

ACCELERATING CONVERGENCE OF THE GLOBALIZED NEWTON METHOD TO CRITICAL SOLUTIONS OF NONLINEAR EQUATIONS

A. Fischer¹, A.F. Izmailov², and M.V. Solodov³

April 7, 2020 (revised September 9, 2020)

ABSTRACT

In the case of singular (and possibly even nonisolated) solutions of nonlinear equations, while superlinear convergence of the Newton method cannot be guaranteed, local linear convergence from large domains of starting points still holds under certain reasonable assumptions. We consider a linesearch globalization of the Newton method, combined with extrapolation and over-relaxation accelerating techniques, aiming at a speed up of convergence to critical solutions (a certain class of singular solutions). Numerical results indicate that an acceleration is observed indeed.

Key words: Nonlinear equation; Newton method; singular solution; critical solution; 2-regularity; linear convergence; superlinear convergence; extrapolation; overrelaxation; linesearch.

AMS subject classifications: 65H10, 65H20.

¹ Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany.
Email: andreas.fischer@tu-dresden.de

² Lomonosov Moscow State University, MSU, Uchebnyy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: izmaf@ccas.ru

³ IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: solodov@impa.br

1 Introduction

In this paper, we are interested in the behavior of various forms of linesearch Newton methods for a system of nonlinear equations

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

where $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$ is a sufficiently smooth mapping. We are especially interested in the case when this system has singular (and possibly even nonisolated) solutions. Here, by singularity of a solution \bar{u} we mean that the Jacobian $\Phi'(\bar{u})$ is a singular matrix.

For a given iterate $u^k \in \mathbb{R}^p$, the basic Newton method generates the next iterate by the update $u^{k+1} = u^k + v^k$, where v^k solves the linear system

$$\Phi(u^k) + \Phi'(u^k)v = 0. \tag{1.2}$$

The classical theory says that when initialized close to a nonsingular solution, this method uniquely defines a sequence of iterates, converging to this solution superlinearly. Moreover, the standard linesearch technique employing the residual of (1.1), used to globalize convergence, accepts the unit stepsize asymptotically, thus ensuring fast local convergence of the overall algorithm. See, e.g., [17, Chapter 5.1.1].

In case of a singular solution \bar{u} , it is evident that one cannot guarantee convergence of the Newton method from all starting points in the entire neighborhood of \bar{u} , as the iteration equation (1.2) need not even be solvable for some u^k arbitrarily close to such \bar{u} . Nevertheless, local convergence at a linear rate can still be established from some large domains of starting points, and under reasonable assumptions. The most relevant (from our perspective) results of this kind are those obtained in [9, 10]. See also the survey in [11], and rich literature on the subject cited therein. These results will be discussed in Section 2, where a certain special pattern of convergence will be highlighted. As demonstrated in [15, 16], the key assumption needed for these results to be valid may only hold at those singular solutions which are called critical. The property of criticality is characterized by violation of the local Lipschitzian error bound near such solution. This implies that critical solutions (when they exist) are specially attractive for Newtonian sequences, even when every neighborhood of such solution contains other (typically noncritical) solutions.

It should be mentioned at this point that there exist some stabilized modifications of the basic Newton scheme, possessing superlinear or even quadratic local convergence to a nearby solution, when initialized close to a noncritical one. And this is so even if this solution is singular, or even nonisolated. The examples are the classical Levenberg–Marquardt method [13, 14] with an appropriate adaptive control of the regularization parameter [5, 7, 21], the stabilized Newton–Lagrange method (stabilized sequential quadratic programming) when (1.1) corresponds to the Lagrange

optimality system for an equality-constrained optimization problem [6, 12, 20], and the LP-Newton method [3, 4]. All these methods possess similar strong local convergence properties specified above. In particular, their attraction domains to critical solutions become smaller than those for the basic Newton method, and in this sense, these stabilization techniques locally serve the goal of avoiding convergence to critical solutions. However, as demonstrated in [15], domains of attraction to critical solutions may still be quite large, and in cases of convergence to such solutions, the superlinear rate is still lost. For that reason, the specified nice local convergence properties of the stabilized Newton-type methods near noncritical solutions may not show up in computation, because their sequences may never enter a sufficiently small neighborhood of any noncritical solution. Especially when considering globalized versions of these methods with initialization at remote starting points. For this reason, here we pursue a different path. We do not try to avoid convergence to a critical solution (which appears difficult in the singular case). Instead, the idea is to make use of the special convergence pattern of the basic Newton method to critical solutions, in order to ensure relatively fast linear convergence to such solutions for versions of this method employing linesearch for globalization. When considering the latter, the key issue is the ultimate acceptance of the unit stepsize: even for linear convergence this question must be answered in the affirmative. This turned out to be highly nontrivial in the context of singular solutions, but the problem was recently successfully resolved in [8]. We shall comment on the corresponding results in Section 2.

