case of constrained equations.

Having in mind typical equation reformulations of complementarity problems, an important issue consists of possible extensions of the results mentioned above to equations with restricted smoothness properties. As one example of this kind, the case of piecewise smooth equations was addressed in [19]. Different reformulations of complementarity lead to equations with different smoothness and regularity properties, and as a result, to different methods for solving complementarity problems, and understanding the relative advantages and disadvantages of these methods is of much interest and importance.

In this work, we focus on nonlinear equations with operators differentiable near the solution in question, and with their derivatives being strongly semismooth at this solution, but when the second derivatives of the operator may not exist. The concept of strong semismoothness was introduced in [44]; see, e.g., [35, Sect. 1.4] for a recent exposition of the related theory. Local convergence properties of the basic Newton method and some acceleration techniques were studied under similar smoothness assumptions in [42]. The main difference between the results in [42] and our development below is that we deal not only with the basic Newton method, but with its perturbed version covering, in particular, some stabilized modifications of the basic Newton scheme, specially intended for tackling singular (and even nonisolated) solutions. Moreover, we consider not only the local convergence properties, but also the issue of the asymptotic acceptance of the unit stepsize by the algorithms equipped with linesearch for globalization of convergence. The latter line of analysis leads to a new result for the perturbed Newton method, even in the case of arbitrary smoothness.

As it will be discussed below, reformulations of complementarity problems, possessing the specified smoothness properties, necessarily give rise to singularity of solutions violating strict complementarity, and hence, serve as a natural source of applications, both in [42] and below.

The rest of the paper is structured as follows. In Sect. 2, we provide the needed preliminaries, and specify the problem setting. Section 3 contains the main result on linear local convergence of the perturbed Newton method framework to singular solution satisfying a certain 2-regularity property that may only hold at solutions called critical. In Sect. 4, we consider a linesearch globalization procedure for the methods

in question, and investigate the issue of asymptotic acceptance of the full step, playing a key role for the potential success of the extrapolation procedure intended for acceleration of convergence to critical solutions. Finally, Sect. 5 contains examples of application of the results obtained to smooth equation reformulations of nonlinear complementarity problems.

Some words about our notation. For any $\bar{u},\ \bar{v} \in \mathbb{R}^p$, and any given scalars $\varepsilon > 0$ and $\delta > 0$, define the set

$$K_{\varepsilon,\delta}(\bar{u};\ \bar{v}) := \{u \in \mathbb{R}^p \mid \|u - \bar{u}\| \le \varepsilon,\ \|\|\bar{v}\|(u - \bar{u}) - \|u - \bar{u}\|\bar{v}\| \le \delta\|u - \bar{u}\|\|\bar{v}\|\}.$$

For a $q \times p$ matrix A, the null space of the corresponding linear operator is $\ker A := \{v \in \mathbb{R}^p \mid Av = 0\}$. For a mapping $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ differentiable at $\bar{u}$, we will make use of the unique decomposition of every $u \in \mathbb{R}^p$ into the sum $u = u_1 + u_2$ with $u_1 \in (\ker \Phi'(\bar{u}))^\perp$ and $u_2 \in \ker \Phi'(\bar{u})$, where $^\perp$ stands for the orthogonal complement of a linear subspace.

2 Preliminaries and Problem Setting

Consider a mapping $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ that is differentiable near a point $\bar{u} \in \mathbb{R}^p$, but not necessarily twice differentiable, even at $\bar{u}$. The analysis in this paper will rely on the assumption that the derivative $\Phi' : \mathbb{R}^p \to \mathbb{R}^{q\times p}$ is strongly semismooth at $\bar{u}$. According to [35, Sect. 1.4.2], this requirement means that Φ' is Lipschitz-continuous near $\bar{u}$, directionally differentiable at $\bar{u}$ in every direction, and the estimate

$$\max_{J\in\partial\Phi'(u)} \|\Phi'(u) - \Phi'(\bar{u}) - J(u - \bar{u})\| = O(\|u - \bar{u}\|^2) \tag{1}$$

holds as $u \in \mathbb{R}^p$ tends to $\bar{u}$. Here, $\partial\Phi'(u)$ stands for Clarke's generalized Jacobian of Φ' at u [6, Definition 2.6.1]. These "smoothness" assumptions can actually be further relaxed: it would be enough to assume that Φ' itself is just calm at $\bar{u}$, while $\Pi\Phi'$ is strongly semismooth at $\bar{u}$, with Π being the orthogonal projector onto $(\mathrm{im}\,\Phi'(\bar{u}))^\perp$ in $\mathbb{R}^q$. We do not pursue this further, in order to keep the presentation reasonably simple.

Let $(\Phi')'(\bar{u};\ v)$ stand for the directional derivative of Φ' at $\bar{u}$ in a direction $v \in \mathbb{R}^p$. Observe that $(\Phi')'(\bar{u};\ \cdot)$ maps $\mathbb{R}^p$ to $\mathbb{R}^{q\times p}$, and is positively homogeneous and Lipschitz-continuous. Define

$$r(u) := \Phi(u) - \Phi(\bar{u}) - \Phi'(\bar{u})(u - \bar{u}) - \frac{1}{2}(\Phi')'(\bar{u};\ u - \bar{u})(u - \bar{u}), \tag{2}$$

$$R(u) := \Phi'(u) - \Phi'(\bar{u}) - (\Phi')'(\bar{u};\ u - \bar{u}). \tag{3}$$

Combining (1) with [35, Proposition 1.71 (c)], from (3), we readily obtain the estimate

$$R(u) = O(\|u - \bar{u}\|^2) \tag{4}$$

as $u \to \bar{u}$. Furthermore, according to (2), by the Newton–Leibniz formula we derive

$$\begin{aligned} r(u) &= \int_0^1 (\Phi'(\tau u + (1-\tau)\bar{u}) - \Phi'(\bar{u}))(u-\bar{u})\,d\tau - \frac{1}{2}(\Phi')'(\bar{u};\ u-\bar{u})(u-\bar{u}) \\ &= \int_0^1 (\Phi'(\bar{u} + \tau(u-\bar{u})) - \Phi'(\bar{u}) - (\Phi')'(\bar{u};\ \tau(u-\bar{u}))(u-\bar{u})\,d\tau \\ &= \int_0^1 R(\bar{u} + \tau(u-\bar{u}))(u-\bar{u})\,d\tau \\ &= O(\|u-\bar{u}\|^3), \end{aligned} \tag{5}$$

where the second equality employs the fact that $(\Phi')'(\bar{u};\ \cdot)$ is positively homogeneous, the third is by (3), while the last one is by (4).

The mapping Φ is said to be 2-regular at $\bar{u}$ in the direction v if the linear operator $B(v) : \ker \Phi'(\bar{u}) \to (\operatorname{im} \Phi'(\bar{u}))^{\perp}$ defined as the restriction of $\Pi(\Phi')'(\bar{u};\ v)$ to $\ker \Phi'(\bar{u})$ is surjective; see the corresponding definitions and their discussion in [32, 33, 42].

Remark 2.1 At this point, we mention that the structure of the set consisting of directions of 2-regularity of Φ at $\bar{u}$ is not arbitrary. For instance, if Φ is twice differentiable at $\bar{u}$, it can only be 2-regular at $\bar{u}$ in every nonzero direction if either $\operatorname{rank} \Phi'(\bar{u}) = q$, or $\Phi'(\bar{u}) = 0$.

Indeed, assuming that $\operatorname{rank} \Phi'(\bar{u}) < q$, fix any $w \in (\operatorname{im} \Phi'(\bar{u}))^{\perp}$, and consider the $p \times p$ matrix $w\Phi''(\bar{u}) := \sum_{i=1}^{q} w_i \Phi_i''(\bar{u})$. Then for any $\widehat{v} \in \mathbb{R}^p$ and $v \in \ker \Phi'(\bar{u})$ it holds that

$$\begin{aligned} \langle w,\ B(\widehat{v})v\rangle &= \langle w,\ \Pi\Phi''(\bar{u})[\widehat{v},\ v]\rangle \\ &= \langle \Pi w,\ \Phi''(\bar{u})[\widehat{v},\ v]\rangle \\ &= \langle w,\ \Phi''(\bar{u})[\widehat{v},\ v]\rangle \\ &= \sum_{i=1}^{q} w_i \langle \Phi_i''(\bar{u})\widehat{v},\ v\rangle \\ &= \left\langle \sum_{i=1}^{q} w_i \Phi_i''(\bar{u})\widehat{v},\ v \right\rangle \\ &= \langle w\Phi''(\bar{u})\widehat{v},\ v\rangle, \end{aligned} \tag{6}$$

where the second equality is due to the symmetry of Π, while the third is because Π acts as the identity on $(\operatorname{im} \Phi'(\bar{u}))^{\perp}$. If the matrix $w\Phi''(\bar{u})$ is singular, then there exists $\widehat{v} \in \ker w\Phi''(\bar{u}) \setminus \{0\}$, and substituting it into (6), we conclude that $w \in (\operatorname{im} B(\widehat{v}))^{\perp}$. On the other hand, if $w\Phi''(\bar{u})$ is nonsingular, and $\Phi'(\bar{u}) \neq 0$, then there exists $\widehat{v} \in \mathbb{R}^p$ such that $w\Phi''(\bar{u})\widehat{v} \in (\ker \Phi'(\bar{u}))^{\perp} \setminus \{0\}$, implying, in particular, that $\widehat{v} \neq 0$, and again by (6) we have that $w \in (\operatorname{im} B(\widehat{v}))^{\perp}$.

Therefore, if $\operatorname{rank} \Phi'(\bar{u}) < q$ and $\Phi'(\bar{u}) \neq 0$, then for any $w \in (\operatorname{im} \Phi'(\bar{u}))^{\perp}$ we have the existence of a nonzero $\widehat{v} \in \mathbb{R}^p$ such that $w \in (\operatorname{im} B(\widehat{v}))^{\perp}$. In particular, if we take $w \neq 0$, this implies that Φ is not 2-regular at $\bar{u}$ in the direction $\widehat{v}$.

