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

Unit stepsize for the Newton method close to critical solutions

A. Fischer · A. F. Izmailov · M. V. Solodov

Received: January 2019 / Revised: December 2019

Abstract As is well known, when initialized close to a nonsingular solution of a smooth nonlinear equation, the Newton method converges to this solution superlinearly. Moreover, the common Armijo linesearch procedure used to globalize the process for convergence from arbitrary starting points, accepts the unit stepsize asymptotically and ensures fast local convergence. In the case of a singular and possibly even nonisolated solution, the situation is much more complicated. Local linear convergence (with asymptotic ratio of 1/2) of the Newton method can still be guaranteed under reasonable assumptions, from a starlike, asymptotically dense set around the solution. Moreover, convergence can be accelerated by extrapolation and overrelaxation techniques. However, nothing was previously known on how the Newton method can be coupled in these circumstances with a linesearch technique for globalization that locally accepts unit stepsize and guarantees linear convergence. It turns out that this is a rather nontrivial issue, requiring a delicate combination of the analyses on acceptance of the unit stepsize and on the iterates staying within the relevant starlike domain of convergence. In addition to these analyses, numerical illustrations and comparisons are presented for the Newton method and the use of extrapolation to accelerate convergence speed.

Keywords Nonlinear equation · Newton method · singular solution · critical solution · 2-regularity · linear convergence · superlinear convergence · extrapolation · linesearch.

The first author is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 409756759. The second author is supported in part by the Russian Foundation for Basic Research Grants 17-01-00125 and 19-51-12003 NNIO_a, and by Volkswagen Foundation. The third author is supported by CNPq Grant 303724/2015-3, by FAPERJ Grant E-26/202.540/2019, and by PRONEX–Optimization.

A. Fischer
Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany, E-mail: andreas.fischer@tu-dresden.de

A. F. Izmailov
VMK Faculty, OR Department, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory, 119991 Moscow, Russia; and Peoples' Friendship University of Russia, Miklukho-Maklaya Str. 6, 117198 Moscow, Russia. E-mail: izmaf@ccas.ru

M. V. Solodov
IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. E-mail: solodov@impa.br

Mathematics Subject Classification (2000) 65H10.

1 Introduction

Consider a system of nonlinear equations

$$\Phi(u) = 0, \tag{1}$$

where $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ is sufficiently smooth. For a current iterate $u^k \in \mathbb{R}^p$, the basic Newton method solves the linear system

$$\Phi(u^k) + \Phi'(u^k)v = 0 \tag{2}$$

to obtain v^k, and defines the next iterate as $u^{k+1} = u^k + v^k$.

In this work, we are interested in the case when (1) has a singular, possibly even nonisolated, solution $\bar{u}$. This means that $\Phi'(\bar{u})$ is a singular matrix. Evidently, one cannot guarantee convergence of the Newton method to a singular solution $\bar{u}$ from all starting points in its entire neighborhood, as the Newton iteration need not be well-defined at some points arbitrarily close to such $\bar{u}$. However, well-definedness and convergence can still be shown from some large domains, and under reasonable assumptions. To discuss the relevant results and put our contributions in perspective, we need to recall some definitions first.

A set $U \subset \mathbb{R}^p$ is called starlike with resect to $\bar{u} \in \mathbb{R}^p$ if, for every $u \in U$ and $t \in (0, 1]$, it holds that $tu + (1-t)\bar{u} \in U$. With $\mathscr{S}$ standing for the unit sphere in $\mathbb{R}^p$, a direction $v \in \mathscr{S}$ is called excluded for such a starlike set U if $\bar{u} + tv \notin U$ for all $t > 0$. A set which is starlike with respect to a given point is called asymptotically dense if the corresponding set of excluded directions is thin, i.e., the complement of the latter is open and dense in $\mathscr{S}$ (with topology induced from $\mathbb{R}^p$).

Assuming that Φ is twice differentiable at $\bar{u}$, we say that it is 2-regular at $\bar{u}$ in the direction $v \in \mathbb{R}^p$ if the matrix $\Phi'(\bar{u}) + \Pi\Phi''(\bar{u})[v]$ is nonsingular, where Π is the orthogonal projector in $\mathbb{R}^p$ onto $(\operatorname{im}\Phi'(\bar{u}))^{\perp}$, and $\Phi''(\bar{u})[v]$ is a matrix of the linear operator defined by $(\Phi''(\bar{u})[v])u = \Phi''(\bar{u})[v, u]$, $u \in \mathbb{R}^p$. This is one of the possible definitions of 2-regularity; an equivalent characterization will be introduced and employed in Section 2. Of course, if $\Phi'(\bar{u})$ is nonsingular, then Φ is 2-regular at $\bar{u}$ in every direction v, including $v = 0$. What is important here is that 2-regularity may hold naturally at singular (and even nonisolated) solutions in nonzero directions and, among them, in directions from $\ker\Phi'(\bar{u})$ (see Examples 2–4 below). This will be the key assumption in our analysis.

Next, we recall that the notion of a noncritical solution of a nonlinear equation, introduced in [10], consists of the following two ingredients: Clarke-regularity of the solution set at $\bar{u}$, and the contingent cone to the solution set at $\bar{u}$ being equal to $\ker\Phi'(\bar{u})$. As demonstrated in [9, 10], noncriticality is equivalent to the local Lipschitzian error bound, and if the key assumption of 2-regularity of Φ at the solution $\bar{u}$ in a direction $v \in \ker\Phi'(\bar{u}) \setminus \{0\}$ holds, then the second ingredient in the definition of noncriticality of $\bar{u}$ is necessarily violated. Thus, the 2-regularity assumption can be expected to hold only at those singular solutions which are critical. Note that such solutions can be stable under a rich class of perturbations [10].

In [7], convergence of the Newton iterates to $\bar{u}$ has been established under the 2-regularity assumption specified above. Under this assumption, convergence holds from all starting points in a set in $\mathbb{R}^p$, which is starlike with respect to $\bar{u}$ and asymptotically dense. It is important to emphasize that this analysis is applicable to solutions which are not isolated. As is shown in [7], the rate of convergence is linear, with the asymptotic ratio of $1/2$, and it cannot be any faster (again, unless $\bar{u}$ is a nonsingular solution). Moreover, there is a rather special convergence pattern specified in detail in [8, Theorem 2.1]. In particular, in addition to the above-mentioned linear convergence rate with ratio of $1/2$, it is established that convergence is along a single direction $v \in \ker \Phi'(\bar{u})$, i.e., the sequence $\{(u^k - \bar{u})/\|u^k - \bar{u}\|\}$ converges to such v. In [16], these results were further extended to the case when Φ has a Lipschitz-continuous first derivative (but is not necessarily twice differentiable). They were also (partially) generalized to various modifications of the basic Newton scheme in [9], and to constrained equations in [4].

At this point, it should be commented that faster local convergence (i.e., superlinear) in the singular/degenerate cases can be guaranteed for some Newton-type methods when they are initialized close to noncritical solutions. These methods modify the basic Newton schemes by incorporating appropriate stabilizing mechanisms. The examples are the classical Levenberg–Marquardt method [14, 15] with an adaptive selection of the regularization parameter [5, 18], the stabilized Newton–Lagrange method (stabilized sequential quadratic programming, when (1) corresponds to the Lagrange optimality system for an equality-constrained optimization problem) [3, 11, 17], and the LP-Newton method [1, 2]. In these references, it is demonstrated that all those methods possess very similar and very strong local convergence properties when initialized close enough to a noncritical solution: in this case, they converge superlinearly or even quadratically to a nearby solution. However, the results in [9] highlight that these nice convergence properties may not show up in computation, if critical solutions exist. This is because critical solutions may be specially attractive for Newtonian sequences initialized within large sets of starting points, even though critical solutions typically form only a thin subset within the whole solution set (this is normally the case, unless a singular solution is isolated). In particular, the attraction phenomenon of critical solutions may not allow the cited methods to enter a sufficiently small neighborhood of any noncritical one, from where superlinear convergence occurs.

Thus, in this work, we follow a different path: instead of trying to avoid convergence to critical solutions, the idea is to exploit the special convergence pattern of the Newton method to critical solution, established in [8, Theorem 2.1], in order to accelerate convergence, and to achieve at least a relatively fast linear rate. Some accelerating techniques are known, like extrapolation [6, 8], which decreases the asymptotic ratio of linear convergence from 1/2 to 1/4, 1/8, etc., and is easy to integrate into globalization schemes. Theoretically clean integration of other existing acceleration techniques seems way more difficult. In any case, preserving accelerating properties of these techniques becomes not a simple issue if linesearch is used for globalization: the ultimate acceptance of the unit stepsize of the Newton method becomes a key question, as this is not at all automatic (in contrast to the case of a nonsingular solution). This question turns out to be highly nontrivial and will be the central issue addressed in this work.

It is also quite possible that other approaches for local acceleration might be developed in the future, but they would all face this principle issue: when combining these essentially local acceleration techniques with globalized versions of the Newton method, the latter must preserve the convergence pattern of the pure Newton method, for the overall algorithm to be well-defined and (potentially) accelerated. Therefore, our work is a basic contribution in this direction, i.e., of consciously exploiting the attraction of Newtonian methods to critical solutions.