Some known approaches for accelerating convergence to singular solutions are extrapolation and overrelaxation techniques, developed in [9, 11]. Incorporating these essentially local constructions into a linesearch globalization framework is the main subject of this paper. Both extrapolation and overrelaxation rely upon the special convergence pattern of the basic Newton method to critical solutions. Hence, this pattern should be preserved within the linesearch framework, thus making the results from [8] on ultimate acceptance of unit stepsize essential for justification of the resulting algorithms.

The rest of the paper is organized as follows. We provide some necessary preliminaries in Section 2. In Section 3, we discuss extrapolation and overrelaxation techniques, as well as linesearch globalization of the Newton method, and the related convergence theory. Finally, Section 4 presents some numerical results highlighting the effect of acceleration techniques. Somewhat surprisingly (and to the best of our knowledge), there appear to be no previous studies in the literature comparing various acceleration options, not even in the local setting. So, we think, what we report is also useful for this reason.

2 Preliminaries

Throughout the paper, $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ stand for the Euclidian inner product and norm, respectively.

Recall that a set $U \subset \mathbb{R}^p$ is called starlike with respect to $\bar{u} \in \mathbb{R}^p$ if for every $u \in U$ and $t \in (0, 1]$, it holds that $tu + (1 - t)\bar{u} \in U$. Any $v \in \mathbb{R}^p$ is called an excluded direction for a set U starlike with respect to \bar{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.

The key role in local analyses of Newton-type methods near singular solutions is played by the notion of 2-regularity. Assuming that Φ is twice differentiable at \bar{u} , we call it 2-regular at \bar{u} in a direction $v \in \mathbb{R}^p$ if

$$\text{im } \Phi'(\bar{u}) + \Phi''(\bar{u})[v, \ker \Phi'(\bar{u})] = \mathbb{R}^p.$$

Evidently, if $\Phi'(\bar{u})$ is nonsingular, then Φ is 2-regular at \bar{u} in every direction, including $v = 0$. What is important, is that 2-regularity may naturally hold at singular (and even nonisolated) solutions in nonzero directions, including those from $\ker \Phi'(\bar{u})$. This plays a key role in the results presented in this section.

The following theorem summarizes the main conclusions from [10, Theorem 6.1]; see also [11, Theorem 2.1].

Theorem 2.1 *Let $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$ 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 $u \rightarrow \bar{u}$. Let \bar{u} be a singular solution of equation (1.1), and assume that there exists $\bar{v} \in \ker \Phi'(\bar{u})$ such that Φ is 2-regular at \bar{u} in this direction.

Then there exists a set $U \subset \mathbb{R}^p$ starlike with respect to \bar{u} , asymptotically dense at \bar{u} , and such that for every starting point $u^0 \in U$, the basic Newton method with subproblems (1.2) uniquely defines the sequence $\{u^k\} \subset \mathbb{R}^p \setminus \{\bar{u}\}$, this sequence converges to \bar{u} ,

$$\lim_{k \rightarrow \infty} \frac{\|u^{k+1} - \bar{u}\|}{\|u^k - \bar{u}\|} = \frac{1}{2}, \quad (2.1)$$

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

The standard approach to globalization of convergence of the Newton method is the Armijo linesearch procedure in the computed direction, for the residual $\|\Phi(\cdot)\|$ or the squared residual. Here, we prefer to stay with the residual (without square), as this allows for somewhat larger stepsizes, in general. Observe that the residual is differentiable at every point u^k which is not a solution of (1.1), and its gradient is

$(\Phi'(u^k))^T \Phi(u^k) / \|\Phi(u^k)\|$. Then, as is easily seen, the inner product of this gradient and the Newton direction v^k solving (1.2) is equal to $-\|\Phi(u^k)\|$. Hence, v^k is a direction of descent for the residual at u^k , which justifies the use of the Armijo linesearch in this direction. Moreover, these calculations imply that condition (2.2) below is precisely the Armijo inequality for the residual and the Newton direction.

We next recall the result on acceptance of the unit stepsize, obtained in [8, Theorem 1]; it is concerned with the following prototype algorithm.

Algorithm 2.1 Choose the parameters $\sigma \in (0, 1)$ and $\theta \in (0, 1)$. Choose $u^0 \in \mathbb{R}^p$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Compute $v^k \in \mathbb{R}^p$ as a solution of (1.2).
3. Set $\alpha = 1$. If the inequality

$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \sigma\alpha)\|\Phi(u^k)\| \quad (2.2)$$

is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check inequality (2.2) again, etc., until (2.2) becomes valid.