The case when $\Phi'(\bar{u}) = 0$ is of course quite a special instance of singularity on its own. Moreover, even in this case, from the considerations above it follows that 2-regularity of Φ in any nonzero direction is only possible if there exists no nonzero $w \in \mathbb{R}^q$ such that the matrix $w\Phi''(\bar{u})$ is singular. But the latter property imposes further restrictions on the dimensions p and q. See, e.g., [1, Theorem 1], implying in particular, that this is not possible when p is odd and $q \geq 2$. A related observation can be found in [3].

In the rest of the paper, we deal with Newton-type methods for the equation

$$\Phi(u) = 0, \tag{7}$$

and to that end, we assume that $p = q$. In this case, $\bar{u}$ is called a singular solution of (7) if $\Phi'(\bar{u})$ is a singular matrix. Observe that every nonisolated solution is necessarily singular. Observe further that if $\bar{u}$ is nonsingular, Φ is 2-regular at $\bar{u}$ in every direction v, including $v = 0$. At the same time, Φ may be 2-regular at $\bar{u}$ in nonzero directions even when $\bar{u}$ is singular, and even when $\bar{u}$ is a nonisolated solution of (7), and even in directions $\bar{v} \in \ker \Phi'(\bar{u})$, which is specially important here as this will play a crucial role in our analysis below, and leads to the

Key assumption: there exist $\bar{v} \in \ker \Phi'(\bar{u})$ such that the mapping Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$.

According to Izmailov et al. [31, Theorem 2], a solution $\bar{u}$ of (7) is regarded as critical if and only if it violates the local Lipschitzian error bound property

$$\operatorname{dist}(u,\ \Phi^{-1}(0)) = O(\|\Phi(u)\|) \tag{8}$$

as $u \in \mathbb{R}^p$ tends to $\bar{u}$. The property in (8) is related to the concept of (weak) sharp minima (see [43, Sect. 5.2.3], and [5]) for the residual function $\|\Phi(\cdot)\|$. By Izmailov et al. [31, Theorem 3], every critical solution is necessarily singular, but generally not the other way round. Moreover, the discussion in [31, p. 497] demonstrates that for a singular (e.g., nonisolated) solution $\bar{u}$, our key assumption may only hold if $\bar{u}$ is a critical solution.

3 Local Convergence of Perturbed Newton Methods to Critical Solutions

As in [30], define the perturbed Newton method (pNM) framework for Equation (7) as follows. For a given iterate $u^k \in \mathbb{R}^p$, the next iterate is $u^{k+1} = u^k + v^k$, where v^k is a solution of the linear equation

$$\Phi(u^k) + (\Phi'(u^k) + \Omega(u^k))v = \omega(u^k), \tag{9}$$

where the mappings $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ are the terms characterizing various kinds of perturbation, and defining specific methods within the pNM framework.

The following is a generalization of Izmailov et al. [30, Lemma 1] and Izmailov et al. [22, Lemma 1] to the case when the first derivative is strongly semismooth, but the second derivative may not exist.

Lemma 3.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be differentiable near $\bar{u} \in \mathbb{R}^p$, and let the derivative of Φ be strongly semismooth at $\bar{u}$. Let $\bar{u}$ be a solution of Equation (7), and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \mathbb{R}^p$. Let $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ satisfy the following properties: there exists $\delta > 0$ such that*

$$\Omega(u) = O(\|u - \bar{u}\|), \quad \omega(u) = O(\|u - \bar{u}\|^2) \tag{10}$$

for $u \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ as $\varepsilon \to 0+$, and

$$\Pi\Omega(u) = o(\|u - \bar{u}\|) \tag{11}$$

for $u \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ as $\varepsilon \to 0+$ and $\delta \to 0+$.

Then there exist $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ such that, for every $u \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u};\ \bar{v})\backslash\{\bar{u}\}$, the linear operator $B(u - \bar{u})$ is invertible,

$$(B(u - \bar{u}))^{-1} = O(\|u - \bar{u}\|^{-1}) \tag{12}$$

as $u \to \bar{u}$, Equation (9) with $u^k = u$ has the unique solution v, and this solution satisfies

$$\begin{aligned} u_1 + v_1 - \bar{u}_1 &= O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|\|\Omega(u)\|) \\ &\quad + O(\|\omega(u)\|) + O(\|u - \bar{u}\|^3), \end{aligned} \tag{13}$$

$$\begin{aligned} u_2 + v_2 - \bar{u}_2 &= \frac{1}{2}(u_2 - \bar{u}_2 + (B(u - \bar{u}))^{-1}\Pi(\Phi')'(\bar{u};\ u - \bar{u})(u_1 - \bar{u}_1)) \\ &\quad + O(\|\Pi\Omega(u)\|) + O(\|u - \bar{u}\|^{-1}\|\Pi\omega(u)\|) \\ &\quad + O(\|u - \bar{u}\|^2) \end{aligned} \tag{14}$$

as $u \to \bar{u}$.

Proof The argument below follows the lines of that in [30, Lemma 1], with modifications needed under the current restricted smoothness assumptions. Without loss of generality assume that $\bar{u} = 0$.

Multiplying (9) by $(I - \Pi)$ and by Π, and employing (2)–(3), Equation (9) with $u^k = u \in \mathbb{R}^p$ is decomposed into the following two equations:

$$\begin{aligned} &(\Phi'(\bar{u}) + (I - \Pi)((\Phi')'(\bar{u};\ u) + R(u) + \Omega(u)))v_1 \\ &\quad = -\Phi'(\bar{u})u_1 - (I - \Pi)\left(\frac{1}{2}(\Phi')'(\bar{u};\ u)u + r(u) - \omega(u)\right) \\ &\qquad - (I - \Pi)((\Phi')'(\bar{u};\ u) + R(u) + \Omega(u)))v_2 \end{aligned} \tag{15}$$

and

$$\Pi((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u))(v_1+v_2)=-\Pi\left(\frac{1}{2}(\Phi')'(\bar{u};\,u)u+r(u)-\omega(u)\right). \quad (16)$$

Let $\bar{\varepsilon}>0$ and $\bar{\delta}>0$ be fixed arbitrarily for now, and from this point on, we consider only those $u\in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u};\,\bar{v})\backslash\{0\}$. Define the linear operator $\mathcal{A}(u):(\ker\Phi'(\bar{u}))^{\perp}\to\operatorname{im}\Phi'(\bar{u})$ as the restriction of $(\Phi'(\bar{u})+(I-\Pi)((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u)))$ to $(\ker\Phi'(\bar{u}))^{\perp}$. Furthermore, let $\widehat{A}:(\ker\Phi'(\bar{u}))^{\perp}\to\operatorname{im}\Phi'(\bar{u})$ be the restriction of $\Phi'(\bar{u})$ to $(\ker\Phi'(\bar{u}))^{\perp}$. Then, taking into account (5), the equality (15) can be written as

$$\begin{aligned}\mathcal{A}(u)v_1=&-\widehat{A}u_1-(I-\Pi)((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u))v_2\\&-(I-\Pi)\left(\frac{1}{2}(\Phi')'(\bar{u};\,u)u-\omega(u)\right)+O(\|u\|^3)\end{aligned} \quad (17)$$

as $u\to 0$.

Evidently, $\widehat{A}$ is invertible, and according to (4) and the first condition in (10),

$$\mathcal{A}(u)=\widehat{A}+O(\|u\|).$$

This implies that if $\bar{\varepsilon}>0$ is small enough, then $\mathcal{A}(u)$ is invertible, and

$$(\mathcal{A}(u))^{-1}=\widehat{A}^{-1}+O(\|u\|) \quad (18)$$

as $u\to 0$; this follows, e.g., from Izmailov and Solodov [35, Lemma A.6]. Therefore, taking also into account the second condition in (10), (17) can be written as

$$v_1=-u_1+\mathcal{M}(u)v_2+O(\|u\|^2), \quad (19)$$

where $\mathcal{M}(u):\ker\Phi'(\bar{u})\to(\ker\Phi'(\bar{u}))^{\perp}$ is defined by

$$\mathcal{M}(u):=-(\mathcal{A}(u))^{-1}(I-\Pi)((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u))=O(\|u\|) \quad (20)$$

as $u\to 0$, where the last estimate is again by (4) and by the first condition in (10).

Substituting (19) into (16), and taking into account (4), we obtain the equation

$$\begin{aligned}&\Pi((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u))(I+\mathcal{M}(u))v_2\\&\quad=-\Pi\left(\frac{1}{2}(\Phi')'(\bar{u};\,u)u-\omega(u)\right)\\&\qquad+\Pi((\Phi')'(\bar{u};\,u)+\Omega(u))u_1+O(\|u\|^3).\end{aligned} \quad (21)$$

Define the linear operator $\mathcal{B}(u):\ker\Phi'(\bar{u})\to(\operatorname{im}\Phi'(\bar{u}))^{\perp}$ as the restriction of $\Pi((\Phi')'(\bar{u};\,u)+R(u)+\Omega(u))(I+\mathcal{M}(u))$ to $\ker\Phi'(\bar{u})$. Then (21) can be written in the form

$$\mathcal{B}(u)v_2=-\frac{1}{2}B(u)u_2+\Pi\left(\left(\frac{1}{2}(\Phi')'(\bar{u};\,u)+\Omega(u)\right)u_1+\omega(u)\right)+O(\|u\|^3) \quad (22)$$

as $u \to 0$.

Observe now that by Izmailov and Solodov [35, Lemma A.6], and by continuity of $(\Phi')'(\bar{u};\ \cdot)$ at $\bar{v}$, 2-regularity of Φ at 0 in the direction $\bar{v}$ implies the existence of $C > 0$ such that $B(u)$ is invertible and

$$\|(B(u))^{-1}\| \le C\|u\|^{-1} \tag{23}$$

provided $\bar{\delta} > 0$ is taken small enough. This yields (12). According to (4), (10), and (20), it further holds that

$$\mathcal{B}(u) = B(u) + \Pi\Omega(u) + O(\|u\|^2).$$

Further reducing $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ if necessary, by (11) and (23), and again by Izmailov and Solodov [35, Lemma A.6], we now obtain that $\mathcal{B}(u)$ is invertible, and

$$(\mathcal{B}(u))^{-1} = (B(u))^{-1} + O(\|u\|^{-2}\|\Pi\Omega(u)\|) + O(1) = O(\|u\|^{-1})$$

as $u \to 0$. Therefore, (22) is uniquely solvable, and its unique solution has the form

$$\begin{aligned} v_2 &= -\frac{1}{2}u_2 + \frac{1}{2}(B(u))^{-1}\Pi(\Phi')'(\bar{u};\ u)u_1 + O(\|\Pi\Omega(u)\|) \qquad (24)\\ &\quad + O(\|u\|^{-1}\|\Pi\omega(u)\|) + O(\|u\|^2)\\ &= O(\|u\|), \qquad (25)\end{aligned}$$

as $u \to 0$, where the last estimate is by (10) and (23).