The rest of the paper is organized as follows. In Section 2, we summarize the results from [7], with some refinements required for our purposes. Our linesearch globalization is dealt with in Section 3, together with the acceleration by extrapolation. Furthermore, we exhibit that, unlike in the case of convergence to a nonsingular solution, the question of acceptance of the unit stepsize by linesearch is highly nontrivial in the singular case. Section 4 provides the analysis showing when the unit stepsize is accepted. We conclude with numerical illustrations and comparisons in Section 5.

2 Preliminaries

After some basic definitions, we provide the key ingredients of the analysis in [7], with some refinements needed for our subsequent developments.

For a symmetric bilinear mapping $B: \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$, and for a given $v \in \mathbb{R}^p$, we denote by $B[v]$ a linear operator from $\mathbb{R}^p$ to $\mathbb{R}^p$ defined by $B[v]u = B[v, u]$, $u \in \mathbb{R}^p$. Assuming that Φ is twice differentiable at $\bar{u}$, define the linear operator $\mathscr{B}(v) : \ker \Phi'(\bar{u}) \to (\operatorname{im} \Phi'(\bar{u}))^\perp$ as the restriction of $\Pi \Phi''(\bar{u})[v]$ to $\ker \Phi'(\bar{u})$. Then 2-regularity of Φ at $\bar{u}$ in a direction v is equivalent to saying that $\mathscr{B}(v)$ is nonsingular. For a given point $\bar{u} \in \mathbb{R}^p$, a given direction $\bar{v} \in \mathscr{S}$, and scalars $\varepsilon > 0$ and $\delta > 0$, define the set

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

Here and throughout the paper, all the norms are Euclidean (i.e., consistent with the Euclidean inner product, denoted by $\langle \cdot, \cdot \rangle$). Observe further that every $u \in \mathbb{R}^p$ can be written as $u = u_1 + u_2$ with uniquely defined $u_1 \in (\ker \Phi'(\bar{u}))^\perp$ and $u_2 \in \ker \Phi'(\bar{u})$.

The following result can be considered as a sharpened version of [7, Lemma 4.1]; see also [9, Lemma 1].

Lemma 1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be twice differentiable near $\bar{u} \in \mathbb{R}^p$, with its second derivative Lipschitz-continuous with respect to $\bar{u}$, that is,*

$$\Phi''(u) - \Phi''(\bar{u}) = O(\|u - \bar{u}\|)$$

as $u \to \bar{u}$. Let $\bar{u}$ be a solution of equation (1)*, and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \mathscr{S}$. Then, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that, for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, the linear operator $\mathscr{B}(u^k - \bar{u})$ is invertible,*

$$(\mathscr{B}(u^k - \bar{u}))^{-1} = O(\|u^k - \bar{u}\|^{-1}) \tag{3}$$

as $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$ tends to $\bar{u}$, the Newton iteration system has a unique solution v^k, and it holds that

$$u_1^k + v_1^k - \bar{u}_1 = O(\|\Pi \Phi''(\bar{u})[u^k - \bar{u}, u_1^k - \bar{u}_1]\|) + O(\|u^k - \bar{u}\|^3), \tag{4}$$

$$u_2^k + v_2^k - \bar{u}_2 = \frac{1}{2}\pi(u^k - \bar{u}) + O(\|u^k - \bar{u}\|^2), \tag{5}$$

where for $v \in \mathbb{R}^p$ we set

$$\pi(v) = v_2 + (\mathscr{B}(v))^{-1} \Pi \Phi''(\bar{u})[v, v_1]. \tag{6}$$

Proof As in the proof in [9, Lemma 1], without loss of generality we can assume that $\bar{u} = 0$ and

$$\Phi(u) = Au + \frac{1}{2}B[u, u] + R(u), \tag{7}$$

where $A = \Phi'(0) \in \mathbb{R}^{p\times p}$, $B = \Phi''(0)$ is a symmetric bilinear mapping from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}^p$, and the mapping $R : \mathbb{R}^p \to \mathbb{R}^p$ is differentiable near 0, with

$$R(u) = O(\|u\|^3), \quad R'(u) = O(\|u\|^2)$$

as $u \to 0$.

Directly following the proof in [9, Lemma 1] we obtain the existence of $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ such that for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$ there exists the unique v^k solving (2), and

$$u_1^k + v_1^k = O(\|u - \bar{u}\|^2) \tag{8}$$

holds, as well as (5), as $u \to \bar{u}$. Therefore, it only remains to show that (8) can be replaced by the sharper estimate (4).

Multiplying (2) by $(I - \Pi)$, and employing (7), we obtain that

$$\begin{aligned}(A + (I-\Pi)(B[u^k] + R'(u^k)))v_1^k &= -Au_1^k - (I-\Pi)\left(\frac{1}{2}B[u^k, u^k] + R(u^k)\right)\\ &\quad -(I-\Pi)(B[u^k] + R'(u^k))v_2^k\\ &= -Au_1^k - (I-\Pi)\left(B\left[u^k, \frac{1}{2}u^k + v_2^k\right]\right.\\ &\quad \left. + R(u^k) + R'(u^k)v_2^k\right). \end{aligned} \tag{9}$$

According to (3), (5), and (6), it follows that

$$\begin{aligned} B\left[u^k, \frac{1}{2}u^k + v_2^k\right] &= B\left[u^k, \frac{1}{2}u_2^k + v_2^k\right] + O(\|\Pi B[u^k, u_1^k]\|)\\ &= O(\|\Pi B[u^k, u_1^k]\|) + O(\|u^k\|^3). \end{aligned}$$

But then, repeating the argument in the proof in [9, Lemma 1], we obtain from (9) that (4) holds. □

Compared to the smoothness of Φ in Lemma 1, a weaker smoothness assumption is used in [16, Assumption 1]. However, to get the refined results in Lemma 1 above, it seems impossible to relax our smoothness assumptions on Φ.

The next result is a counterpart of [7, Lemma 5.1]; see also [6, Lemma 2.2] and [9, Theorem 1 and Remark 1]. It characterizes convergence from points in the set $K_{\varepsilon,\delta}(\bar{v})$ with sufficiently small $\varepsilon > 0$ and $\delta > 0$, where $\bar{v}$ is from the key 2-regularity assumption.

Proposition 1 *In addition to the assumptions of Lemma 1, let $\bar{v} \in \ker \Phi'(\bar{u})$. Then, for every $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that, for every starting point $u^0 \in K_{\varepsilon,\delta}(\bar{v})$, the unique sequence $\{u^k\} \subset \mathbb{R}^p$ exists such that $v^k = u^{k+1} - u^k$ solves (2) for each k, and this sequence possesses the following properties: $\{u^k\} \subset K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, $\{u^k\}$ converges to $\bar{u}$,*

$$\lim_{k\to\infty} \frac{\|u^{k+1} - \bar{u}\|}{\|u^k - \bar{u}\|} = \frac{1}{2}, \tag{10}$$

and the sequence $\{(u^k - \bar{u})/\|u^k - \bar{u}\|\}$ converges to some $v \in \ker \Phi'(\bar{u})$.

The key assumption needed for applicability of Proposition 1 consists of the existence of $\bar{v} \in \ker \Phi'(\bar{u})$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. This assumption will be playing the central role in the subsequent analysis as well. We refer the reader to [13, Section 1.3.4] for some analysis of how wide are the classes of mappings with 2-regularity properties. In particular, it is demonstrated there that the set of pairs (Q, v) such that a quadratic mapping $Q : \mathbb{R}^p \to \mathbb{R}^p$ is 2-regular at 0 in a direction $v \in \mathbb{R}^p$ is open and dense in the normed linear space of such pairs.

3 Peculiarities of linesearch near singular solutions

A standard approach to achieving global convergence of Newtonian methods relies on linesearch with a suitable residual function. To this end, we shall employ the function $\varphi : \mathbb{R}^p \to [0, +\infty)$ with

$$\varphi(u) = \|\Phi(u)\|.$$

Observe that if $\Phi(u^k) \neq 0$ for some iterate u^k, then φ is differentiable at u^k with

$$\varphi'(u^k) = \|\Phi(u^k)\|^{-1}(\Phi'(u^k))^\top \Phi(u^k).$$

Therefore, for v^k defined by (2), it holds that

$$\langle \varphi'(u^k), v^k \rangle = -\|\Phi(u^k)\| < 0, \tag{11}$$

and hence, v^k is a direction of descent of φ at u^k.

Based on the residual function φ, we first state the following model linesearch globalization scheme.

Algorithm 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 (2).
3. Set $\alpha = 1$. If the inequality

$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \sigma\alpha)\|\Phi(u^k)\| \tag{12}$$

   is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (12) again, etc., until (12) becomes valid.
4. Set $u^{k+1} = u^k + \alpha_k v^k$.
5. Increase k by 1 and go to step 1.