4. Set $u^{k+1} = u^k + \alpha_k v^k$.
5. Increase k by 1 and go to Step 1.

Theorem 2.2 *Under the assumptions of Theorem 2.1, there exists a set $U \subset \mathbb{R}^p$ starlike with respect to \bar{u} , asymptotically dense at \bar{u} , and such that for every starting point $u^0 \in U$, Algorithm 2.1 with $\sigma \in (0, 3/4)$ uniquely defines the sequence $\{u^k\}$, and $\alpha_k = 1$ is accepted at Step 3 of this algorithm for all k large enough.*

3 Extrapolation, overrelaxation, and globalization

Under its assumptions, Theorem 2.1 reveals a very special pattern of convergence of the basic Newton method to singular solutions: the convergence rate is linear with the exact asymptotic ratio equal to $1/2$, and convergence is along a single direction $v \in \ker \Phi'(\bar{u})$.

One technique for accelerating the Newton method in case of convergence to a singular solution, naturally suggested by the specified convergence pattern, is extrapolation [11]. In its simplest form, it consists of generating, along with the Newton sequence $\{u^k\}$, an auxiliary sequence $\{\hat{u}^k\}$ obtained by doubling the Newton step: for each k compute

$$\hat{u}^{k+1} = u^k + 2v^k. \quad (3.1)$$

There are no reasons to expect that this would lead to superlinear convergence, but it follows from [11, Theorem 4.1] that under the assumptions of Theorem 2.1 above, $\{\hat{u}^k\}$ 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 $\{u^k\}$. Observe that the main iterative sequences $\{u^k\}$ are themselves not affected by this acceleration technique in any way, and that generating the auxiliary sequence $\{\hat{u}^k\}$ costs (basically) nothing. Therefore, extrapolation can be easily incorporated into any globalization of the Newton method, like in Algorithm 2.1: the extrapolated iterates \hat{u}^k can be generated according to (3.1) all the way from the beginning of the process, in parallel with the main iterates u^k . Moreover, Theorem 2.2 allows to expect that extrapolation will take effect when combined with linesearch globalization in such a way.

At this point, we must observe that “deeper” extrapolation provides further acceleration of convergence, and in principle, gives arbitrarily fast linear convergence [9, 11]. However, and unfortunately, this faster convergence of the iterates is not directly reflected in the rate of decrease of the residual of the equation. More in detail, close to a singular solution, the residual need not be proportional to the distance to the solution. According to our experience, when using the residual-based stopping rules like (4.2) below, the algorithms always employing extrapolation, say “of depth 2”, do not terminate any faster (on average) than those using the simplest extrapolation, even though the iterates at termination produced by “deeper” extrapolation are usually (much) closer to the solution. For this reason, numerical results for “deeper” extrapolation would be more-or-less the same as for the simplest extrapolation, like those reported in Section 4 of this paper (with understanding that the quality of approximate solution obtained would be higher). An observation about numerical behavior of residuals for “deeper” extrapolation in [9, p. 150] might be helpful to overcome this difficulty. But, at the moment, it is not clear how this observation can be reasonably incorporated into globalization schemes. Finally, note also that increasing the “depth” of extrapolation requires more evaluations of Φ . Thus, for the reasons specified above, in what follows we deal with the simplest extrapolation only.

Another acceleration technique is 2- or 3-step overrelaxation [11]. Unlike extrapolation, these techniques do not generate any auxiliary sequences in parallel with the usual Newton sequence, but rather modify the latter by (almost) doubling every second or every third step, respectively. A seemingly appealing idea of simply doubling every step would not work, because this may result in leaving the convergence domain. And in any case, this would destroy the convergence pattern of the Newton method. In case of doubling the step, some intermediate basic Newton iterations are needed to restore this pattern. As demonstrated in [11, Theorems 4.2, 4.3], under the appropriate assumptions one can expect 2-step superlinear convergence rate for the method with 2-step overrelaxation, and 3-step quadratic rate for the method with 3-step overrelaxation.

Incorporating overrelaxation into globalized algorithms is more tricky than for extrapolation, in particular because it should only be initiated when the needed conver-

gence pattern to a singular solution is detected, as in cases of convergence to nonsingular solutions increasing the length of v^k at every second or third iteration of Algorithm 2.1 might evidently be only harmful. Since correct identification of the convergence pattern can never be guaranteed, full theoretical justification of a globalized algorithm equipped with overrelaxation can hardly be possible. Nevertheless, Theorem 2.2 gives it some chances to be successful, and the numerical results in Section 4 confirm that this is often the case.

Interesting acceleration strategies may come out of combining extrapolation and overrelaxation, but we do not elaborate on this idea in the current work.