Substituting (24) into (17), and employing (4) again, we obtain the equation

$$\mathcal{A}(u)v_1 = -\widehat{A}u_1 + O(\|u\|\|u_1\|) + O(\|u\|\|\Omega(u)\|) + O(\|\omega(u)\|) + O(\|u\|^3)$$

and hence, by (18),

$$v_1 = -u_1 + O(\|u\|\|u_1\|) + O(\|u\|\|\Omega(u)\|) + O(\|\omega(u)\|) + O(\|u\|^3) \tag{26}$$

as $u \to 0$.

From (24) and (26), and from (10), we have the needed estimates (13) and (14). □

The next example demonstrates that even in the case of twice continuous differentiability of Φ, and even in the absence of perturbations, *strong* semismoothness of Φ' is essential for the conclusion of Lemma 3.1 to be valid.

Example 3.1 Let $p = 2$, $\Phi(u) = (u_1 + 3u_2^{7/3}/7,\ u_2^2/2)$. Then Φ is everywhere twice continuously differentiable, and the unique solution of (7) is $\bar{u} = 0$. Furthermore, for any $u,\ v \in \mathbb{R}^p$

$$\Phi'(u) = \begin{pmatrix} 1 & u_2^{4/3} \\ 0 & u_2 \end{pmatrix}, \quad \Phi'(0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \Phi''(0)[v] = \begin{pmatrix} 0 & 0 \\ 0 & v_2 \end{pmatrix}.$$

Therefore, Φ is 2-regular at 0 in every nonzero direction in $\ker \Phi'(0) = \{0\} \times \mathbb{R}$.

Assuming that $u_2 \neq 0$, the basic Newton step from $u^k = u$ (i.e., the unique solution of (9) with $\Omega \equiv 0$ and $\omega \equiv 0$) is $v = (-u_1 + u_2^{7/3}/14,\ -u_2/2)$. In particular, (14) is valid, while (13) (and even a weaker estimate from Izmailov et al. [30, Lemma 1]) is not. The reason is violation of (1).

Theorem 3.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be differentiable near $\bar{u} \in \mathbb{R}^p$, and let the derivative of Φ be strongly semismooth at $\bar{u}$. Let $\bar{u}$ be a solution of Equation (7), and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \ker \Phi'(\bar{u}) \backslash \{0\}$. Moreover, let $\Omega : \mathbb{R}^p \to \mathbb{R}^{p \times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ satisfy the following properties: there exists $\delta > 0$ such that, along with (10), the estimates*

$$\Pi\Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \tag{27}$$

and

$$\Pi\omega(u) = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3) \tag{28}$$

hold for $u \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ as $\varepsilon \to 0+$.

Then, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ there exists the unique sequence $\{u^k\} \subset \mathbb{R}^p$ such that for each k it holds that $u^{k+1} = u^k + v^k$, where v^k satisfies (9), and for this sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u};\ \bar{v})$, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically,

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_2^{k+1} - \bar{u}_2\|} = O(\|u^k - \bar{u}\|) \tag{29}$$

as $k \to \infty$, and

$$\lim_{k\to\infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|} = \frac{1}{2}. \tag{30}$$

Proof Under the assumption (10), estimates (12)–(14) in Lemma 3.1 further imply that

$$u_1 + v_1 - \bar{u}_1 = O(\|u - \bar{u}\|^2), \tag{31}$$

$$\begin{aligned} u_2 + v_2 - \bar{u}_2 = {} & \frac{1}{2}(u_2 - \bar{u}_2) + O(\|u_1 - \bar{u}_1\|) \\ & + O(\|\Pi\Omega(u)\|) + O(\|u - \bar{u}\|^{-1}\|\Pi\omega(u)\|) \\ & + O(\|u - \bar{u}\|^2) \end{aligned} \tag{32}$$

as $u \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u};\ \bar{v}) \backslash \{\bar{u}\}$ tends to $\bar{u}$, where $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are defined according to Lemma 3.1.

Assuming further that there exists $\delta > 0$ such that (27), (28) hold for $u \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ as $\varepsilon \to 0+$, the estimate (32) is further simplified to

$$u_2 + v_2 - \bar{u}_2 = \frac{1}{2}(u_2 - \bar{u}_2) + O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \tag{33}$$

as $u \to \bar{u}$, and the subsequent analysis in the proof of Izmailov et al. [30, Theorem 1] goes through, as it does not further rely on any smoothness assumptions but only on the estimates (31) and (33). This yields the needed result. □

Remark 3.1 The flexibility of the assumption on perturbation terms $\Omega(\cdot)$ and $\omega(\cdot)$ allows for applications of Theorem 3.1 to various specific Newton-type methods, including those equipped with stabilizing features intended specially for finding singular and even nonisolated solutions. To begin with, taking $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$ recovers the classical Newton method for Equation (7), with the subproblem

$$\Phi(u^k) + \Phi'(u^k)v = 0. \tag{34}$$

Furthermore, consider the Levenberg–Marquardt method [38, 39] (see also [41, Sect. 10.3]) with the subproblem of the form

$$\text{minimize } \frac{1}{2}\|\Phi(u^k) + \Phi'(u^k)v\|^2 + \frac{1}{2}\rho(u^k)\|v\|^2, \quad v \in \mathbb{R}^p, \tag{35}$$

where $\rho : \mathbb{R}^p \to \mathbb{R}_+$ defines the regularization parameter. For modern local quadratic convergence theories for this method under the local Lipschitzian error bound condition (8) (i.e., noncriticality of the solution in question), and including the associated rules to control the regularization parameter, see [10, 13, 14, 16, 17, 24, 46].

Passing to the case of a critical solution, observe that the subproblem (35) employing the Euclidean norm is equivalent to the linear equation

$$(\Phi'(u^k))^\top \Phi(u^k) + ((\Phi'(u^k))^\top \Phi'(u^k) + \rho(u^k)I)v = 0, \tag{36}$$

and the constructions in [30, Sect. 3.1] allow to interpret this equation as the subproblem (9) with $\Omega(\cdot)$ and $\omega(\cdot)$ possessing the needed properties when $\rho(\cdot) := \|\Phi(\cdot)\|^\tau$, with $\tau \geq 2$. This yields a counterpart of Izmailov et al. [30, Corollary 1], saying essentially that under the smoothness and 2-regularity assumptions in Theorem 3.1, the conclusion of this theorem is valid for the Levenberg–Marquardt method.

Another relevant algorithm in this context is the LP-Newton method introduced in [11], with the iteration subproblem of the form

$$\begin{aligned} &\text{minimize } \gamma \\ &\text{subject to } \|\Phi(u^k) + \Phi'(u^k)v\| \leq \gamma\|\Phi(u^k)\|^2, \\ &\qquad\quad \|v\| \leq \gamma\|\Phi(u^k)\|, \\ &\qquad\quad (v,\ \gamma) \in \mathbb{R}^p \times \mathbb{R}. \end{aligned} \tag{37}$$

As demonstrated in [10, 11] (see also [17]), local convergence properties of this method near noncritical solutions are the same as for the Levenberg–Marquardt method. Yet

again, thinking of critical solutions, and following the development in [30, Sect. 3.2], one can embed the LP-Newton method into the pNM framework above, and obtain counterpart of Izmailov et al. [30, Corollary 2], saying that under the smoothness and 2-regularity assumptions in Theorem 3.1, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\delta}(\bar{u};\ \bar{v})$ there exists a sequence $\{u^k\} \subset \mathbb{R}^p$ such that for each k the pair $(u^{k+1} - u^k,\ \gamma_{k+1})$ with some γ_{k+1} solves (37), and for any such sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u};\ \bar{v})$, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically, and (29) and (30) hold. (Observe that uniqueness of $\{u^k\}$ is not claimed in this case, and indeed, (37) may have nonunique solutions.)

We finally mention the stabilized Newton–Lagrange (sequential quadratic programming) method for equality-constrained optimization problems [15, 28, 34, 45]; see also [35, Chapter 7]. It can also be covered by Theorem 3.1, thus relaxing the smoothness hypothesis in [30, Sect. 3.3], thus generalizing [30, Corollary 3]. We do not go into more detail regarding this issue as this would require an extensive discussion, including introducing terminology not needed in this paper otherwise.

Remark 3.2 An extension of Izmailov et al. [30, Theorem 1] to the case of a constrained equation as in [20, Theorem 3.1] is also possible under the smoothness hypothesis of this work. Consider the problem

$$\Phi(u) = 0, \quad u \in P,$$

where $P \subset \mathbb{R}^p$ is a given closed convex set. Then the analysis in [20, Sect. 3] allows to conclude that under the assumptions of Theorem 3.1, with the additional requirement that $\bar{v}$ belongs to the interior of the radial cone to P at $\bar{u}$, the iterates u^k in that theorem can be additionally claimed to stay feasible (i.e., to belong to P for all k). This allows to cover the constrained Gauss–Newton method with the subproblem

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

the constrained Levenberg–Marquardt method [4, 12, 37, 47] with the subproblem

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

(cf. (35)); the version of the LP-Newton method with the additional constraint [11], with the subproblem

$$\begin{aligned} &\text{minimize } \gamma \\ &\text{subject to } \|\Phi(u^k) + \Phi'(u^k)v\| \leq \gamma\|\Phi(u^k)\|^2, \\ &\qquad\qquad \|v\| \leq \gamma\|\Phi(u^k)\|, \\ &\qquad\qquad u^k + v \in P \end{aligned}$$

(cf. (37)); as well as projected version of these methods; see [20, Sect.s 1.1, 3] for details.