Instead of (12), one might also consider the Armijo linesearch rule for the squared residual $\psi : \mathbb{R}^p \to [0, +\infty)$,

$$\psi(u) = \|\Phi(u)\|^2,$$

which would mean replacing (12) by the inequality

$$\psi(u^k + \alpha v^k) \leq \psi(u^k) + \sigma\alpha\langle \psi'(u^k), v^k \rangle.$$

Since

$$\psi'(u^k) = 2(\Phi'(u^k))^\top \Phi(u^k),$$

for v^k defined by (2) we then have that the Armijo inequality above takes the form

$$\|\Phi(u^k + \alpha v^k)\|^2 \leq (1 - 2\sigma\alpha)\|\Phi(u^k)\|^2. \tag{13}$$

The two rules defined by (12) and (13) are closely related (see [12, Section 5.1]), but they are not the same: in general, (12) allows for larger stepsizes than (13). For instance, in Example 1, (13) is satisfied with $\alpha = 1$ only if $\sigma \leq 15/32$.

Note that we do not discuss what can be done in cases where the Newton equation (2) has no solution, or when the Newton direction v^k exists but provides a poor descent. The inclusion of corresponding remedies, like safeguarding techniques, would shift the focus of this paper away from our main aim. Thus, global convergence of Algorithm 1 is not an issue here. Rather, in Section 4, we will concentrate on the development of conditions that, close to a singular solution, guarantee the well-definedness of the Newton direction and its acceptance by the linesearch in Step 3 of Algorithm 1 with unit stepsize.

We next discuss an extrapolation technique, as a tool for accelerating the local rate of linear convergence. This underlines that the unit stepsize is important to have. Moreover, to prepare the corresponding developments in Section 4, we provide examples demonstrating that obtaining a unit stepsize is not at all automatic as it is for nonsingular solutions.

Extrapolation for accelerating the Newton method in the case of singular solutions is naturally suggested by the convergence pattern in Proposition 1, and is well-known from [8]. In its simplest form, it consists of generating an auxiliary sequence $\{\widehat{u}^k\}$ obtained by doubling the Newton step, i.e.,

$$\widehat{u}^{k+1} = u^k + 2v^k. \tag{14}$$

Of course, this would not help to achieve the superlinear convergence rate (of the nonsingular case). However, it follows from [8, Theorem 4.1] that under the key 2-regularity assumption above, $\{\widehat{u}^k\}$ converges linearly with the asymptotic ratio of $1/4$ (instead of $1/2$ for $\{u^k\}$), from all points in the domain of convergence of Newtonian sequences $\{u^k\}$, which are themselves not affected by this acceleration technique in any way. Moreover, by increasing the "depth" of extrapolation, the asymptotic ratio of the linear convergence of the corresponding auxiliary sequences can be made arbitrarily small. Algorithm 1 can be easily modified to incorporate extrapolation: the latter is not concerned with any serious computational overhead (just one extra evaluation of Φ is needed to assess if the obtained $\widehat{u}^{k+1}$ satisfies the stopping criterion), and does not affect the main iteration sequence.

The numerical experiments in Section 5 show that already the simple extrapolation of "depth" 1 based on formula (14) may lead to significant improvements, provided that the unit stepsize is ultimately accepted.

Yet another existing acceleration technique is 2- or 3-step overrelaxation [8], but we will not discuss any details in the current paper, as our focus here is on principal difficulties to be encountered by any acceleration technique (whether existing or expected in the future) when it comes to globalization of convergence.

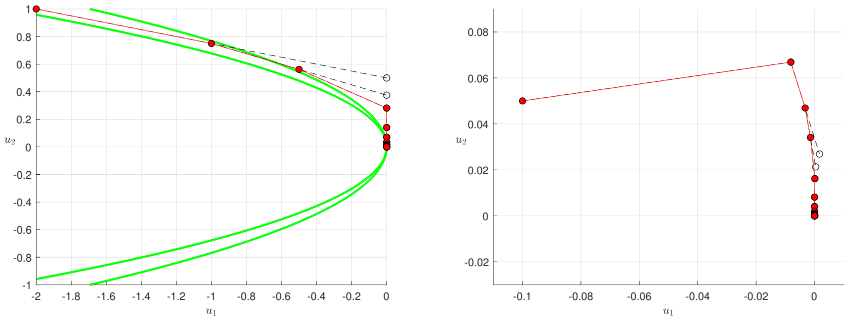
In any case, the key question for establishing properties of Algorithm 1 on local convergence and its rate, even together with some acceleration technique, is the ultimate acceptance of the unit stepsize. Convergence rate and overall success of extrapolation and overrelaxation depends on following the convergence pattern of the basic Newton method, established in [6, Lemma 3.2] and [8, Theorem 2.1] (and summarized in the statement of Proposition 1). As we demonstrate now by examples, this question is highly nontrivial, unlike in cases of convergence to a nonsingular solution (for the latter, see, e.g., [12, Theorem 5.4]), and requires special attention.

We start with a very simple example showing that one should further restrict the values of σ in (12), in order to expect acceptance of the unit stepsize in this setting.

Example 1 Let $p = 1$, and $\Phi(u) = u^2$. The unique solution of (1) with this Φ is $\bar{u} = 0$, where $\Phi'(\bar{u}) = 0$, and Φ is 2-regular at $\bar{u}$ in every nonzero direction v.

By direct computation we obtain that for every $u^k \neq 0$, the unique solution of (2) is $v^k = -u^k/2$. Therefore, $u^k + v^k = u^k/2$, $\Phi(u^k + v^k) = (u^k)^2/4$, and it is now evident that (12) holds with $\alpha = 1$ if and only if $\sigma \leq 3/4$.

Let us mention that this restriction on σ would fully agree with the theoretical justification in Section 4.



(a) Example 2 (b) Example 3

Fig. 1: Iterative sequences.

Example 2 Let $p = 2$, $\Phi(u) = (u_1 + au_2^2/2,\, u_2^2/2)$, where $a \in \mathbb{R}$ is a parameter. In this example, and in all the examples below, we use subscripts "1" and "2" to denote components of vectors. This notation usually agrees with our use of these subscripts in decompositions of vectors into the sum of orthogonal components. The unique solution of (1) with this Φ is $\bar{u} = 0$, $\ker \Phi'(\bar{u}) = \{v \in \mathbb{R}^2 \mid v_1 = 0\}$, and Φ is 2-regular at $\bar{u}$ in every direction $v \in \mathbb{R}^2$ with $v_2 \neq 0$, and in particular, in every nonzero direction in $\ker \Phi'(\bar{u})$.

For simplicity, the iteration index k is omitted within this example. By direct computation we obtain that, for every $u \in \mathbb{R}^2$ with $u_2 \neq 0$, the unique solution of (2) with $u^k = u$ is $v_u = (-u_1, -u_2/2)$. Therefore, $u + v_u = (0, u_2/2)$, $\Phi(u + v_u) = (au_2^2/8,\, u_2^2/8)$, and hence,

$$\|\Phi(u + v_u)\| = \frac{1}{8}\sqrt{a^2 + 1}u_2^2.$$

Then, it can be seen that the set

$$W = \{u \in \mathbb{R}^2 \mid u_2 \neq 0,\ \|\Phi(u + v_u)\| > (1 - \sigma)\|\Phi(u)\|\}$$

of points in $\mathbb{R}^2$ such that the stepsize condition (12) does not hold with $\alpha = 1$ is nonempty if and only if $a^2 - 16(1 - \sigma)^2 + 1 > 0$. In this case, W consists of u satisfying

$$a_- u_2^2 < u_1 < a_+ u_2^2, \tag{15}$$

where

$$a_\pm = -\frac{a}{2} \pm \frac{\sqrt{a^2 - 16(1 - \sigma)^2 + 1}}{8(1 - \sigma)}.$$

The set W is shown in Figure 1a (for $a = \sqrt{15}$ and $\sigma = 0.1$) as the area between the two parabolas. Observe that, when nonempty, W is asymptotically thin, i.e., its complement contains a set which is starlike with respect to $\bar{u}$ and asymptotically dense. However, note that W contains points arbitrarily close to $\bar{u}$. Hence, acceptance of the unit stepsize fails at those points.

What is even worse is that the set W has a nonempty intersection with $K_{\varepsilon,\delta}(\bar{v})$ for $\bar{v} = (0, \pm 1)$ spanning $\ker \Phi'(\bar{u})$, for any choices of $\varepsilon > 0$ and $\delta > 0$. Recall that according to Proposition 1, the sequences of the Newton method are well-defined when initialized within $K_{\varepsilon,\delta}(\bar{v})$ with $\varepsilon > 0$ and $\delta > 0$ small enough, and moreover, the latter can be chosen in such a way that these sequences do not leave $K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$ with any pre-fixed $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$. Unfortunately, acceptance of the unit stepsize along such sequences cannot be guaranteed.

Another important observation is that $a_+ < 0$ holds if and only if $\sigma \in (0, 3/4)$, which fully agrees with part b) of Lemma 2 below. If $a_+ > 0$, then the set W contains $\ker \Phi'(\bar{u}) \setminus \{0\}$ in its interior, and in view of the last assertion of Proposition 1, even "typical" acceptance of the unit stepsize might be problematic in this case, even though the set in question is asymptotically thin.