Observe now that the prototype Algorithm 2.1 is of course impractical, in particular because the Newton direction may not exist at remote points, as well as at points arbitrarily close to singular solutions. The following algorithm complements the prototype Algorithm 2.1 by introducing safeguards for the cases when the Newton direction does not exist or is too large (the latter indicating that it might not be a “good choice”). In those cases, the method resorts to the gradient step for the merit function $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$,

$$\varphi(u) = \frac{1}{2} \|\Phi(u)\|^2, \quad (3.2)$$

with the gradient given by

$$\varphi'(u) = (\Phi'(u))^T \Phi(u). \quad (3.3)$$

By itself, this construction is of course quite standard; see, e.g., [17, Section 5.1]. However, here we highlight the important feature that if the Newton direction is used infinitely often, and the generated iterative sequence is bounded, then all limit points of this sequence are necessarily solutions of the equation (1.1). In particular, using Newton directions always leads to solutions, and cannot lead to stationary points of the residual which are not solutions of (1.1), even when the Jacobian of Φ at such point is singular (if the Jacobian is nonsingular, this property is well-known and obvious). We also emphasize that the behavior specified in Theorem 2.2 subsumes that the Newton direction actually tends to zero as the sequence of iterates converges to the solution possessing the needed properties, and thus, the test on the size of the direction, (3.4) below, is always satisfied in cases of such convergence, at least from some iteration on.

Algorithm 3.1 Choose the parameters $C > 0$, $\tau > 0$, $\sigma \in (0, 1)$, and $\theta \in (0, 1)$. Choose $u^0 \in \mathbb{R}^n$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.
2. Compute $v^k \in \mathbb{R}^n$ as a solution of (1.2). If such v^k cannot be found, or violates

$$\|v^k\| \leq \max\{C, 1/\|\Phi(u^k)\|^\tau\}, \quad (3.4)$$

go to Step 4.

3. Set $\alpha = 1$. If inequality (2.2) is satisfied, set $\alpha_k = \alpha$. Otherwise replace α by $\theta\alpha$, check inequality (2.2) again, etc., until (2.2) becomes valid, and then set $\alpha_k = \alpha$, and go to Step 6.
4. Set $v^k = -\varphi'(u^k)$, with φ defined by (3.2) (see (3.3)). If $v^k = 0$, stop.
5. Set $\alpha = 1$. If the inequality

$$\varphi(u^k + \alpha v^k) \leq \varphi(u^k) - \sigma\alpha\|v^k\|^2 \quad (3.5)$$

is satisfied, set $\alpha_k = \alpha$. Otherwise replace α by $\theta\alpha$, check the inequality again, etc., until (3.5) becomes valid.

6. Set $u^{k+1} = u^k + \alpha_k v^k$.
7. Increase k by 1 and go to Step 1.

Global convergence properties of Algorithm 3.1 are as follows.

Theorem 3.1 *Let $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be continuously differentiable.*

Then, for any starting point $u^0 \in \mathbb{R}^n$, Algorithm 3.1 either terminates with some iterate u^k satisfying

$$(\Phi'(u^k))^T \Phi(u^k) = 0, \quad (3.6)$$

or generates an infinite sequence $\{u^k\}$ such that every accumulation point \bar{u} of this sequence satisfies

$$(\Phi'(\bar{u}))^T \Phi(\bar{u}) = 0. \quad (3.7)$$

Moreover, if for some \bar{u} there exists an infinite subsequence $\{u^{k_j}\}$ convergent to \bar{u} such that the Newton direction is used by Algorithm 3.1 for all indices $k = k_j$ (i.e., for all j , for $k = k_j$ equation (1.2) is solvable, and the computed solution satisfies (3.4)), then

$$\{\Phi(u^k)\} \rightarrow 0 \quad (3.8)$$

as $k \rightarrow \infty$, and in particular, all accumulation points of $\{u^k\}$ are solutions of (1.1).

Proof. If for a current iterate u^k it holds that $\Phi(u^k) = 0$, the algorithm terminates, and (3.6) is evidently satisfied in this case. Otherwise, if the algorithm accepts the Newton direction v^k , then the discussion before Algorithm 2.1 implies that (2.2) is always satisfied after a finite number of backtrackings, i.e., the linesearch procedure at Step 4 of Algorithm 3.1 terminates with some $\alpha_k > 0$, as this is the standard Armijo linesearch for a function which is smooth at u^k , and in a direction of descent for this function at u^k . If the Newton direction is not used, then $v^k = -\varphi'(u^k)$. In this case, the linesearch rule (3.5) at Step 5 of the algorithm terminates finitely for the

same reason if $\varphi'(u^k) \neq 0$, while otherwise the algorithm terminates with (3.6) being satisfied according to (3.3).

Thus, Algorithm 3.1 either terminates with some iterate u^k satisfying (3.6), or generates an infinite sequence $\{u^k\}$. We proceed with analysis of the latter case. Observe that the sequence $\{\|\Phi(u^k)\|\}$ is monotonically non-increasing, whichever search directions are used. Then, since it is bounded below (by zero), it converges.

Consider first the case when there exists an infinite subsequence $\{u^{k_j}\}$ convergent to some \bar{u} , and such that the Newton direction is accepted by Algorithm 3.1 at every point of this subsequence.