Remark 3.3 According to Izmailov et al. [30, Remark 2], the estimates (29)–(30) in Theorem 3.1 imply that

$$\lim_{k\to\infty} \frac{\|u^{k+1} - \bar{u}\|}{\|u^k - \bar{u}\|} = \frac{1}{2},$$

i.e., $\{u^k\}$ converges to $\bar{u}$ linearly, with an asymptotic ratio exactly equal to $1/2$.

This convergence pattern serves as a basis for convergence acceleration techniques [25, 27], one of them being the so-called extrapolation. The simplest variant of it consists of generating an auxiliary sequence $\{\widehat{u}^k\}$ by doubling the Newtonian step: for each k, set

$$\widehat{u}^{k+1} = u^k + 2v^k. \tag{38}$$

According to Griewank [27, Theorem 4.1], one may expect $\{\widehat{u}^k\}$ to converge linearly with the asymptotic ratio of $1/4$, instead of $1/2$ for $\{u^k\}$, at least for the basic Newton method with the subproblem (34). Observe that this procedure can be easily incorporated into any implementations of the algorithms discussed above: (38) does not affect the main iteration sequence $\{u^k\}$, and is not concerned with any computational overhead except for one extra evaluation of Φ needed to assess the quality of the obtained $\widehat{u}^{k+1}$. The specified extrapolation procedure will be employed in Sect. 5.

4 Asymptotic Acceptance of the Full Step

We will deal with the issue specified in the title of this section for the following prototype algorithm combining the local perturbed Newton method framework with a linesearch globalization technique.

Algorithm 4.1 Choose $u^0 \in \mathbb{R}^p$, $\sigma \in (0,\ 1)$, $\theta \in (0,\ 1)$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Compute $v^k \in \mathbb{R}^p$ as a solution of (9).
3. Set $\alpha = 1$. If the inequality

$$\|\Phi(u^k + \alpha v^k)\| \le (1 - \sigma\alpha)\|\Phi(u^k)\| \tag{39}$$

   is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (39) again, etc., until (39) becomes valid.
4. Set $u^{k+1} = u^k + \alpha_k v^k$.
5. Increase k by 1 and go to Step 1.

The fact that Algorithm 4.1 (equipped with some further safeguards for the cases when Step 2 fails or produces a direction "of poor quality" [21]) is well-defined and possesses reasonable global convergence properties is supposed to be established for the specific instances of (9) at Step 2. The role of the perturbed Newton method framework is only local, which conforms with the local nature of our analysis, and in

principle, those global issues are not the subject of this work, but we will give some related comments in Remark 4.1 below.

Theorem 4.1 *Under the assumptions of Theorem 3.1, let the estimates (27) and (28) hold with removed* Π*, i.e.,*

$$\Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \tag{40}$$

and

$$\omega(u) = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3) \tag{41}$$

for $u \in K_{\varepsilon,\,\delta}(\bar{u};\ \bar{v})$ *as* $\varepsilon \to 0+$.

Then, for every $\widehat{\varepsilon} > 0$ *and* $\widehat{\delta} > 0$*, there exist* $\varepsilon = \varepsilon(\bar{v}) > 0$ *and* $\delta = \delta(\bar{v}) > 0$ *such that for any starting point* $u^0 \in K_{\varepsilon,\,\delta}(\bar{u};\ \bar{v})$ *Algorithm 4.1 with* $\sigma \in (0,\ 3/4)$ *uniquely defines the sequence* $\{u^k\}$, $u^k \in K_{\widehat{\varepsilon},\,\widehat{\delta}}(\bar{u},\ \bar{v})$ *for all k, and* $\alpha_k = 1$ *holds for all k large enough.*

Observe that conditions (40) and (41) imply both (10) and (11), and of course cover the case when $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$, and (9) turns into the basic (unperturbed) Newton scheme (34), while Algorithm 4.1 turns into its instance considered in [22, Algorithm 1]. Therefore, Theorem 4.1 generalizes [22, Proposition 3], both in a sense of weaker smoothness assumptions, and of allowed perturbations of the basic Newton scheme.

Proof As in Lemma 3.1, let $\bar{u} = 0$, and let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0,\ 1)$ be chosen according to that lemma. Then for $u \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u};\ \bar{v})\backslash\{\bar{u}\}$, there exists the unique solution v of (9). Moreover, by the argument in the proof of Izmailov et al. [30, Theorem 1] we then have

$$\|u_1\| \le \bar{\delta}\|u\| \le \frac{\bar{\delta}}{1-\bar{\delta}}\|u_2\|, \tag{42}$$

and hence, estimates (31), (33) yield

$$u + v = \frac{1}{2}u_2 + O(\|u_1\|) + O(\|u\|^2) = \frac{1}{2}u_2 + O(\bar{\delta}\|u_2\|) + O(\|u_2\|^2), \tag{43}$$

$$\frac{1}{2}u + v = O(\|u_1\|) + O(\|u\|^2) = O(\bar{\delta}\|u_2\|) + O(\|u_2\|^2) \tag{44}$$

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

According to (2) and (5),

$$\begin{aligned}\Phi(u) &= \Phi'(\bar{u})u + \frac{1}{2}(\Phi')'(\bar{u};\ u)u + O(\|u\|^3)\\ &= \Phi'(\bar{u})u_1 + \frac{1}{2}(\Phi')'(\bar{u};\ u_2)u_2 + O(\|u_1\|^2) + O(\|u_1\|\|u_2\|) + O(\|u\|^3)\end{aligned}$$

$$= \Phi'(\bar{u})u_1 + \frac{1}{2}(\Phi')'(\bar{u};\, u_2)u_2 + O(\bar{\delta}\|u_2\|^2) + O(\|u_2\|^3) \tag{45}$$

as $u \to 0$ and $\bar{\delta} \to 0+$, where the second equality is by Lipschitz continuity of $(\Phi')'(\bar{u};\, \cdot)$, while the last one is by (42). Furthermore, by the same reasoning, but also employing (43), we obtain that

$$\begin{aligned}\Phi(u+v) &= \Phi'(\bar{u})(u+v) + \frac{1}{2}(\Phi')'(\bar{u};\, u+v)(u+v) + O(\|u+v\|^3)\\ &= \Phi'(\bar{u})(u+v) + \frac{1}{8}(\Phi')'(\bar{u};\, u_2)u_2 + O(\bar{\delta}\|u_2\|^2) + O(\|u_2\|^3)\end{aligned} \tag{46}$$

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

Since v is a solution of (9), by (2)–(5) and (40)–(41), we conclude that

$$\begin{aligned}0 &= -\Phi(u) - \Phi'(u)v - \Omega(u)v + \omega(u)\\ &= -\Phi'(\bar{u})u - \frac{1}{2}(\Phi')'(\bar{u};\, u)u - \Phi'(\bar{u})v - (\Phi')'(\bar{u};\, u)v\\ &\quad + O(\|u\|^3) + O(\|u\|^2\|v\|) + O(\|u_1\|\|v\|) + O(\|u\|\|u_1\|),\end{aligned}$$

which by (42), (44) implies that

$$\begin{aligned}\Phi'(\bar{u})(u+v) &= -(\Phi')'(\bar{u};\, u)\left(\frac{1}{2}u + v\right)\\ &\quad + O(\|u^3\|) + O(\|u\|^2\|v\|) + O(\|u_1\|\|v\|) + O(\|u\|\|u_1\|)\\ &= O(\bar{\delta}\|u_2\|^2) + O(\|u_2\|^3).\end{aligned}$$

Substituting the latter into (46) yields

$$\Phi(u+v) = \frac{1}{8}(\Phi')'(\bar{u};\, u_2)u_2 + O(\bar{\delta}\|u_2\|^2) + O(\|u_2\|^3) \tag{47}$$

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

Estimates (45) and (47) comprise what is needed for the analysis leading to Fischer et al. [22, Proposition 3] to go through when combined with the following additional facts none of which requires stronger smoothness assumptions. First, 2-regularity of Φ in a direction $\bar{v} \in \ker \Phi'(\bar{u})\backslash\{0\}$ implies that $\Pi(\Phi')'(\bar{u};\, \bar{v})\bar{v} = B(\bar{v})\bar{v} \neq 0$, and then it can be seen that $\bar{\delta} > 0$ can be chosen in such a way that there exists $\gamma > 0$ such that

$$\|(\Phi')'(\bar{u};\, u_2)u_2\| \geq \|\Pi(\Phi')'(\bar{u};\, u_2)u_2\| \geq \gamma\|u_2\|^2.$$

Second, (13), (33), and (42) imply the estimates

$$u_1 + v_1 = O(\|u\|\|u_1\|) + O(\|u\|^3) = O(\bar{\delta}\|u_2\|^2) + O(\|u_2\|^3), \tag{48}$$

$$u_2 + v_2 = \frac{1}{2}u_2 + O(\|u_1\|) + O(\|u\|^2) = \frac{1}{2}u_2 + O(\bar{\delta}\|u_2\|) + O(\|u_2\|^2) \tag{49}$$

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

Observe that unlike for the local convergence result in Theorem 3.1, the estimate (48) (that is sharper than (31)) is essential here, as together with (49), it allows to conclude that for every $\bar{\gamma} > 0$, one can chose $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ in such a way that

$$\|u_1 + v_1\| \le \bar{\gamma}\|u_2 + v_2\|^2,$$

yielding another key ingredient of this analysis. □

Remark 4.1 Algorithm 4.1 makes perfect sense when used with the basic Newton scheme (34) at Step 2 (i.e., with $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$ in (9)), and with Euclidean norm in (39) at Step 3; see the related comments in [22]. In some sense, this remains true for the Levenberg–Marquardt method with the iteration system (36) as well, since the function $\varphi : \mathbb{R}^p \to \mathbb{R}_+$, $\varphi(u) := \|\Phi(u)\|$, defined using the Euclidian norm, is differentiable at any point u^k such that $\Phi(u^k) \ne 0$ (cf. Step 1 of Algorithm 4.1), and

$$\varphi'(u^k) = (\Phi'(u^k))^\top \Phi(u^k)/\|\Phi(u^k)\|,$$

and hence, for the solution v^k of (36) it holds that

$$\langle \varphi'(u^k),\ v^k\rangle = -\langle((\Phi'(u^k))^\top \Phi'(u^k) + \rho(u^k)I)v^k,\ v^k\rangle/\|\Phi(u^k)\| < 0.$$

Therefore, v^k is a direction of descent for φ at u^k. That said, we emphasize that here we only discuss a principal possibility of using the Levenberg–Marquardt directions with linesearch tests like (39), i.e., we only consider the descent property of these directions for the residual. In particular, we do not discuss finite termination of backtracking procedures using this test, as this would still not guarantee global convergence of the overall algorithm, and we do not state here any formal results of this kind, as this is beyond the scope of this paper focusing on local analysis. Moreover, actual linesearch algorithms with known global convergence guarantees, involving the Levenberg–Marquardt directions, either employ, in a hybrid manner, some kind of safeguards for the case when the quality of descent is insufficient (like in [7, 13]), or special linesearch tests (like in [23]).