Figure 1a also shows one particular sequence generated by the damped Newton method in Algorithm 1. Solid lines always show a single step. The starting point $u^0 = (-2, 1)$ and the following iterate u^1 lie in the area between the two parabolas, meaning that the step length is less than 1. The dashed lines show where the corresponding full Newton steps would end. All further steps of the algorithm are Newton steps with unit stepsize.

The next modification of Example 2 demonstrates that, unlike in that example, even when the full Newton step is accepted, there is no guarantee in general that it will be accepted at the next iteration.

Example 3 Let $p = 2$, $\Phi(u) = (u_1 + \sqrt{15}u_2^2/2,\, u_1u_2 + u_2^2/2)$. With this Φ, (1) has two solutions; the one of interest is $\bar{u} = 0$, where $\ker \Phi'(\bar{u}) = \{v \in \mathbb{R}^2 \mid v_1 = 0\}$, and Φ is 2-regular at $\bar{u}$ in every direction $v \in \mathbb{R}^2$ with $v_1 + v_2 \neq 0$, and in particular, in every nonzero direction in $\ker \Phi'(\bar{u})$.

Figure 1b shows some quite typical run of the Newton method, with the same meaning of dashed and solid lines as in Figure 1a. One can see that the method enters, after one step, the area where the full Newton step is not accepted, but then leaves this area after two halved steps.

The examples above put in evidence that acceptance of the unit stepsize by Algorithm 1 is not at all automatic, even when the Newton direction is well defined, and even arbitrarily close to the solution. Nevertheless, in the next section, we demonstrate that under the assumptions of Proposition 1, there exists a set which is starlike with respect to $\bar{u}$ and asymptotically dense, and such that when initialized within this set, Algorithm 1 is well-defined, converges to $\bar{u}$, and ultimately accepts the unit stepsize.

4 Asymptotic acceptance of the unit stepsize

The next result might be regarded quite expected, in view of Example 2 (compare (15) with (16) and (18) below). Nevertheless, the argument involves a rather subtle interplay between first- and second-order terms of Φ.

Lemma 2 *Under the assumptions of Proposition 1, the following assertions hold:*

(a) For every $\sigma \in (0, 1)$*, there exist* $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$*,* $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$*, and* $\bar{\Gamma} > 0$*, such that for every* $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$ *satisfying*

$$\|u_1^k - \bar{u}_1\| \geq \bar{\Gamma} \|u_2^k - \bar{u}_2\|^2, \tag{16}$$

there exists a unique solution v^k *of the Newton equation* (2)*, and it holds that*

$$\|\Phi(u^k + v^k)\| \leq (1 - \sigma)\|\Phi(u^k)\|. \tag{17}$$

(b) For every $\sigma \in (0, 3/4)$*, there exist* $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$*,* $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$*, and* $\bar{\gamma} = \bar{\gamma}(\bar{v}) > 0$*, such that for every* $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$ *satisfying*

$$\|u_1^k - \bar{u}_1\| \leq \bar{\gamma} \|u_2^k - \bar{u}_2\|^2, \tag{18}$$

there exists a unique solution v^k *of the Newton equation* (2)*, and* (17) *holds.*

Proof To simplify the presentation, assume without loss of generality that $\bar{u} = 0$. Let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0, 1)$ be chosen according to Lemma 1. Then, for every $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$, there exists the unique v^k solving (2), and this v^k satisfies (3)–(6). Furthermore, condition $\bar{v} \in \ker \Phi'(0)$ means that $\bar{v}_1 = 0$, and hence,

$$\frac{\|u_1^k\|}{\|u^k\|} = \left\| \frac{u_1^k}{\|u^k\|} - \bar{v}_1 \right\| \leq \left\| \frac{u^k}{\|u^k\|} - \bar{v} \right\| \leq \bar{\delta},$$

implying that

$$\|u_1^k\| \leq \bar{\delta}\|u^k\| \leq \bar{\delta}(\|u_1^k\| + \|u_2^k\|).$$

Therefore,

$$\|u_1^k\| \leq \frac{\bar{\delta}}{1 - \bar{\delta}} \|u_2^k\|. \tag{19}$$

Combined with (3)–(6), this further implies that

$$u^k = O(\|u_2^k\|), \quad v^k = O(\|u_2^k\|) \tag{20}$$

as $u^k \to 0$.

Furthermore, since v^k solves (2), employing Taylor's formula, (3)–(6), (19), and (20), we obtain that

$$\begin{aligned}
\Phi(u^k + v^k) &= \Phi(u^k) + \Phi'(u^k)v^k + \frac{1}{2}\Phi''(u^k)[v^k, v^k] + O(\|v^k\|^3) \\
&= \frac{1}{2}\Phi''(u^k)[v^k, v^k] + O(\|v^k\|^3) \\
&= \frac{1}{2}\Phi''(0)[v^k, v^k] + O(\|u^k\|\|v^k\|^2) + O(\|v^k\|^3) \\
&= \frac{1}{8}\Phi''(0)[u_2^k, u_2^k] + O(\|u_1^k\|\|u_2^k\|) + O(\|u_1^k\|^2) + O(\|u_2^k\|^3) \\
&= \frac{1}{8}\Phi''(0)[u_2^k, u_2^k] + O(\bar{\delta}\|u_2^k\|^2) + O(\|u_2^k\|^3),
\end{aligned} \tag{21}$$

and

$$\begin{aligned}\Phi(u^k) &= \Phi'(0)u^k + \frac{1}{2}\Phi''(0)[u^k, u^k] + O(\|u^k\|^3)\\ &= \Phi'(0)u_1^k + \frac{1}{2}\Phi''(0)[u_2^k, u_2^k] + O(\|u_1^k\|\|u_2^k\|) + O(\|u_1^k\|^2) + O(\|u_2^k\|^3)\\ &= \Phi'(0)u_1^k + \frac{1}{2}\Phi''(0)[u_2^k, u_2^k] + O(\bar{\delta}\|u_2^k\|^2) + O(\|u_2^k\|^3). \end{aligned} \tag{22}$$

We first prove item (a). Suppose that (16) is satisfied. Let $\bar{\nu} > 0$ be the smallest nonzero singular value of $\Phi'(0)$. Then, from (22), we have that

$$\|\Phi(u^k)\| \geq \bar{\nu}\bar{\Gamma}\|u_2^k\|^2 + O(\|u_2^k\|^2) \geq \nu\bar{\Gamma}\|u_2^k\|^2$$

for every pre-fixed $\nu \in (0, \bar{\nu})$, provided $\bar{\varepsilon} > 0$ is small enough while $\bar{\Gamma} > 0$ is large enough. On the other hand, from (21) it follows that

$$\Phi(u^k + v^k) = O(\|u_2^k\|^2).$$

Hence, (17) holds provided $\bar{\varepsilon} > 0$ is small enough while $\bar{\Gamma} > 0$ is large enough.

We now proceed with the proof of item (b). Observe first that 2-regularity of Φ at 0 in the direction $\bar{v} \in \ker \Phi'(0) \setminus \{0\}$ implies that

$$\Phi''(0)[\bar{v}, \bar{v}] \notin \operatorname{im} \Phi'(0),$$

and hence,

$$\Pi\Phi''(0)[\bar{v}, \bar{v}] \neq 0. \tag{23}$$

Furthermore, since $\bar{v}_1 = 0$,

$$\left\| \frac{u_2^k}{\|u^k\|} - \bar{v} \right\| = \left\| \frac{u_2^k}{\|u^k\|} - \bar{v}_2 \right\| \leq \left\| \frac{u^k}{\|u^k\|} - \bar{v} \right\| \leq \bar{\delta}.$$

Taking into account that $\|\bar{v}\| = 1$, the latter further implies that $u_2^k/\|u_2^k\| \to \bar{v}$ as $\bar{\delta} \to 0$. Therefore, by (23), there exists $\gamma = \gamma(\bar{v}) > 0$ such that

$$\|\Phi''(0)[u_2^k, u_2^k]\| \geq \|\Pi\Phi''(0)[u_2^k, u_2^k]\| \geq \gamma\|u_2^k\|^2, \tag{24}$$

provided $\bar{\delta} > 0$ is small enough.

Next, fix any

$$\nu \in \left(0, \frac{3 - 4\sigma}{8(2 - \sigma)}\right) \tag{25}$$

(recall that within (b), $\sigma \in (0, 3/4)$). By further decreasing $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ if necessary, we obtain from (21) and (24) that

$$\|\Phi(u^k + v^k)\| \leq \left(\frac{1}{8} + \nu\right) \|\Phi''(0)[u_2^k, u_2^k]\|. \tag{26}$$

Suppose now that (18) is satisfied. Then, again employing (24), from (22) we obtain that

$$\|\Phi(u^k)\| \geq \left(\frac{1}{2} - \nu\right) \|\Phi''(0)[u_2^k, u_2^k]\| \tag{27}$$

provided $\bar{\varepsilon} > 0$, $\bar{\delta} > 0$ and $\bar{\gamma} > 0$ are small enough.