If

$$\bar{\alpha} = \limsup_{j \rightarrow \infty} \alpha_{k_j} > 0, \quad (3.9)$$

then (2.2) implies that for infinitely many indices k_j , the residual $\|\Phi(u^{k_j})\|$ is reduced at least linearly (by a factor of at least $(1 - \sigma\bar{\alpha}/2)$). Recalling the monotonicity of $\{\|\Phi(u^k)\|\}$, we conclude that (3.8) holds.

On the other hand, if we do not have (3.9), then it holds that

$$\lim_{j \rightarrow \infty} \alpha_{k_j} = 0, \quad (3.10)$$

implying that for each j large enough the initial stepsize value had been reduced at least once, i.e., the value $\alpha_{k_j}/\theta > \alpha_{k_j}$ does not satisfy (2.2). Thus, it holds that

$$\frac{\|\Phi(u^{k_j} + \alpha_{k_j}v^{k_j}/\theta)\| - \|\Phi(u^{k_j})\|}{\alpha_{k_j}/\theta} > -\sigma\|\Phi(u^{k_j})\|. \quad (3.11)$$

Observe again that $\|\Phi(\cdot)\|$ can only be nondifferentiable at solutions of (1.1). However, if for infinitely many j the line segments $[u^{k_j}, u^{k_j} + \alpha_{k_j}v^{k_j}/\theta]$ contain points \tilde{u}^{k_j} such that $\Phi(\tilde{u}^{k_j}) = 0$, then we obtain (3.8) yet again, since $\{\tilde{u}^{k_j}\} \rightarrow \bar{u}$ by $\{u^{k_j}\} \rightarrow \bar{u}$ and (3.10).

So let $\|\Phi(\cdot)\|$ take no zero values in the intervals $[u^{k_j}, u^{k_j} + \alpha_{k_j}v^{k_j}/\theta]$ for all j . In this case, $\|\Phi(\cdot)\|$ is differentiable on these sets, with the gradient of the form $(\Phi'(\cdot))^T\Phi(\cdot)/\|\Phi(\cdot)\|$. Employing the mean-value theorem [17, Theorem A.10(a)], we then obtain from (3.11) that for each j there exists $t_{k_j} \in [0, 1]$ such that

$$\left\langle \frac{(\Phi'(u^{k_j} + t_{k_j}\alpha_{k_j}v^{k_j}/\theta))^T\Phi(u^{k_j} + t_{k_j}\alpha_{k_j}v^{k_j}/\theta)}{\|\Phi(u^{k_j} + t_{k_j}\alpha_{k_j}v^{k_j}/\theta)\|}, v^{k_j} \right\rangle > -\sigma\|\Phi(u^{k_j})\|. \quad (3.12)$$

If the sequence $\{v^{k_j}\}$ of Newton directions were to be unbounded, condition (3.4) implies that

$$\liminf_{j \rightarrow \infty} \|\Phi(u^{k_j})\| = 0.$$

This, in view of the monotonicity of the sequence in question, again yields (3.8). Therefore, it remains to consider the case when $\{v^{k_j}\}$ is bounded. Taking a further subsequence, if necessary, we may assume that $\{v^{k_j}\}$ converges to some \tilde{v} , and therefore, by (1.2),

$$\Phi(\bar{u}) = -\Phi'(\bar{u})\tilde{v}. \quad (3.13)$$

Suppose that $\Phi(\bar{u}) \neq 0$. Then by (3.10) and (3.13), passing onto the limit in (3.12) as $j \rightarrow \infty$, we obtain that

$$-\|\Phi(\bar{u})\| = -\frac{\langle \Phi(\bar{u}), \Phi(\bar{u}) \rangle}{\|\Phi(\bar{u})\|} = \left\langle \frac{\Phi(\bar{u})}{\|\Phi(\bar{u})\|}, \Phi'(\bar{u})\tilde{v} \right\rangle = \left\langle \frac{(\Phi'(\bar{u}))^T \Phi(\bar{u})}{\|\Phi(\bar{u})\|}, \tilde{v} \right\rangle \geq -\sigma \|\Phi(\bar{u})\|,$$

in contradiction with $\Phi(\bar{u}) \neq 0$.

It remains to consider the case when there exists an infinite subsequence $\{u^{k_j}\}$ convergent to some \bar{u} , and such that the Newton direction is not accepted by Algorithm 3.1 at every point of this subsequence. In this case, the iterates $\{u^{k_{j+1}}\}$ are generated by the gradient steps with Armijo linesearch for the merit function (3.2). That being the case, (3.7) follows by standard argument (see, e.g., [1, Proposition 1.2.3]). \blacksquare