Observe that the result on asymptotic acceptance of the full step for the Levenberg–Marquardt method with $\rho(\cdot) := \|\Phi(\cdot)\|^\tau$, $\tau \ge 2$, in cases of convergence to (critical) solutions with the needed 2-regularity property, following from Theorem 4.1 and considerations in [30, Sect. 3.1] (recall also Remark 3.1), is new even in the case of twice differentiable Φ.

As for the LP-Newton method, the natural choice of the norm in the subproblem (37) is the infinity-norm, as it makes (37) a linear programming problem. In any case, the globalization procedure proposed in [18, Algorithm 1] employs the stepsize test

of the form

$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \sigma\alpha)\|\Phi(u^k)\| + \sigma\alpha\gamma_{k+1}\|\Phi(u^k)\|^2$$

with the same norm as the one appearing in (37). This test is evidently weaker than (39) (with the same norm), and hence, accepts the unit stepsize once (39) does. Therefore, Theorem 4.1 and considerations in [30, Sect. 3.2] (recall also Remark 3.1 again) yield the result on asymptotic acceptance of the full step for the LP-Newton method, under the needed assumptions.

In completion of this section we note that, unlike in [22], under the current smoothness assumptions one cannot expect the set of excluded directions for starlike domains of convergence and asymptotic acceptance of the full step to be thin, even for the basic (unperturbed) Newton method; see Examples 5.1–5.3 below.

5 Applications to a Smooth Reformulation of Nonlinear Complementarity Problems and Numerical Results

Consider the nonlinear complementarity problem (NCP)

$$u \geq 0, \quad F(u) \geq 0, \quad \langle u,\ F(u)\rangle = 0, \tag{50}$$

where $F : \mathbb{R}^p \to \mathbb{R}^p$ is a given smooth mapping. Using the complementarity function $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$\psi(a,\ b) := 2ab - (\min\{0,\ a + b\})^2 \tag{51}$$

(originally introduced in [9]), NCP (50) is equivalently reformulated as (7) with

$$\Phi(u) := \psi(u,\ F(u)), \tag{52}$$

where ψ is applied componentwise. The function ψ in (51) is one of known smooth complementarity functions [36, 40]: assuming that F is differentiable at $u \in \mathbb{R}^p$, the corresponding mapping defined in (52) is also differentiable at u, with the Jacobian $\Phi'(u)$ having the rows

$$\Phi_i'(u) = 2u_i F_i'(u) + 2F_i(u)e^i - 2\min\{0,\ u_i + F_i(u)\}(F_i'(u) + e^i), \quad i = 1,\ \ldots,\ p, \tag{53}$$

where $e^1,\ \ldots,\ e^p$ is the standard basis in $\mathbb{R}^p$. From [35, Proposition 1.75] it then follows that if F' is strongly semismooth at $\bar{u} \in \mathbb{R}^p$ (in particular, if it is twice differentiable near $\bar{u}$, with its second derivative being Lipschitz-continuous near $\bar{u}$), then Φ' is strongly semismooth at $\bar{u}$.

If $\bar{u}$ is a solution of NCP (50), then the disjoint index sets

$$\begin{aligned} I_0(\bar{u}) &:= \{i = 1, \ldots, p \mid \bar{u}_i = F_i(\bar{u}) = 0\}, \\ I_1(\bar{u}) &:= \{i = 1, \ldots, p \mid \bar{u}_i > 0, \ F_i(\bar{u}) = 0\}, \\ I_2(\bar{u}) &:= \{i = 1, \ldots, p \mid \bar{u}_i = 0, \ F_i(\bar{u}) > 0\}, \end{aligned}$$

provide a partition of $\{1, \ldots, p\}$, and from (53) we have

$$\Phi_i'(\bar{u}) = \begin{cases} 0 & \text{if } i \in I_0(\bar{u}), \\ 2\bar{u}_i F_i'(\bar{u}) & \text{if } i \in I_1(\bar{u}), \\ 2F_i(\bar{u})e^i & \text{if } i \in I_2(\bar{u}). \end{cases} \tag{54}$$

This implies that if $I_0(\bar{u}) \neq \emptyset$, meaning violation of the strict complementarity condition at $\bar{u}$, then $\bar{u}$ is necessarily a singular solution of Equation (7).

From (53) one can easily obtain that for any $v \in \mathbb{R}^p$ and $i \in I_0(\bar{u})$

$$\begin{aligned} (\Phi_i')'(\bar{u};\, v) &= 2(v_i - \min\{0,\, v_i + \langle F_i'(\bar{u}),\, v\rangle\})F_i'(\bar{u}) \\ &\quad + 2(\langle F_i'(\bar{u}),\, v\rangle - \min\{0,\, v_i + \langle F_i'(\bar{u}),\, v\rangle\})e^i \\ &= 2\max\{v_i,\, -\langle F_i'(\bar{u}),\, v\rangle\}F_i'(\bar{u}) - 2\min\{v_i,\, -\langle F_i'(\bar{u}),\, v\rangle\}e^i. \end{aligned}$$

Then from (54), we derive that the key assumption of 2-regularity of Φ at $\bar{u}$ in some direction $\bar{v} \in \ker \Phi'(\bar{u})$ automatically holds for any $\bar{v} \in \mathbb{R}^p$ such that

$$\langle F_i'(\bar{u}),\, \bar{v}\rangle = 0, \ i \in I_1(\bar{u}), \quad \bar{v}_i = 0, \ i \in I_2(\bar{u}), \tag{55}$$

and the matrix with the rows

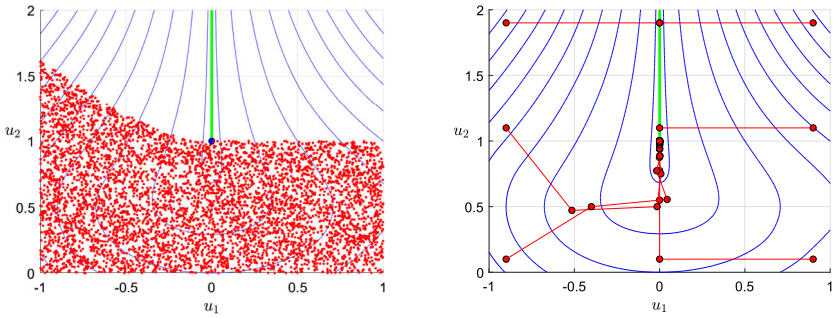
$$\begin{matrix} \max\{\bar{v}_i,\, -\langle F_i'(\bar{u}),\, \bar{v}\rangle\}F_i'(\bar{u}) - \min\{\bar{v}_i,\, -\langle F_i'(\bar{u}),\, \bar{v}\rangle\})e^i, \ i \in I_0(\bar{u}), \\ F_i'(\bar{u}), \ i \in I_1(\bar{u}), \\ e^i, \ i \in I_2(\bar{u}), \end{matrix} \tag{56}$$

is nonsingular. The latter sufficient condition for 2-regularity of Φ at $\bar{u}$ in a direction $\bar{v}$ evidently implies that

$$F_i'(\bar{u}), \ i \in I_1(\bar{u}), \quad e^i, \ i \in I_2(\bar{u}), \quad \text{are linearly independent,} \tag{57}$$

and moreover, this sufficient condition also becomes necessary under (57). The general characterization of 2-regularity in the current context, not assuming (57), can be found in [42]. For easier understanding of the essence of the properties in question, here we restrict ourselves to the case when singularity is imposed in a natural way, i.e., only by violation of strict complementarity at $\bar{u}$, or, in other words, when (57) holds. That said, see Example 5.3 below, demonstrating the case when the key assumption holds in the absence of (57).

Example 5.1 [19, Example 1] Let $p = 1$, $F(u) = u^2$. Then NCP (50) has the unique solution $\bar{u} = 0$, with $I_1(\bar{u}) = I_2(\bar{u}) = \emptyset$, $F'(\bar{u}) = 0$, and the first line in (56) is



(a) Domain of attraction to critical solution (b) Iterative sequences

Fig. 1 Example 5.2: the Newton method

positive if $\bar{v} < 0$, and equals 0 otherwise. Therefore, the key assumption holds with any $\bar{v} < 0$, but not with $\bar{v} \geq 0$.

Being initialized at $u^0 < 0$, Algorithm 4.1 employing the basic Newton method, and with $\sigma < 3/4$, converges to $\bar{u}$ by full steps (from some iteration on), and the rate of convergence is linear with the asymptotic ratio $1/2$. For $\sigma \geq 3/4$, the full step is never accepted (the ultimate stepsize value is $\alpha = 0.5$ for $\sigma = 3/4$, and approaches 0 as σ approaches 1), and the linear convergence rate is lower (with the asymptotic ratio $3/4$ for $\sigma = 3/4$, and approaching 1 as σ approaches 1).

The case when $u^0 > 0$ is not covered by the theory above, and the method ultimately accepts the unit stepsize for sufficiently small values of σ (only for those smaller than some threshold $\bar{\sigma} \in (0,\ 3/4)$), but in such cases the rate of convergence is linear with the asymptotic ratio $2/3$. This specific rate is explained by the fact that for $u > 0$, it holds that $\Phi(u) = 2u^3$, and the Newton iteration at $u^k > 0$ produces $u^{k+1} = 2u^k/3$. This also agrees with the theory developed in [26] for arbitrarily smooth equations and for the basic Newton method, allowing for higher-order regularity when $\Phi''(\bar{u}) = 0$ (as it essentially happens in this case).