Combining (26) and (27), and employing (25), we now conclude that (17) is satisfied. This completes the proof. □

Note that Example 2 demonstrates, in particular, that assertion (b) of Lemma 2 does not hold for $\sigma \geq 3/4$.

The next example shows that the requirement of 2-regularity is essential in Lemma 2. Observe that in the case of full singularity, i.e. when $\Phi'(0) = 0$, condition (18) holds automatically (with any $\bar{\gamma} > 0$) for every $u^k \in \mathbb{R}^p$, since $u_1^k = 0$.

Example 4 Let $p = 2$, $\Phi(u) = (u_1^2 + u_2^q, u_1 u_2)$, where $q \geq 3$ is an integer. Then, the unique solution of (1) is $\bar{u} = 0$, $\Phi'(0) = 0$, and Φ is 2-regular at 0 in every direction $\bar{v} \in \mathbb{R}^2$ with $\bar{v}_1 \neq 0$. The iteration system (2) of the Newton method takes the form

$$(u_1^k)^2 + (u_2^k)^q + 2u_1^k v_1 + q(u_2^k)^{q-1} v_2 = 0, \quad u_1^k u_2^k + u_2^k v_1 + u_1^k v_2 = 0. \tag{28}$$

Suppose that $u_1^k = 0$, $u_2^k \neq 0$, which implies that u^k does not need to belong to any domain of acceptance of the full step as specified in Lemma 2, since Φ is not 2-regular at 0 in this direction u^k. Then, (28) is uniquely solved by $v_1^k = 0$, $v_2^k = -u_2^k/q$, and hence,

$$\Phi(u^k + v^k) = \left(\left(1 - \frac{1}{q}\right)^q (u_2^k)^q, 0 \right), \quad \Phi(u^k) = ((u_2^k)^q, 0).$$

Therefore, (17) holds if and only if

$$\left(1 - \frac{1}{q}\right)^q \leq 1 - \sigma,$$

implying

$$\sigma \leq 1 - \left(1 - \frac{1}{q}\right)^q \leq 1 - \left(1 - \frac{1}{3}\right)^3 < \frac{3}{4}$$

(recall that $q \geq 3$).

Moreover, the Jacobian $\Phi'(u)$ is singular at points $u \in \mathbb{R}^2$ satisfying $2u_1^2 - qu_2^q = 0$. Consider now Newton steps computed at points of a sequence $\{u^k\}$ approaching 0 and staying close to the curve of singular points: let $u_1^k = at_k^{q/2}$, $u_2^k = t_k > 0$, where the scalar a is taken close to $\sqrt{q/2}$. By similar considerations as above, it can be seen that

$$\|\Phi(u^k + v^k)\| = |(1 - A)A| a t_k^{(q+2)/2} + o(t_k^{(q+2)/2}), \quad \|\Phi(u^k)\| = a t_k^{(q+2)/2} + o(t_k^{(q+2)/2})$$

as $t_k \to 0$, where

$$A = A(a) = \frac{q - 1 - a^2}{q - 2a^2}.$$

Therefore, (17) may hold for small $t_k > 0$ only provided $|(1 - A)A| < 1$, which is not the case for a close to $\sqrt{q/2}$.

According to Lemma 2, the full Newton step is guaranteed to be accepted by linesearch (12) for any iterate $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}$ outside of the set

$$\mathscr{W} = \{u \in \mathbb{R}^p \mid \bar{\gamma} \|u_2 - \bar{u}_2\|^2 < \|u_1 - \bar{u}_1\| < \bar{\Gamma} \|u_2 - \bar{u}_2\|^2\}. \tag{29}$$

Inside of this set, we have no guarantee that the full step is accepted; we shall thus refer to $\mathscr{W}$ as the "troublesome" set. Note that the set $\mathscr{W}$ is asymptotically thin: its complement contains a subset defined by the inequality $\|u_1 - \bar{u}_1\| \geq \bar{\Gamma} \|u_2 - \bar{u}_2\|^2$, which is starlike with respect to $\bar{u}$ and asymptotically dense, with the only excluded directions being those in $\ker \Phi'(\bar{u})$ (unless $\Phi'(\bar{u}) = 0$, in which case $\mathscr{W} = \emptyset$); recall that $u - \bar{u} \in \ker \Phi'(\bar{u})$ implies

$u_1 - \bar{u}_1 = 0$. Moreover, the intersection of this subset with $K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$ is asymptotically dense within $K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$. Nevertheless, as long as $\mathscr{W}$ contains points arbitrarily close to $\bar{u}$, there is always a risk that the full Newton step will not be accepted on some iteration, no matter how close u^k is to $\bar{u}$. Therefore, we proceed with the investigation of how the iterates can (or better cannot) enter $\mathscr{W}$. The key role in these considerations will be played by inequality (18), as we shall show that, in our setting, it always holds after a finite number of iterations. This, in turn, means that the iterates eventually leave the troublesome set $\mathscr{W}$. Nevertheless, assertion (a) of Lemma 2 will be needed in Lemma 6 and in Remark 1 below.

Lemma 3 *Under the assumptions of Proposition 1, for every $\bar{\gamma} > 0$, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that, for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, the Newton equation* (2) *has a unique solution v^k, and it holds that*

$$\|u_1^k + v_1^k - \bar{u}_1\| \leq \bar{\gamma}\|u_2^k + v_2^k - \bar{u}_2\|^2. \tag{30}$$

Proof Assume again for simplicity that $\bar{u} = 0$, and let $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0,\,1)$ be first chosen according to Lemma 1. Then, for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, there exists the unique v^k solving (2), this v^k satisfies (3)–(6), and relations (19)–(20) hold. Therefore, we get

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

and

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

which evidently implies (30) provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. □

Combining Proposition 1 with Lemmas 2 and 3, we come to the following statement.

Proposition 2 *Let the assumptions of Proposition 1 be satisfied. Then, for every $\sigma \in (0,\,3/4)$, one can choose $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ according to this proposition in such a way that if* (17) *holds for $k = 0$, it is valid for all $k \in \mathbb{N}$.*

Our next goal is to understand what can be done with this "if (17) holds for $k = 0$" in Proposition 2. Example 2 demonstrates that (17) cannot be guaranteed for $u^0 \in K_{\varepsilon,\delta}(\bar{v})$, no matter how small $\varepsilon > 0$ and $\delta > 0$ are. At the same time, Example 3 suggests that the iterates might have the chance to leave the troublesome set $\mathscr{W}$ (given by (29)), where the full step may not be accepted. And indeed, we are going to show now that, after some finite number of steps, the iterates must leave the set $\mathscr{W}$, while still staying in $K_{\varepsilon,\delta}(\bar{v})$ with the needed values of $\varepsilon > 0$ and $\delta > 0$. Then, at that moment of exit from the set $\mathscr{W}$ where the full step may not be accepted, Proposition 2 will come into play.

This part of the analysis requires several steps before Proposition 3 is obtained.

Lemma 4 *Let the assumptions of Proposition 1 be satisfied. Then, for every $\sigma \in (0,\,1)$, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$, $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$, and $\bar{\alpha} = \bar{\alpha}(\bar{v}) > 0$ such that, for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, the Newton equation* (2) *has a unique solution v^k, and* (12) *holds for all $\alpha \in (0,\,\bar{\alpha}]$.*

Proof Under the smoothness assumptions of Proposition 1, Φ' is Lipschitz-continuous near $\bar{u}$ with some constant $\ell > 0$, and hence,

$$\|\Phi(u+v) - \Phi(u) - \Phi'(u)v\| \leq \frac{\ell}{2}\|v\|^2$$

holds for all $u \in \mathbb{R}^p$ close to $\bar{u}$, and all $v \in \mathbb{R}^p$ close enough to 0 (see, e.g., [12, Lemma A.11]). Therefore, assuming again that $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0, 1)$ are chosen according to Lemma 1, we have that for every $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$ there exists the unique v^k solving (2), it satisfies the second relation in (20), and then for all $\alpha \in (0, 1]$ it holds that

$$\|\Phi(u^k + \alpha v^k)\| \leq (1-\alpha)\|\Phi(u^k)\| + L\alpha^2\|u_2^k\|^2,$$

with some $L > 0$ independent of u^k and α, provided $\bar{\varepsilon} > 0$ is small enough. This implies that (12) always holds if

$$\alpha \leq \frac{(1-\sigma)}{L}\frac{\|\Phi(u^k)\|}{\|u_2^k\|^2}. \tag{31}$$

It remains to employ the expansion in (22) and (24), implying that

$$\|\Phi(u^k)\| \geq \|\Pi\Phi(u^k)\| \geq \gamma\|u_2^k\|^2,$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. This yields that the right-hand side of (31) is separated from zero by some positive constant not depending on u^k. □

With Lemma 4 at hand, we are now ready to show that the iterates generated by Algorithm 1 cannot stay in the troublesome set $\mathscr{W}$ infinitely long. Specifically, we show that (32) below must hold on some iteration eventually, which means that the corresponding iterate is not in $\mathscr{W}$, see (29).