Remark 3.1 An alternative to using the safeguard gradient direction of the (squared) residual in our context is the Levenberg–Marquardt direction (see [13, 14]), i.e., v^k being the (unique) solution of the linear system

$$(\Phi'(u^k))^T \Phi(u^k) + ((\Phi'(u^k))^T \Phi'(u^k) + \rho_k I)v = 0, \quad (3.14)$$

where $\rho_k > 0$ is the regularization parameter. Employing (3.3), for this direction we have that

$$\begin{aligned} \langle \varphi'(u^k), v^k \rangle &= \langle (\Phi'(u^k))^T \Phi(u^k), v^k \rangle \\ &= -\langle ((\Phi'(u^k))^T \Phi'(u^k) + \rho_k I)v^k, v^k \rangle \\ &\leq -\rho_k \|v^k\|^2. \end{aligned}$$

In particular, the Levenberg–Marquardt direction v^k is always of descent for the merit function defined in (3.2), unless $v^k = 0$ (in which case (3.6) holds). With proper choices of the parameter ρ_k , global convergence analysis of the corresponding linesearch algorithms can be found in [5, 21], and it can be easily combined with the one in Theorem 3.1 above, leading to the same conclusions.

Algorithm 3.1 allows to incorporate extrapolation or overrelaxation, in a way discussed above for Algorithm 2.1. This is precisely what we shall do in the numerical testing in the next section.

4 Numerical results

We next provide some numerical results demonstrating the global performance of Algorithm 3.1 and the effect of accelerating techniques. We note that, to the best of our knowledge, there are no studies in the literature comparing various acceleration options, not even in the local setting.

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 figures):

- NM (for “Newton Method”) is Algorithm 3.1 without any modifications, and with parameter values $C = 10^7$, $\tau = 2$, $\sigma = 0.01$, and $\theta = 0.5$.
- NM-EP (for “ExtraPolation”) is Algorithm 3.1, generating also an auxiliary sequence $\{\widehat{u}^k\}$ according to (3.1).
- NM-OR2 (for “2-step OverRelaxation”) is Algorithm 3.1, but with v^k obtained at Step 2 replaced (when accepted) by $(2 - \|v^k\|^{1/2})v^k$ for all $k = \widehat{k} + 2i - 1$, $i = 1, 2, \dots$, where the choice of \widehat{k} will be discussed below.
- NM-OR3 (for “3-step OverRelaxation”) is Algorithm 3.1, but with v^k obtained at Step 2 replaced (when accepted) by $2v^k$ for all $k = \widehat{k} + 3i - 2$, $i = 1, 2, \dots$
- LM (for “Levenberg–Marquardt method”) can be seen as Algorithm 3.1 without Steps 2 and 4, and with v^k at Step 3 replaced by the solution of (3.14) (see Remark 3.1), with the following choice of the regularization parameter:

$$\rho_k = \frac{\|\Phi(u^k)\|^2}{1 + \|\Phi(u^k)\|^2}.$$

Locally this rule agrees with the one in [21], while the form of the denominator allows to avoid large values of this parameter, potentially blocking long steps of the algorithm.

The number \widehat{k} for OR2 and OR3 is the first iteration number $k = 1, 2, \dots$ such that $\alpha_k = \alpha_{k-1} = 1$, and the following tests are passed:

$$\left\| \frac{v^k}{\|v^k\|} - \frac{v^{k-1}}{\|v^{k-1}\|} \right\| \leq \varepsilon_{\text{stab}},$$

$$\left\| \Phi'(u^k) \frac{v^k}{\|v^k\|} \right\| \leq \varepsilon_{\text{ker}},$$

$$\left| \frac{\|v^k\|}{\|v^{k-1}\|} - \frac{1}{2} \right| \leq \varepsilon_{\text{rate}},$$

where the tolerances were taken as $\varepsilon_{\text{stab}} = \varepsilon_{\text{ker}} = \varepsilon_{\text{rate}} = 0.1$.

Algorithms terminate, with success declared, after an iterate u^k is generated satisfying

$$\|\Phi(u^k)\| \leq 10^{-8}. \quad (4.1)$$

The only exception is EP, for which, for $k = 1, 2, \dots$, we first generate \hat{u}^k , and terminate with success if

$$\|\Phi(\hat{u}^k)\| \leq 10^{-8}. \quad (4.2)$$

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 when termination happened because of (4.2)). If successful termination did not occur after 100 iterations, or any of the backtracking procedures in Algorithm 3.1 produced a trial value α such that $\alpha\|v^k\| \leq 10^{-10}$, the process was terminated declaring failure.

We report on the results for a set of nonlinear equations possessing singular solutions, obtained the following way. We used all test problems from [18] for which an exact solution is known, except for the three linear problems **Linear - full rank**, **Linear - rank 1**, and **Linear - rank 1 with zero columns and rows**. Some of thus selected problems have more equations than variables, in which cases some extra equations were removed (always linear, except for **Beal** for which the choice of an equation to be removed does not seem to affect the performance in any significant way). The resulting test problems are listed in Table 1, supplied with information on dimensions of the problems, on removed equations, and on the choice of starting points when nonstandard. All other test problems were used with standard starting points provided in [18]. Nonstandard starting points were used for **Freudenstein and Roth** and for **Helical valley** in order to force the algorithm to converge to singular solutions; the specific choices were $u^0 = (4.5, 3.5)$ and $u^0 = (2, 1, 1)$, respectively.