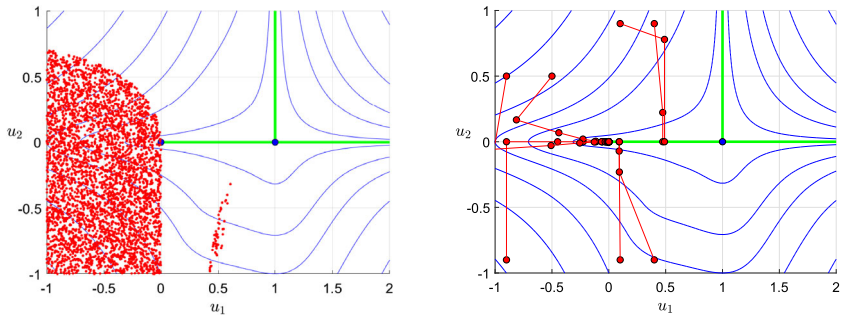
Example 5.2 (Test problem affknot1 in [42]) Let $p = 2$, $F(u) = (u_2 - 1,\ u_1)$. Then NCP (50) has the solution set $\{0\} \times [1,\ +\infty)$ (thick line in Fig. 1, where thin lines are contours of $\|\Phi(\cdot)\|$), with $\bar{u} = (0,\ 1)$ (thick dot in Fig. 1) being the unique critical solution, and $I_0(\bar{u}) = \{1\}$, $I_1(\bar{u}) = \{2\}$, $I_2(\bar{u}) = \emptyset$, $F_1'(\bar{u}) = e^2$, $F_2'(\bar{u}) = e^1$. Condition (55) yields $\bar{v}_1 = 0$, while the matrix with rows given by (56) takes the form

$$\begin{pmatrix} 0 & -\bar{v}_2 \\ 1 & 0 \end{pmatrix}$$

when $\bar{v}_2 < 0$, and

$$\begin{pmatrix} \bar{v}_2 & 0 \\ 1 & 0 \end{pmatrix}$$

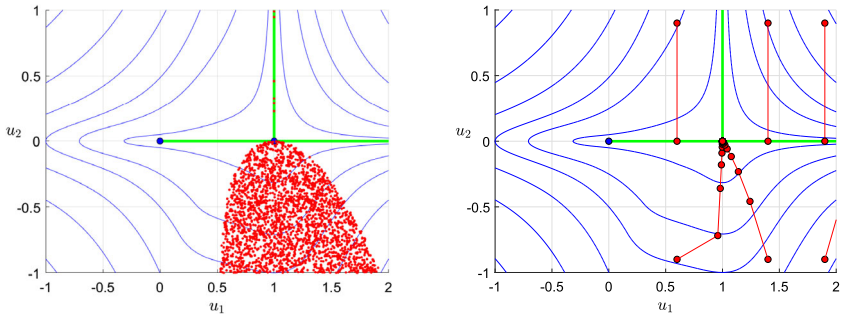
otherwise. Therefore, the key assumption holds with $\bar{v} = (0,\ \bar{v}_2)$ for any $\bar{v}_2 < 0$, but does not hold for $\bar{v}_2 \geq 0$.



(a) Domain of attraction to critical solution (0, 0)

(b) Iterative sequences

Fig. 2 Example 5.3: the Newton method



(a) Domain of attraction to critical solution (1, 0)

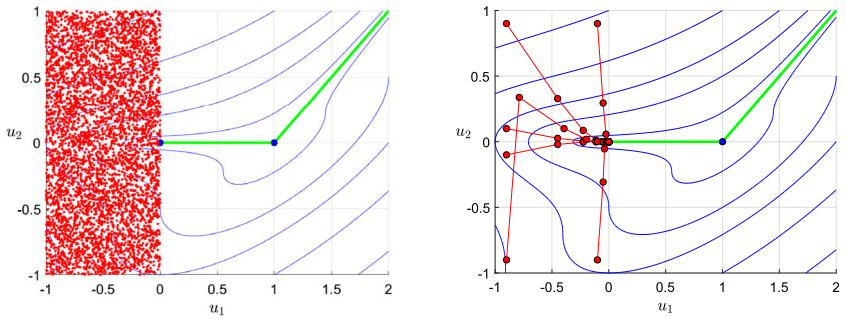
(b) Iterative sequences

Fig. 3 Example 5.3: the Newton method

Figure 1a demonstrates the starting points from which convergence of the basic Newton method to the critical solution $\bar{u}$ was detected. In order to obtain this figure, we initialized the method at 10000 random starting points distributed uniformly in the cubic neighborhood of $\bar{u}$ with the half-edge equal to 1. The runs were terminated with success when the residual $\|\Phi(u^k)\|$ was achieving a value below 10^{-11}, and out of these cases, convergence to $\bar{u}$ was claimed when $\|u^k - \bar{u}\|$ at successful termination was smaller than 10^{-3}. Figures with domains of attraction for other examples below were generated similarly. Note that the tolerance 10^{-3} is a compromise between the tasks of numerically detecting the cases of convergence and non-convergence to the solution of interest.

Figure 1b shows some typical iterative sequences. The observed pattern of convergence to $\bar{u}$ agrees with the developed theory, and the full step is ultimately accepted.

Example 5.3 [2, Example 3.3] Let $p = 2$, $F(u) = ((u_1 - 1)u_2, (u_1 - 1)^2)$. Then NCP (50) has the solution set $(\mathbb{R}_+ \times \{0\}) \cup (\{1\} \times \mathbb{R}_+)$, with (0, 0) and (1, 0) being the only critical solutions. Figures 2 and 3 provide the same kind of information as Fig. 1 above.



(a) Domain of attraction to critical solution (0, 0)

(b) Iterative sequences

Fig. 4 Example 5.4: the Newton method

For $\bar{u} = (0,\ 0)$ we have $I_0(\bar{u}) = \{1\}$, $I_1(\bar{u}) = \emptyset$, $I_2(\bar{u}) = \{2\}$, $F_1'(\bar{u}) = -e^2$, $F_2'(\bar{u}) = 2e^1$. Condition (55) yields $\bar{v}_2 = 0$, while the matrix with rows given by (56) takes the form

$$\begin{pmatrix} -\bar{v}_1 & 0 \\ 0 & 1 \end{pmatrix}$$

when $\bar{v}_1 < 0$, and

$$\begin{pmatrix} 0 & -\bar{v}_1 \\ 0 & 1 \end{pmatrix}$$

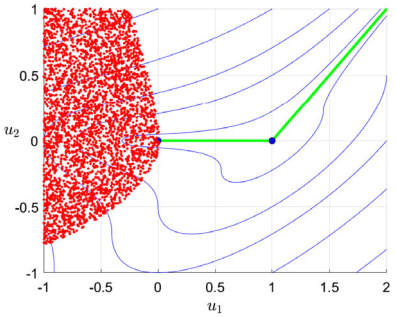
otherwise. Therefore, the key assumption holds with $\bar{v} = (\bar{v}_1,\ 0)$ for any $\bar{v}_1 < 0$, but does not hold for $\bar{v}_1 \geq 0$.

Figure 2a demonstrates the starting points from which convergence of the basic Newton method to the critical solution $\bar{u} = (0,\ 0)$ was detected, while Fig. 2b shows some typical iterative sequences. The observed pattern of convergence to $\bar{u}$ agrees with the developed theory, and the full step is ultimately accepted.

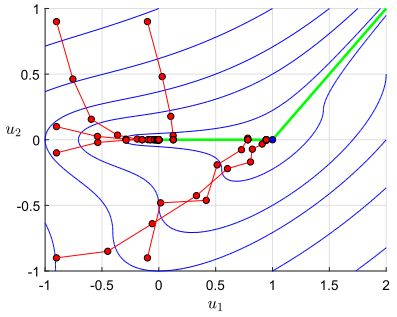
For $\bar{u} = (1,\ 0)$ we have $I_0(\bar{u}) = \{2\}$, $I_1(\bar{u}) = \{1\}$, $I_2(\bar{u}) = \emptyset$, $F_1'(\bar{u}) = F_2'(\bar{u}) = (0,\ 0)$, implying, in particular, that (57) does not hold. Nevertheless, it can be seen that the key assumption holds with any $\bar{v}$ such that $\bar{v}_2 < 0$.

Figure 3 is intended to emphasize the role of the critical solution $\bar{u} = (1,\ 0)$. Observe that $\varepsilon(\bar{v}) \to 0$ as $\bar{v}_2 \to 0-$, i.e., as $\bar{v}$ approaches nonzero directions in the u_1-axis, in which 2-regularity is violated. That said, the boundary of the attraction domain in Fig. 3a is tangential to the u_1-axis at $(1,\ 0)$, and $\varepsilon(\bar{v})$ is positive for every direction $\bar{v}$ with $\bar{v}_2 < 0$. A similar effect is observed in other examples.

Example 5.4 [2, Example 3.2] Let $p = 2$, $F(u) = (0,\ -u_1 + u_2 + 1)$. Then NCP (50) has the solution set $([0,\ 1] \times \{0\}) \cup \{(t + 1,\ t) \mid t \geq 0\}$, with $(0,\ 0)$ and $(1,\ 0)$ being the only critical solutions. Figures 4, 5 and 6 provide the same kind of information as Figs. 1, 2 and 3 above, though Figs. 5 and 6 are for the Levenberg–Marquardt method rather than the basic Newton method.