Lemma 5 *Let the assumptions of Proposition 1 be satisfied. Then, for every $\sigma \in (0, 1)$ and $\theta \in (0, 1)$, there exist $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$ and $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$ such that, if a sequence $\{u^k\} \subset K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$ is generated by Algorithm 1, then*

$$\|u_1^k - \bar{u}_1\| \leq \bar{\gamma}\|u_2^k - \bar{u}_2\|^2 \tag{32}$$

holds for some k.

Proof Let us assume again that $\bar{u} = 0$, and $\bar{\varepsilon} > 0$ and $\bar{\delta} \in (0, 1)$ are chosen according to Lemma 1. Then, for every $u^k \in K_{\bar{\varepsilon}, \bar{\delta}}(\bar{v})$ there exists the unique v^k solving (2), this v^k satisfies (3)–(6), and (20) holds. This leads to

$$u_1^{k+1} = u_1^k + \alpha_k v_1^k = (1-\alpha_k)u_1^k + O(\|u_1^k\|\|u_2^k\|) + O(\|u_2^k\|^3), \tag{33}$$

and

$$u_2^{k+1} = u_2^k + \alpha_k v_2^k = \left(1 - \frac{1}{2}\alpha_k\right)u_2^k + O(\|u_1^k\|) + O(\|u_2^k\|^2). \tag{34}$$

Let us now suppose that u^k violates (32) for some $k \in \mathbb{N}$. Then, (33) and (34) imply the existence of $C > 0$ and $c > 0$ such that

$$\|u_1^{k+1}\| \leq (1-\alpha_k)\|u_1^k\| + C(\|u_1^k\|\|u_2^k\| + \|u_2^k\|^3) \leq \left(1 - \alpha_k + \frac{2C}{\bar{\gamma}}\|u_2^k\|\right)\|u_1^k\|$$

and, with (19),

$$\|u_2^{k+1}\| \geq \left(1 - \frac{1}{2}\alpha_k\right)\|u_2^k\| - c(\|u_1^k\| + \|u_2^k\|^2) \geq \left(1 - \frac{1}{2}\alpha_k - c\left(\frac{\bar{\delta}}{1-\bar{\delta}} + \|u_2^k\|\right)\right)\|u_2^k\|$$

follows. Therefore, we obtain

$$\frac{\|u_1^{k+1}\|}{\|u_2^{k+1}\|^2} \leq \frac{1-\alpha_k+\dfrac{2C}{\bar{\gamma}}\|u_2^k\|}{\left(1-\dfrac{1}{2}\alpha_k - c\left(\dfrac{\bar{\delta}}{1-\bar{\delta}}+\|u_2^k\|\right)\right)^2}\frac{\|u_1^k\|}{\|u_2^k\|^2}. \tag{35}$$

According to Lemma 4, we may assume that there exists $\widehat{\alpha} \in (0, 1]$ independent of u^k such that $\alpha_k \in [\widehat{\alpha}, 1]$. It can be easily seen that the function $\varphi : [0, 1] \to \mathbb{R}$, $\varphi(\alpha) = (1-\alpha)/(1-\alpha/2)^2$, is monotonically decreasing with $\varphi(0) = 1$, $\varphi(1) = 0$. Therefore, $\widehat{q} = \varphi(\widehat{\alpha}) \in [0, 1)$, and $\varphi(\alpha) \leq \widehat{q}$ for all $\alpha \in [\widehat{\alpha}, 1]$. Then, by the uniform continuity argument, we obtain that for every $q \in (\widehat{q}, 1)$ it holds that

$$\frac{1-\alpha_k+\dfrac{2C}{\bar{\gamma}}\|u_2^k\|}{\left(1-\dfrac{1}{2}\alpha_k - c\left(\dfrac{\bar{\delta}}{1-\bar{\delta}}+\|u_2^k\|\right)\right)^2} \in (0, q],$$

provided $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are small enough. Inequality (35) now yields

$$\frac{\|u_1^{k+1}\|}{\|u_2^{k+1}\|^2} \leq q\frac{\|u_1^k\|}{\|u_2^k\|^2}.$$

Since $q \in (0, 1)$, this implies that if u^k violates (32) for all sufficiently large k, then we have that $\|u_1^k\|/\|u_2^k\|^2 \to 0$ as $k \to \infty$. But then $\|u_1^k\|/\|u_2^k\|^2 \leq \bar{\gamma}$ for all sufficiently large k, which contradicts the violation of (32). □

Remark 1 If we assume in addition that there exists $\bar{\Gamma} > 0$ such that the sequence $\{u^k\}$ in Lemma 5 satisfies

$$\|u_1^k - \bar{u}_1\| \leq \bar{\Gamma}\|u_2^k - \bar{u}_2\|^2 \tag{36}$$

for all k (cf. Lemma 2), then it can be seen from the proof of Lemma 5 that the number of steps before exiting the troublesome set $\mathscr{W}$ is actually uniformly bounded: there exists a positive integer κ independent of $\{u^k\}$, such that (32) holds for some $k \in \{1, \ldots, \kappa\}$.

Lemma 6 *Let the assumptions of Proposition 1 be satisfied. Then, for every $\bar{\varepsilon} > 0$, $\bar{\delta} > 0$, $\sigma \in (0, 1)$, and $\theta \in (0, 1)$, and for every $\bar{\gamma} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that, for every starting point $u^0 \in K_{\varepsilon,\delta}(\bar{v})$, Algorithm 1 uniquely defines the sequence $\{u^k\}$, and $\{u^k\} \subset K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$.*

Proof Assume again that $\bar{u} = 0$. Without loss of generality, we can further assume that $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ are chosen according to both Lemmas 1 and 4 (an explanation can be found in the proof in [9, Theorem 1]). Then, for every $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, there exists the unique v^k solving (2), and this v^k satisfies (3)–(6), implying (33)–(34).

The rest of the proof is by following the lines of the proof in [9, Theorem 1], with the necessary modifications related to the fact that in [9, Theorem 1], (33)–(34) are used with $\alpha_k = 1$, while here possible values of α_k are characterized by Lemma 4. It is also important to observe that α_k can be less than 1 only when (36) holds with $\bar{\Gamma} > 0$ from Lemma 2. Then (33) gives

$$u_1^{k+1} = O(\|u_2^k\|^2),$$

and hence, this estimate holds for $u^k \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$, no matter which α_k is accepted at this iterate. □

By putting together Lemma 2, Proposition 2, and Lemmas 5–6, we obtain the following result.

Proposition 3 *Let the assumptions of Proposition 1 be satisfied. Then, for every* $\bar{\varepsilon} > 0$*,* $\bar{\delta} > 0$*,* $\sigma \in (0, 3/4)$*, and* $\theta \in (0, 1)$*, there exist* $\varepsilon = \varepsilon(\bar{v}) > 0$ *and* $\delta = \delta(\bar{v}) > 0$ *such that, for every starting point* $u^0 \in K_{\varepsilon,\delta}(\bar{v})$*, Algorithm 1 uniquely defines the sequence* $\{u^k\}$*,* $\{u^k\} \subset K_{\bar{\varepsilon},\bar{\delta}}(\bar{v})$*, and* (17) *holds for all* k *large enough.*

We next analyze what happens if the starting point of the algorithm is close to $\bar{u}$ but perhaps is not in $K_{\varepsilon,\delta}(\bar{v})$, with $\varepsilon > 0$ and $\delta > 0$ from Proposition 3. We first show that in this case the full Newton step typically leads to $K_{\varepsilon,\delta}(\bar{v})$ for an appropriate direction $\bar{v}$.

Lemma 7 *Let* $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ *be twice differentiable near* $\bar{u} \in \mathbb{R}^p$*, with its second derivative Lipschitz-continuous with respect to* $\bar{u}$*. Let* $\bar{u}$ *be a solution of* (1)*.*

Then, for every $\varepsilon > 0$ *and* $\delta > 0$*, and every* $v \in \mathbb{R}^p$ *such that* Φ *is 2-regular at* $\bar{u}$ *in the direction* v*, and* $\pi(v) \neq 0$ *(with* π *defined by* (6)*), there exists* $\tau > 0$ *such that, for every* $t \in (0, \tau]$*, the Newton equation* (2) *for* $k = 0$ *and* $u^0 = \bar{u} + tv$ *has a unique solution* v^0*, and* $u^0 + v^0 \in K_{\varepsilon,\delta}(\pi(v)/\|\pi(v)\|)$*.*

Proof Again assume for simplicity that $\bar{u} = 0$. The existence and uniqueness of v^0 solving (2) with $k = 0$ for $\tau > 0$ small enough follows by Lemma 1, and this v^0 satisfies (4) and (5). Therefore, taking into account that according to (6) and the definition of $\mathscr{B}(\cdot)$, $\pi(\cdot)$ is homogeneous on its domain, we obtain that

$$u^0 + v^0 = \frac{1}{2} t\pi(v) + O(t^2), \tag{37}$$

which further implies that

$$\|u^0 + v^0\| = \frac{1}{2} t\|\pi(v)\| + O(t^2)$$

as $t \to 0$. Then clearly $\|u^0 + v^0\| \leq \varepsilon$ if $\tau > 0$ is small enough, and

$$\left\| \frac{u^0 + v^0}{\|u^0 + v^0\|} - \frac{\pi(v)}{\|\pi(v)\|} \right\| = \left\| \frac{\pi(v) + O(t)}{\|\pi(v)\| + O(t)} - \frac{\pi(v)}{\|\pi(v)\|} \right\| = O(t)$$

as $t \to 0$. This implies that

$$\left\| \frac{u^0 + v^0}{\|u^0 + v^0\|} - \frac{\pi(v)}{\|\pi(v)\|} \right\| \leq \delta$$

if $\tau > 0$ is small enough. □

We finally show that the full Newton step from such starting points (close to $\bar{u}$ but perhaps not in $K_{\varepsilon,\delta}(\bar{v})$) is indeed accepted by linesearch in Algorithm 1.