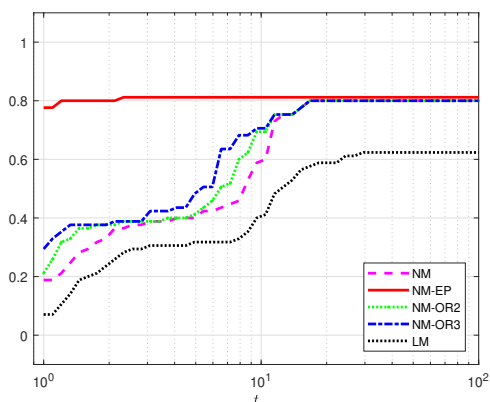
Furthermore, using the technique proposed in [19], selected test problems were transformed into systems of equations of the same sizes, in such a way that a known solution \bar{u} remains a solution of the modified system, but the rank of the Jacobian of the modified system at \bar{u} becomes equal to $n - 1$. This transformation was not applied to **Powell singular**, **Extended Powell singular**, and **Variably dimensioned**, as these test problems possess known singular solutions in their original form.

In order to get more statistics, apart from the starting points u^0 discussed above, we have also used $-10u^0$, $-u^0$ and $10u^0$, $100u^0$ for the choices of u^0 resulting in convergence to the solution of interest.

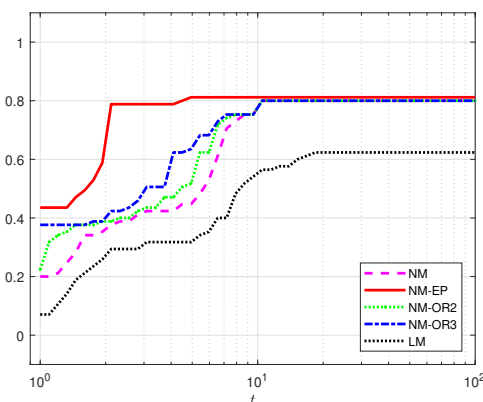
The obtained results are presented in Figure 1, in the form of performance profiles [2]. In Figure 1a, the measure of efficiency is the iteration count, while in Figure 1b, it is the number of evaluations of Φ (for successful runs). These figures demonstrate the accelerating effect of OR, and especially of EP. Moreover, the use of the former does

Table 1: Selected test problems from [18]

Test	n	Comments
1. Rosenbrock	2	
2. Freudenstein and Roth	2	Nonstandard u^0
4. Brown badly scaled	2	Eq. # 2 removed
5. Beale	2	Eq. # 2 removed
7. Helical valley	3	Nonstandard u^0
11. Gulf research and development	3	
12. Box three-dimensional	3	
13. Powell singular	4	
14. Wood	4	Eqs. # 2, 4 removed
18. Biggs EXP6	6	
21. Extended Rosenbrock	500	
22. Extended Powell singular	500	
25. Variably dimensioned	10, 500	Eqs. # $n - 2, n - 1$ removed
26. Trigonometric	30	
27. Brown almost-linear	10, 500	



(a) Iterations



(b) Evaluations

Figure 1: Performance profiles: number of iterations and evaluations of Φ .

not alter robustness of NM, which is somewhat surprising, considering that our ways of combining OR with NM do not rely on any strict theoretical justification. On the other hand, the accelerating effect of EP is much higher than that of OR, which might indicate that OR should be combined with NM in some other (smarter) way. This will be the subject of our future research.

As for the number of evaluations of Φ , observe that if, say, the unit stepsize is

accepted at some iteration of NM, and the corresponding iteration of NM-EP does not end up with termination, the latter requires two evaluations of Φ , while the former only one. This is the main reason why the effect of EP on efficiency in this sense is somehow less strong than on iteration count, on this test set. However, the effect on efficiency is still evident.

Let us underline again that Newton-type methods (like NM and LM) may not show a superlinear or quadratic rate of convergence if, as in our experiments, they operate in a neighborhood of a critical solution. Regardless of this, the reason for the difference of performance between NM and LM in Figure 1 is not fully understood at this time; it might depend on globalization techniques and is a subject for further research.