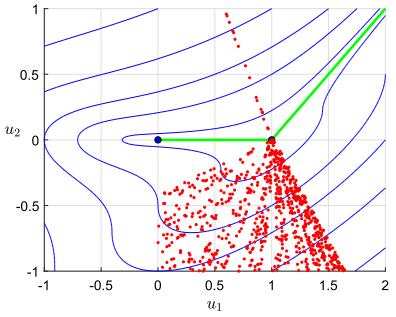


(a) Domain of attraction to critical solution $(0,\ 0)$

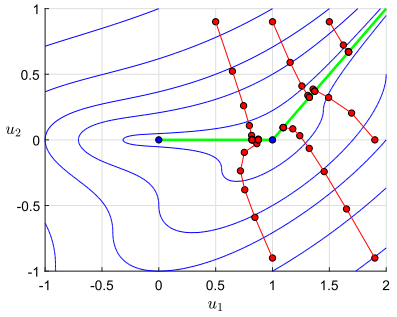


(b) Iterative sequences

Fig. 5 Example 5.4: the Levenberg–Marquardt method



(a) Domain of attraction to critical solution $(1,\ 0)$



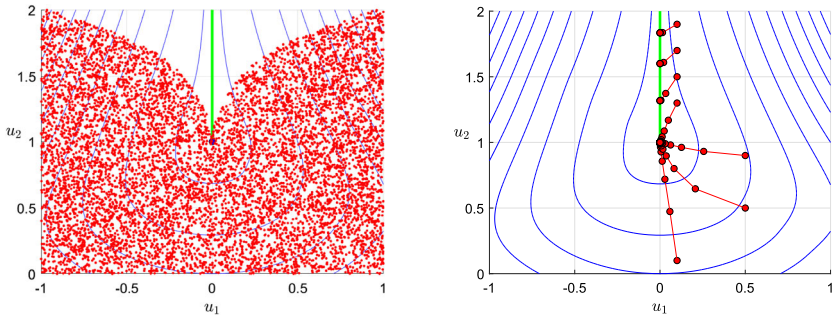
(b) Iterative sequences

Fig. 6 Example 5.4: the Levenberg–Marquardt method

For $\bar{u} = (0,\ 0)$ as in the previous examples one can check that the key assumption holds with $\bar{v} = (\bar{v}_1,\ 0)$ for any $\bar{v}_1 < 0$, and Figs. 4 and 5 reflect this fact.

Furthermore, one can see that $\Phi_1'(u) = -2\min\{0,\ u_1\}e^1 = 0$ for all $u \in \mathbb{R}^2$ with $u_1 \geq 0$, implying that $\Phi'(u)$ is singular for all such u, and in particular, it is singular in a neighborhood of $\bar{u} = (1,\ 0)$. Therefore, the key assumption cannot hold at this $\bar{u}$, and the Newton method is not well-defined near this solution. At the same time, the Levenberg–Marquardt method behaves nicely near this solution, and does not exhibit any tendency of convergence to it; see Fig. 6. In particular, the sparse set in Fig. 6a is actually a result of using an approximate test on closeness of the iterate at termination to $\bar{u}$, with rather rough tolerance 10^{-3}. Further reducing this tolerance makes the "domain of attraction" being shown more and more sparse, and eventually eliminates it completely at the level 10^{-6}.

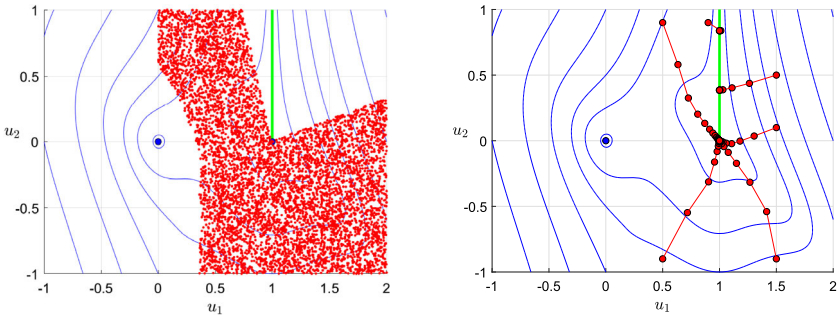
Example 5.5 (Test problem quadknot in [42]) Let $p = 2$, $F(u) = (u_2 - 1,\ u_1^2)$. Then NCP (50) has the solution set $\{0\} \times [1,\ +\infty)$, with $\bar{u} = (0,\ 1)$ being the only critical solution. See Fig. 7.



(a) Domain of attraction to critical solution (0, 1)

(b) Iterative sequences

Fig. 7 Example 5.5: the Levenberg–Marquardt method



(a) Domain of attraction to critical solution (1, 0)

(b) Iterative sequences

Fig. 8 Example 5.6: the Levenberg–Marquardt method

Example 5.6 [2, Example 3.4] Let $p = 2$, $F(u) = ((u_1-1)^2+(u_1-1)u_2,\ (u_1-1)^2)$. Then apart from a strictly complementary solution (0, 0), NCP (50) has the solution set $\{1\} \times \mathbb{R}_+$, with $\bar{u} = (1,\ 0)$ being the only critical solution. See Fig. 8.

We complete the paper with numerical results for a collection of small NCPs taken from Oberlin and Wright [42], and for some other examples of NCP with solutions violating strict complementarity, taken from various sources. The algorithms being tested were applied to (7) with Φ defined according to (51)–(52).

Table 1 presents the results for Algorithm 4.1 employing the basic Newton method with the subproblem (34), and with $\sigma = 0.01$ and $\theta = 1/2$ (abbreviated below as "NM"), as well as for the version of the method supplied with the simplest extrapolation procedure defined according to (38) (abbreviated as "NM-EP"). Successful termination was declared when the Euclidean residual of (7) at the main or extrapolated iterate was getting smaller or equal to 10^{-11}, within 50 iterations. The identifiers of test problems with the key assumption satisfied at the singular solution in question are **boldfaced**. Some of the test problems have two solutions of interest, and then their identifiers have additional attributes 1 or 2. For each test problem, we performed a single run from the "recommended" starting point (when available; abbreviated as

Table 1 Numerical results for NCPs: the Newton method

Test problems	NM		NM-EP	
	Rec	Rand	Rec	Rand
quarp	15	18 (100%, 5.1e−4)	14	16 (100%, 6.7e−4)
DIS61	19	17 (100%, 1.7e−6)	10	9 (100%, 1.2e−6)
quarquad, 1	16	20 (100%, 6.7e−7)	8	13 (100%, 4.0e−7)
quarquad, 2	18	20 (97%, 5.6e−4)	17	17 (97%, 6.2e−4)
affknot1	20	11 (100%, 4.9e−6)	2	5 (100%, 2.8e−7)
affknot2	18	19 (100%, 4.5e−6)	1	9 (100%, 5.7e−13)
quadknot	18	19 (100%, 2.1e−6)	9	9 (100%, 1.3e−6)
munson4	19	19 (100%, 1.7e−6)	10	9 (100%, 1.5e−6)
DIS64	–	19 (100%, 1.7e−6)	–	1 (100%, 1.2e−16)
ne-hard	25	22 (100%, 2.1e−6)	16	15 (100%, 2.7e−7)
doubleknot	21	19 (88%, 2.6e−6)	11	10 (88%, 6.6e−13)
quad1	15	19 (100%, 7.2e−5)	5	13 (100%, 7.0e−5)
quad2	20	19 (100%, 6.3e−5)	9	12 (100%, 7.0e−5)
quarn	15	19 (100%, 1.0e−4)	14	18 (100%, 7.3e−4)
[2, **Example 3.1]**	–	–	–	–
[2, **Example 3.2]**	18	18 (50%, 2.3e−6)	9	8 (51%, 8.4e−13)
[2, **Example 3.3]**, 1	20	10 (100%, 2.8e−6)	11	4 (100%, 3.4e−7)
[2, **Example 3.3]**, 2	19	5 (100%, 2.3e−6)	5	2 (100%, 3.0e−5)
[2, **Example 3.4]**	20	17 (100%, 7.0e−5)	11	11 (100%, 6.0e−5)
[8, **Example 6]**	19	19 (100%, 1.9e−6)	1	1 (100%, 2.2e−16)
[8, **Example 7]**	19	18 (100%, 2.8e−6)	1	1 (100%, 2.7e−16)
[8, **Example 8]**	20	17 (100%, 1.9e−6)	11	8 (100%, 1.3e−6)
[8, **Example 9]**	19	19 (100%, 2.1e−6)	1	1 (100%, 6.0e−16)
[19, **Example 1]**	19	19 (100%, 7.2e−5)	8	12 (100%, 7.0e−5)
[19, **Example 3]**, 1	21	19 (100%, 6.9e−5)	16	13 (100%, 3.8e−9)
[19, **Example 3]**, 2	19	19 (100%, 6.9e−5)	10	13 (100%, 6.2e−5)
[20, **Example 4.2]**	19	18 (100%, 2.1e−6)	1	1 (100%, 1.4e−16)
[20, **Example 4.3]**	19	19 (100%, 1.6e−6)	9	8 (100%, 1.3e−10)
[29, **Example 3.4]**	19	18 (100%, 1.6e−6)	1	1 (100%, 6.3e−18)
[29, **Example 3.5]**	20	19 (100%, 1.6e−6)	1	1 (100%, 8.4e−16)
[32, **Example 1]**	19	19 (100%, 1.7e−6)	7	9 (100%, 1.2e−6)
[32, **Example 2]**	19	18 (100%, 2.0e−6)	1	1 (100%, 7.2e−16)

"Rec"), and also 1000 runs from randomly generated starting points distributed uniformly in the cubic neighborhood of the solution in question, with a half-edge equal to 1 (abbreviated as "Rand"). For the former case, we report only the iteration count, while for the latter we report the average iteration count over successful runs (rounded up to the nearest integer), and additionally the percentage of successful runs and the