Lemma 8 *Let* $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ *be twice differentiable near* $\bar{u} \in \mathbb{R}^p$*, with its second derivative Lipschitz-continuous with respect to* $\bar{u}$*. Let* $\bar{u}$ *be a solution of* (1)*. Then, for every* $v \in \mathbb{R}^p$ *such that* Φ *is 2-regular at* $\bar{u}$ *in the direction* v*, and* $v \notin \ker \Phi'(\bar{u})$*, there exists* $\tau > 0$ *such that, for every* $t \in (0, \tau]$*, the Newton equation* (2) *for* $k = 0$ *and* $u^0 = \bar{u} + tv$ *has a unique solution* v^0*, and* (17) *is satisfied with* $k = 0$ *for this* v^0*.*

Proof We again assume for simplicity that $\bar{u} = 0$. By the same argument as in the proof of Lemma 7, we obtain the existence and uniqueness of v^0, and the relation (37). Implying, in particular, that

$$v^0 = -tv + \frac{1}{2}t\pi(v) + O(t^2) = O(t)$$

as $t \to 0$. Then, by (2), we obtain that

$$\Phi(u^0 + v^0) = \Phi(u^0) + \Phi'(u^0)v^0 + O(\|v^0\|^2) = O(t^2),$$

while

$$\Phi(u^0) = \Phi'(0)u^0 + O(\|u^0\|^2) = t\Phi'(0)v + O(t^2)$$

as $t \to 0$. Since $\Phi'(0)v \neq 0$, the last two relations imply the needed assertion. □

Example 3 demonstrates that, when initialized with $u^0 - \bar{u}$ far from $\ker \Phi'(\bar{u})$, it can be quite typical that the Newton method enters the troublesome set $\mathscr{W}$ (given by (29)) in one step. It is interesting to note that such behavior in this example is caused solely by the presence of the mixed term u_1u_2 in the second component of Φ, which does not allow the step to pass further and to "slip through" $\mathscr{W}$. In the absence of this term (like in Example 2), adding any higher- (than second-) order terms would not result in this effect: using (3)–(6), it can be seen that entering $\mathscr{W}$ can be avoided if, say, the restriction of $\Pi\Phi''(\bar{u})[u^0]$ on $(\ker \Phi'(\bar{u}))^{\perp}$ is identically zero.

However, in general, entering $K_{\varepsilon,\delta}(\pi(v)/\|\pi(v)\|)$ by one step of Algorithm 1 (i.e., the behavior established in Lemmas 7 and 8) does not solve the problem of subsequent possible rejection of the full step by itself. Fortunately, when complemented by Proposition 3, Lemmas 7 and 8 give the needed result, also employing Proposition 1. Specifically, our final result reads as follows.

Theorem 1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be twice differentiable near $\bar{u} \in \mathbb{R}^p$, with its second derivative Lipschitz-continuous with respect to $\bar{u}$. Let $\bar{u}$ be a solution of* (1)*, and assume that there exists $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. Then, there exists a set $U \subset \mathbb{R}^p$ starlike with respect to $\bar{u}$, with possibly excluded directions being only those $v \in \mathscr{S}$ for which Φ is not 2-regular at $\bar{u}$ in the direction v or in the direction $\pi(v)$ (with π given by* (6)*), and such that, for every starting point $u^0 \in U$, Algorithm 1 with $\sigma \in (0, 3/4)$ uniquely defines the sequence $\{u^k\}$, and* (17) *holds for all k large enough. Moreover, $\{u^k\}$ converges to $\bar{u}$ with the linear rate* (10)*, and the sequence $\{(u^k - \bar{u})/\|u^k - \bar{u}\|\}$ converges to some $v \in \ker \Phi'(\bar{u})$.*

Proof Take any $v \in \mathscr{S}$ such that Φ is 2-regular at $\bar{u}$ in both directions v and $\pi(v)$. Since $\Phi'(\bar{u})$ is singular (as it has a nontrivial null space), this implies that $\pi(v) \neq 0$.

Select $\varepsilon = \varepsilon(\pi(v)/\|\pi(v)\|) > 0$ and $\delta = \delta(\pi(v)/\|\pi(v)\|) > 0$ according to Proposition 1 applied with $\bar{v}$ substituted by $\pi(v)/\|\pi(v)\|$ (and with any pre-fixed $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$). Redefine $\bar{\varepsilon}$ as $\varepsilon(\pi(v)/\|\pi(v)\|)$, and $\bar{\delta}$ as $\delta(\pi(v)/\|\pi(v)\|)$.

Select $\varepsilon = \varepsilon(\pi(v)/\|\pi(v)\|) > 0$ and $\delta = \delta(\pi(v)/\|\pi(v)\|) > 0$ according to Proposition 3 applied with $\bar{v}$ substituted by $\pi(v)/\|\pi(v)\|$, and with $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ specified above.

Finally, if $v \notin \ker \Phi'(\bar{u})$, select $\tau(v) > 0$ according to Lemmas 7 and 8 applied with the specified $\varepsilon > 0$ and $\delta > 0$; otherwise put $\tau(v) = \varepsilon(v)$ (observe that in this case, according to (6), $v = \pi(v) = \pi(v)/\|\pi(v)\|$). It remains to define U as the union of the sets $\{\bar{u} + tv \mid 0 < t \leq \tau(v)\}$ over all $v \in \mathscr{S}$ such that Φ is 2-regular at $\bar{u}$ in both directions v and $\pi(v)$. By the definition of $\tau(v)$, this set possesses all the needed properties. □

Observe that the excluded directions of the set U are those v satisfying either $\det \mathscr{B}(v) = 0$, or $\det \mathscr{B}(v) \neq 0$ but $\det \mathscr{B}(\det \mathscr{B}(v)\pi(v)) = 0$, and all such v are contained in the null sets of the corresponding two homogeneous polynomials which are both nontrivial provided Φ is 2-regular in at least one direction in $\ker \Phi'(\bar{u})$. This implies that under the assumptions of Theorem 1, the set of excluded directions is thin, which means that the convergence domain is "large".

5 Numerical illustrations

In this section, we provide some numerical examples and the related statistics concerning ultimate acceptance of the unit stepsize, and the effect of acceleration by extrapolation. The algorithms being tested are the following (the abbreviations correspond to the names of rows in the tables below, and also to what appears in the captions of figures):

- NM (for "Newton Method") is Algorithm 1 without any modifications, and with parameter values $\sigma = 0.01$ and $\theta = 0.5$.
- EP (for "ExtraPolation") is NM including the generation of an auxiliary sequence $\{\hat{u}^k\}$ according to (14).

Algorithm NM terminates after an iterate u^k is generated satisfying

$$\|\Phi(u^k)\| \leq 10^{-14}. \tag{38}$$

For Algorithm EP, and $k \geq 1$, (38) is replaced by

$$\min\{\|\Phi(u^k)\|, \|\Phi(\hat{u}^k)\|\} \leq 10^{-14}. \tag{39}$$

Convergence to the primal solution of interest $\bar{u}$ is declared when, at successful termination, it holds that

$$\|u^k - \bar{u}\| \leq 10^{-4}$$

(with u^k replaced by $\hat{u}^k$ for EP if the minimum in (39) is achieved by the second number). If successful termination did not occur after 200 iterations, or if at some iteration the linear solver failed to solve the Newton equation (2), or the backtracking procedure in Algorithm 1 produced a trial value α such that $\alpha\|v^k\| \leq 10^{-10}$, the process was terminated declaring failure.

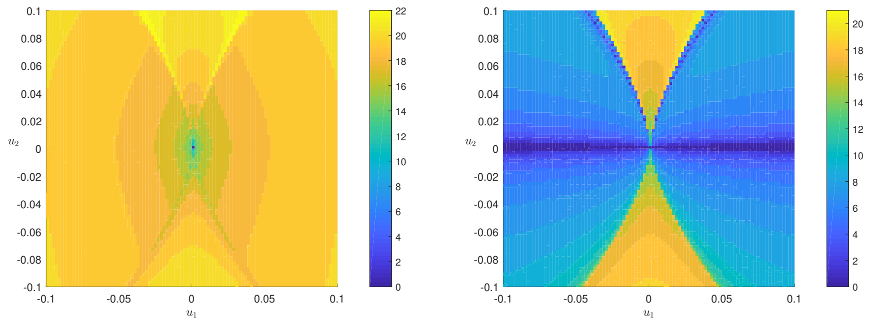
We start with runs for some specific test problems with singular solutions. In our experiments we used about 20 problems, taken from various sources in the literature. Here, we opted to report only on a selection of them, to demonstrate some representative patterns of the local convergence behavior. Namely, we report on problems from Examples 2 (with $a = \sqrt{15}$), 3, and 4 (with $q = 3$), presented above, and also the following problem which is of interest because it violates the key assumption of Proposition 1 and Theorem 1.