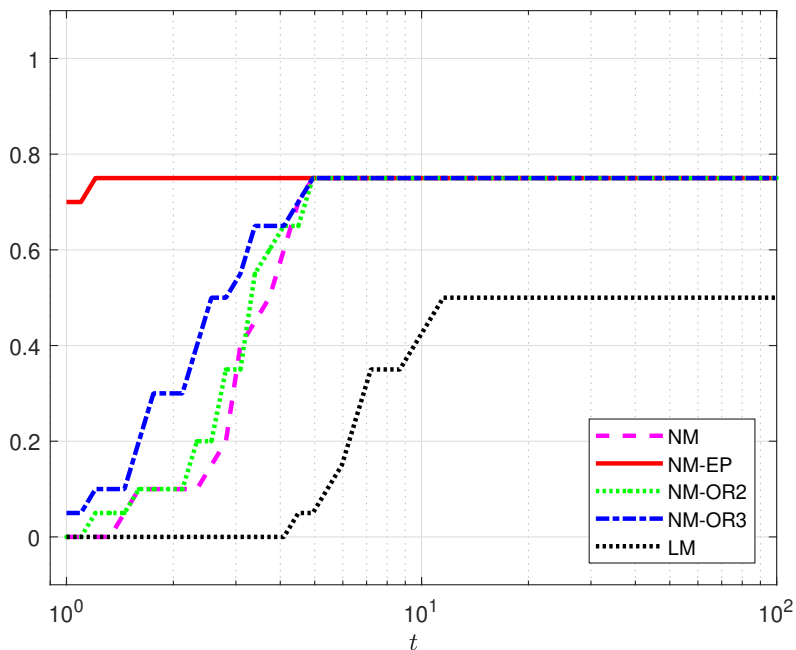


Figure 2: Performance profiles: CPU times for selected problems.

Finally, in Figure 2 we present results in terms of CPU times for problem instances Extended Rosenbrock, Extended Powell singular, Variably dimensioned, and Brown almost-linear, all with $n = 500$. Since we perform 5 runs for each problem, it makes 20 runs altogether. We do not report CPU times for other instances, because for small(er) problems this characteristic can hardly be reliable or informative.

Acknowledgments

This research was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 409756759, by the Russian Foundation for Basic Research Grants 19-51-12003 NNIO_a and 20-01-00106, by CNPq Grant 303913/2019-3, by FAPERJ Grant E-26/202.540/2019, by PRONEX–Optimization, and by Volkswagen Foundation. The authors thank Ivan Rodin and Dmitriy Bannikov for their assistance with numerical experiments, and the two anonymous referees for helpful comments on the original version of the paper.

References

- [1] D.P. Bertsekas. *Nonlinear Programming*. 2nd edition. Athena, Belmont, 1999.
- [2] E.D. Dolan and J.J. More. Benchmarking optimization software with performance profiles. *Math. Program.* 91 (2002), 201–213.
- [3] F. Facchinei, A. Fischer, and M. Herrich. A family of Newton methods for non-smooth constrained systems with nonisolated solutions. *Math. Methods Oper. Res.* 77 (2013), 433–443.
- [4] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions. *Math. Program.* 146 (2014), 1–36.
- [5] J.-Y. Fan and Y.-X. Yuan. On the quadratic convergence of the Levenberg–Marquardt method. *Computing.* 74 (2005), 23–39.
- [6] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program.* 125 (2010), 47–73.
- [7] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program.* 94 (2002), 91–124.
- [8] A. Fischer, A.F. Izmailov, and M.V. Solodov. Unit stepsize for the Newton method close to critical solutions. *Math. Program.* (2020). <https://doi.org/10.1007/s10107-020-01496-z>.
- [9] A. Griewank. *Analysis and modification of Newton’s method at singularities*. PhD Thesis. Australian National University, Canberra, 1980.
- [10] A. Griewank. Starlike domains of convergence for Newton’s method at singularities. *Numer. Math.* 35 (1980), 95–111.

- [11] A. Griewank. On solving nonlinear equations with simple singularities or nearly singular solutions. *SIAM Rev.* 27 (1985), 537–563.
- [12] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. *Math. Program.* 122 (2012), 93–120.
- [13] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Appl. Math.* 2 (1944), 164–168.
- [14] D.W. Marquardt. An algorithm for least squares estimation of non-linear parameters. *SIAM J.* 11 (1963), 431–441.
- [15] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Critical solutions of nonlinear equations: Local attraction for Newton-type methods. *Math. Program.* 167 (2018), 355–379.
- [16] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Critical solutions of nonlinear equations: Stability issues. *Math. Program.* 168 (2018), 475–507.
- [17] A.F. Izmailov and M.V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering, Springer International Publishing, Switzerland, 2014.
- [18] J.J. Moré, B.S. Garbow, and K.E. Hillstom. Testing unconstrained optimization software. *ACM Trans. Math. Softw.* 7 (1981), 17–41.
- [19] R.B. Schnabel and P.D. Frank. Tensor methods for nonlinear equations. *SIAM J. Numer. Anal.* 21 (1984), 815–843.
- [20] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.* 11 (1998), 253–275.
- [21] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg–Marquardt method. In: G. Alefeld and X. Chen (eds.) *Topics in Numerical Analysis*, Springer, Vienna, 2001, pp. 239–249.