Table 2 Numerical results for NCPs: the Levenverg–Marquardt method

Test problems	LMM		LMM-EP	
	Rec	Rand	Rec	Rand
quarp	15	18 (100%, 9.5e−4)	14	16 (100%, 5.3e−4)
DIS61	19	18 (100%, 3.8e−6)	10	10 (100%, 1.2e−6)
quarquad, 1	17	21 (100%, 1.3e−6)	8	14 (100%, 4.1e−7)
quarquad, 2	16	20 (100%, 6.4e−4)	15	17 (100%, 6.8e−4)
affknot1	20	13 (100%, 2.3e−6)	11	7 (100%, 5.6e−7)
affknot2	18	19 (100%, 2.2e−6)	7	9 (100%, 2.3e−8)
quadknot	19	17 (100%, 4.3e−6)	9	9 (100%, 1.4e−6)
munson4	20	19 (100%, 3.6e−6)	11	10 (100%, 1.5e−6)
DIS64	28	19 (100%, 3.9e−6)	16	7 (100%, 1.2e−6)
ne-hard	-	22 (29%, 1.3e−6)	43	14 (28%, 8.7e−7)
doubleknot	21	20 (100%, 5.4e−6)	12	10 (100%, 6.5e−8)
quad1	16	20 (100%, 1.4e−4)	5	13 (100%, 7.0e−5)
quad2	20	20 (100%, 1.4e−4)	9	13 (100%, 7.2e−5)
quarn	15	19 (100%, 3.5e−4)	14	18 (100%, 6.3e−4)
[2, **Example 3.1**]	20	11 (100%, 2.1e−6)	7	5 (100%, 1.1e−6)
[2, **Example 3.2**]	19	12 (99%, 3.0e−6)	10	8 (99%, 3.3e−7)
[2, **Example 3.3**], 1	20	11 (100%, 3.4e−6)	11	7 (100%, 2.8e−07)
[2, **Example 3.3**], 2	19	8 (100%, 2.3e−6)	7	5 (100%, 1.0e−5)
[2, **Example 3.4**]	20	14 (100%, 8.0e−5)	11	10 (100%, 6.7e−5)
[8, **Example 6**]	20	19 (100%, 1.7e−3)	8	7 (100%, 1.1e−6)
[8, **Example 7**]	19	18 (100%, 2.4e−6)	8	7 (100%, 1.6e−6)
[8, **Example 8**]	21	18 (100%, 2.0e−6)	11	9 (100%, 1.3e−6)
[8, **Example 9**]	20	19 (100%, 1.9e−6)	9	7 (100%, 1.3e−6)
[19, **Example 1**]	20	19 (100%, 7.1e−5)	9	12 (100%, 7.1e−5)
[19, **Example 3**], 1	21	19 (100%, 1.4e−6)	16	14 (100%, 3.8e−9)
[19, **Example 3**], 2	20	19 (100%, 5.8e−5)	10	14 (100%, 7.1e−5)
[20, **Example 4.2**]	20	17 (100%, 2.2e−6)	11	7 (100%, 4.5e−6)
[20, **Example 4.3**]	20	19 (100%, 1.6e−6)	10	9 (100%, 7.9e−9)
[29, **Example 3.4**]	19	18 (100%, 1.6e−6)	7	6 (100%, 9.9e−7)
[29, **Example 3.5**]	20	19 (100%, 3.1e−6)	8	8 (100%, 1.9e−6)
[32, **Example 1**]	19	19 (100%, 1.8e−6)	9	9 (100%, 1.3e−6)
[32, **Example 2**]	19	19 (100%, 2.0e−6)	7	7 (100%, 1.2e−6)

average distance to the solution of interest over cases when this distance at successful termination was no greater than 10^{-3} (in parenthesis, separated by commas). The cases when there were no successful runs are marked by "–".

Table 2 reports the same kind of information as Table 1 for Algorithm 4.1 with the same parameter values, but employing the Levenberg–Marquardt method with

the subproblem (36) making use of the regularization parameter $\rho(\cdot) := \|\Phi(\cdot)\|^2$ (abbreviated as LMM, and as LMM-EP for a version supplied with extrapolation).

The asymptotic acceptance of the full step was encountered in all runs of these experiments. Moreover, the full step was accepted almost always, except for some rare cases when it was not accepted on some early iteration (usually once per run, if at all). Moreover, for LMM, the iterations where the full step was not accepted were systematically encountered for DIS61 and quarquad, 2, only. These observations confirm the conclusions of Theorem 4.1: despite convergence to singular solutions, the full step is asymptotically accepted.

Furthermore, the results reported in Tables 1 and 2 clearly demonstrate the accelerating effect of the extrapolation procedure for problems satisfying the key assumption, both for the Newton and the Levenberg–Marquardt methods. This can be considered as an indirect evidence of the convergence pattern established in Theorem 3.1.

6 Conclusions

We have extended some known results on behavior of Newton-type methods (including the Levenberg–Marquardt and the LP-Newton methods) near singular (and perhaps nonisolated) solutions of nonlinear equations to the case when the operator of the equation possesses a strongly semismooth derivative, but is not necessarily twice differentiable. Specifically, we have presented the results on local linear convergence, and on asymptotic acceptance of the full step by linesearch versions of such algorithms. The results were further applied to nonlinear complementarity problems violating strict complementarity, and a collection of examples was presented demonstrating peculiarities of the smoothness assumptions used in this work.

Acknowledgements The work of the first and the third author was supported by the Volkswagen Foundation—97775, and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—409756759. The second author was funded by the Russian Science Foundation Grant 23-11-20020 (https://rscf.ru/en/project/23-11-20020/).

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

References

1. Adams, J.F., Lax, P.D., Phillips, R.S.: On matrices whose real linear combinations are nonsingular. Proc. Am. Math. Soc. **16**(2), 318–322 (1965)
2. Arutyunov, A.V., Izmailov, A.F.: Stability of possibly nonisolated solutions of constrained equations with applications to complementarity and equilibrium problems. Set-Valued Var. Anal. **26**, 327–352 (2018)
3. Avakov, E.R., Agrachev, A.A., Arutyunov, A.V.: The level set of a smooth mapping in a neighborhood of a singular point, and zeros of quadratic mapping. Math. USSR Sbornik **73**, 455–466 (1992)
4. Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. **6**, 927–940 (2012)
5. Burke, J.V., Ferris, M.C.: Weak sharp minima in mathematical programming. SIAM J. Control. Optim. **31**(5), 1340–1359 (1993)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)

7. 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)
8. Daryina, A.N., Izmailov, A.F.: On identification of active indices in mixed complementarity problems. In: Bereznev, V.A. (ed.) Questions of Modelling and Analysis in Decision Making Problems, pp. 72–87. CCAS, Moscow (2004)
9. Evtushenko, Y.G., Purtov, V.A.: Sufficient conditions for a minimum for nonlinear programming problems. Soviet Math. Dokl. **30**, 313–316 (1984)
10. 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)
11. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**, 1–36 (2014)
12. Facchinei, F., Fischer, A., Kanzow, C., Peng, J.-M.: A simply constrained optimization reformulation of KKT systems arising from variational inequalities. Appl. Math. Optim. **40**, 19–37 (1999)
13. Fan, J., Pan, J.: Inexact Levenberg–Marquardt method for nonlinear equations. Discrete Contin. Dyn. Syst. Ser. B **4**(4), 1223–1232 (2004)
14. Fan, J.-Y., Yuan, Y.-X.: On the quadratic convergence of the Levenberg–Marquardt method. Computing **74**, 23–39 (2005)
15. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**, 47–73 (2010)
16. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**, 91–124 (2002)
17. 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)
18. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: A globally convergent LP-Newton method. SIAM J. Optim. **26**(4), 2012–2033 (2016)
19. Fischer, A., Izmailov, A.F., Jelitte, M.: Newton-type methods near critical solutions of piecewise smooth nonlinear equations. Comput. Optim. Appl. **80**, 587–615 (2021)
20. Fischer, A., Izmailov, A.F., Solodov, M.V.: Local attractors of Newton-type methods for constrained equations and complementarity problems with nonisolated solutions. J. Optim. Theory Appl. **180**, 140–169 (2019)
21. Fischer, A., Izmailov, A.F., Solodov, M.V.: Accelerating convergence of the globalized Newton method to critical solutions of nonlinear equations. Comput. Optim. Appl. **78**, 273–286 (2021)
22. Fischer, A., Izmailov, A.F., Solodov, M.V.: Unit stepsize for the Newton method close to critical solutions. Math. Program. **187**, 697–721 (2021)
23. Fischer, A., Shukla, P.K.: A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. **36**(5), 643–646 (2008)
24. Fischer, A., Shukla, P.K., Wang, M.: On the inexactness level of robust Levenberg–Marquardt methods. Optimization **59**(2), 273–287 (2010)
25. Griewank, A.: Analysis and modification of Newton's method at singularities. PhD Thesis, Australian National University, Canberra (1980)
26. Griewank, A.: Starlike domains of convergence for Newton's method at singularities. Numer. Math. **35**, 95–111 (1980)
27. Griewank, A.: On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. **27**, 537–563 (1985)
28. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**, 253–273 (1999)
29. Izmailov, A.F., Kurennoy, A.S.: On regularity conditions for complementarity problems. Comput. Optim. Appl. **57**, 667–684 (2014)
30. 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)
31. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: Critical solutions of nonlinear equations: stability issues. Math. Program. **168**, 475–507 (2018)
32. Izmailov, A.F., Solodov, M.V.: Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. Math. Program. **89**, 413–435 (2001)
33. Izmailov, A.F., Solodov, M.V.: The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. Math. Oper. Res. **27**(3), 614–635 (2002)
34. Izmailov, A.F., Solodov, M.V.: Stabilized SQP revisited. Math. Program. **122**, 93–120 (2012)

35. 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)
36. Kanzow, C.: Some equation-based methods for the nonlinear complementarity problem. Optim. Methods Softw. **3**(4), 327–340 (1994)
37. 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**(2), 375–397 (2004)
38. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. **2**(2), 164–168 (1944)
39. Marquardt, D.W.: An algorithm for least squares estimation of non-linear parameters. SIAM J. **11**(2), 431–441 (1963)
40. Mangasarian, O.L.: Equivalence of the complementarity problem to a system of nonlinear equations. SIAM J. Appl. Math. **31**, 89–92 (1976)
41. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
42. Oberlin, C., Wright, S.J.: An accelerated Newton method for equations with semismooth Jacobians and nonlinear complementarity problems. Math. Program. **117**, 355–386 (2009)
43. Polyak, B.T.: Introduction to Optimization. Optimization Software Inc, New York (1987)
44. Qi, L., Sun, J.: A nonsmooth version of Newton's method. Math. Program. **58**, 353–367 (1993)
45. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**, 253–275 (1998)
46. 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, vol. 15, pp. 239–249. Springer, Vienna (2001)
47. Zhang, J.-L.: On the convergence properties of the Levenberg–Marquardt method. Optimization **52**(6), 739–756 (2003)