Example 5 Let $p = 2$, $\Phi(u) = (u_1(u_1^2 + u_2), u_2(1 + u_2))$. In this case, the unique singular solution of (1) is $\bar{u} = 0$ (while there are other nonsingular solutions $(-1, -1)$, $(0, -1)$, $(1, -1)$). We have that $\ker \Phi'(\bar{u}) = \{v \in \mathbb{R}^2 \mid v_2 = 0\}$, and that Φ is not 2-regular at $\bar{u}$ in any direction from $\ker \Phi'(\bar{u})$.

For each of these problems, we performed 100 runs from random starting points uniformly distributed in a box centered at the solution of interest, with the edge lengths of 0.2. The results are summarized in Table 1. The meaning of the first two columns in the tables is obvious; the other columns are as follows:

Table 1: Results for selected test problems

Problem	Method	% runs with convergence	# iterations			# last full steps			% last full steps		
			min	∅	max	min	∅	max	min	∅	max
Ex. 2	NM	100	13	19.3	21	13	19.3	21	100	100	100
$a=\sqrt{15}$	EP	100	2	2.0	2	2	2.0	2	100	100	100
Ex. 3	NM	96.0	13	20.0	24	12	19.2	22	80.0	96.3	100
	EP	96.0	5	8.3	12	3	7.6	10	50.0	91.5	100
Ex. 4	NM	100	15	19.8	23	15	19.7	23	94.4	99.9	100
$q=3$	EP	100	1	9.3	21	1	9.3	21	80.0	99.9	100
Ex. 5	NM	100	6	20.4	24	6	20.4	24	94.4	99.9	100
	EP	100	5	18.7	22	5	18.7	22	93.8	99.8	100



(a) NM (b) EP

Fig. 2: Example 4: iteration count.

- Column 3: percentage of runs with convergence to the solution of interest $\bar{u}=0$.
- Columns 4–6: mimimum/average/maximum number of iterations over cases of convergence to $\bar{u}=0$.
- Columns 7–9: mimimum/average/maximum number of last full steps over cases of convergence to $\bar{u}=0$.
- Columns 10–12: mimimum/average/maximum percentage of last full steps with respect to the number of all iterations, over cases of convergence to $\bar{u}=0$.

The results in Table 1 for Examples 2 and 3 fully agree with the theory presented above. The overall accelerating effect of EP in these examples is evident. For Example 3, the rare cases when convergence to the solution of interest was not observed are concerned with convergence to another solution.

For Example 4, in Figure 2 we additionally show the dependence of the iteration count until successful termination on starting points. Observe that in this example, 2-regularity does not hold in a direction v with $v_1=0$, which belongs to $\ker \Phi'(\bar{u}) = \mathbb{R}^2$, and once the NM iterates get close to this direction, the convergence pattern of Proposition 1 becomes violated, and stays so, and the acceleration effect of EP is lost in such cases. However, the overall accelerating effect of EP is still evident.

Furthermore, Example 5 violates the key assumption of Proposition 1, and no accelerating effect is observed for EP, even though the full step is ultimately accepted.

Table 2: Results for random test problems

(p, r)	Method	% runs with	# iterations			# last full steps			% last full steps		
		convergence	min	∅	max	min	∅	max	min	∅	max
$(2, 0)$	NM	100	16	20.7	22	16	20.7	22	100	100	100
	EP	100	1	1.0	1	1	1.0	1	100	100	100
$(2, 1)$	NM	98.7	13	21.0	38	13	20.8	25	41.7	99.0	100
	EP	98.7	4	6.6	27	3	6.3	18	22.2	97.7	100
$(5, 0)$	NM	100	20	22.0	23	20	22.0	23	100	100	100
	EP	100	1	1.0	1	1	1.0	1	100	100	100
$(5, 2)$	NM	99.2	20	23.0	34	19	22.8	26	58.8	99.1	100
	EP	99.2	5	7.9	21	5	7.7	15	33.3	97.7	100
$(5, 4)$	NM	93.1	9	22.4	44	9	21.8	27	40.9	97.8	100
	EP	93.1	4	8.4	32	4	7.8	18	18.8	95.4	100
$(10, 0)$	NM	100	22	22.9	24	22	22.9	24	100	100	100
	EP	100	1	1.0	1	1	1.0	1	100	100	100
$(10, 3)$	NM	99.5	22	24.2	31	20	23.8	28	74.2	98.6	100
	EP	99.5	6	8.6	18	5	8.3	15	53.8	96.5	100
$(10, 7)$	NM	93.1	20	24.4	32	18	23.7	28	62.1	97.2	100
	EP	93.1	7	9.9	20	6	9.1	17	35.3	93.7	100
$(10, 9)$	NM	82.3	16	23.7	77	16	22.8	28	24.7	96.9	100
	EP	82.3	6	9.8	65	5	8.9	17	10.8	93.7	100

We next report numerical results for randomly generated linear-quadratic test problems of the form

$$\Phi(u) = Au + \frac{1}{2}B[u, u].$$

The generation is as follows: $A \in \mathbb{R}^{p\times p}$ is a matrix of a specified rank r, with random entries uniformly distributed in $[-10, 10]$, $B[u, u] = (\langle B_1 u, u\rangle, \ldots, \langle B_p u, u\rangle)$ with $B_i \in \mathbb{R}^{p\times p}$ being symmetric matrices with random entries uniformly distributed in $[-10, 10]$, $i = 1, \ldots, p$. With these choices, $\bar{u} = 0$ is always a solution of (1), and it is a singular solution if $r < p$. Observe, however, that this solution does not need to be unique, and convergence to other solutions is certainly possible. This is the main reason why the numbers in column PC of Table 2 are sometimes less than 100%.

For each pair (p, r), we generated 100 problems, and for each of these problems, we performed 10 runs from random starting points generated exactly as before. The results of these runs for some pairs (p, r) (the specific choice of which does not affect the overall picture in any serious way) are summarized in Table 2. The meaning of the columns in this table is as before.

The behavior demonstrated by Table 2 is nearly ideal, with full step being accepted ultimately almost always (at least in cases of convergence to the solution of interest), and with clear accelerating effect of EP.

The very special behavior in cases when $r = 0$ is explained by the fact that in these cases, Φ is a homogeneous polynomial of order 2. This implies that the Newton step from any point $u^k \in \mathbb{R}^p$ yields exactly $u^k/2$, and hence, the EP step yields exactly 0.

Acknowledgment. The authors thank Ivan Rodin for his assistance with numerical experiments, and the two anonymous referees for their useful comments on the original version of this article.

References

1. 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)
2. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. 146, 1–36 (2014)
3. 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)
4. Fischer, A., Izmailov, A.F., M.V. Solodov.: Local attractors of Newton-type methods for constrained equations and complementarity problems with nonisolated solutions. J. Optim. Theory Appl. 180, 140–169 (2019)
5. Fan, J.-Y., Yuan, Y.-X.: On the quadratic convergence of the Levenberg–Marquardt method. Computing. 74, 23–39 (2005)
6. Griewank, A.: Analysis and modification of Newton's method at singularities. PhD Thesis. Australian National University, Canberra (1980)
7. Griewank, A.: Starlike domains of convergence for Newton's method at singularities. Numer. Math. 35, 95–111 (1980)
8. Griewank, A.: On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. 27, 537–563 (1985)
9. 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)
10. Izmailov, A.F.; Kurennoy, A.S., Solodov, M.V.: Critical solutions of nonlinear equations: Stability issues. Math. Program. 168, 475–507 (2018)
11. Izmailov, A.F., Solodov, M.V.: Stabilized SQP revisited. Math. Program. 122, 93–120 (2012)
12. Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, Springer International Publishing, Switzerland (2014)
13. Izmailov, A.F.; Tretyakov, A.A.: 2-Regular Solutions of Nonlinear Problems. Theory and Numerical Methods. In Russian. Fizmatlit, Moscow (1999)
14. Levenberg, K.: A method for the solution of certain non-linear problems in least squares. Quarterly of Appl. Math. 2, 164–168 (1944)
15. Marquardt, D.W.: An algorithm for least squares estimation of non-linear parameters. SIAM J. 11, 431–441 (1963)
16. Oberlin, C., Wright, S.J.: An accelerated Newton method for equations with semismooth Jacobians and nonlinear complementarity problems. Math. Program. 117, 355–386 (2009)
17. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. 11, 253–275 (1998)
18. Yamashita, N., Fukushima, M.: On the rate of convergence of the Levenberg–Marquardt method. Computing. Suppl. 15, 237–249 (2001)