LOCAL CONVERGENCE OF NEWTON-TYPE METHODS WITH CUBIC REGULARIZATION FOR CONSTRAINED NONLINEAR EQUATIONS*

Alexey F. Izmailov[1] and Mikhail V. Solodov[2]

July 21, 2026

ABSTRACT

We present results on local quadratic convergence of the constrained Gauss–Newton method with cubic regularization, for nonlinear equations with the variables restricted to a given closed convex set. The key assumption is the constrained local Lipschitzian error bound allowing, in particular, solutions to be nonisolated. We also specify the level of inexactness in solving the subproblems, which preserves the local quadratic convergence rate. For the special case of an unconstrained system with the number of equations equal to the number of variables, but assuming that the solution in question is singular (in particular, the error bound condition is typically violated), we establish conditions ensuring linear convergence of the method, with the exact asymptotic common ratio 1/2.

Key words: nonlinear equation; constrained equation; Gauss-Newton method; cubic regularization; error bound; singular solution; 2-regularity; superlinear convergence; linear convergence.
AMS subject classifications. 65J15, 49M15.

---

* Research of the second author was supported by CNPq Grant 306775/2023-9 and by FAPERJ Grant E-26/200.165/2026.

[1] VMK Faculty, OR Department, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory, 119991 Moscow, Russia.
Email: izmaf@cs.msu.ru

[2] 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

Constrained equations of the form

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

with $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ being a given smooth mapping, and $P \subset \mathbb{R}^p$ a given nonempty closed convex set, is a widely used modeling paradigm with multiple applications. Constraints may arise in the equations setting as a natural part of the model itself, when, say, the values of the variable u, which do not belong to the domain P, make no "physical" sense. Constraints also often arise in reformulations of the model, either needed for the equivalence of such reformulations to the original problem, or for enforcing some properties needed for developing computational methods with desirable convergence guarantees. See, e.g., [1] for a discussion of various reformulations of systems involving complementarity conditions, which lead to (1.1). This problem setting includes, in particular, the KKT-type systems for generalized Nash equilibrium problems [2] and mathematical programs with complementarity constraints [3]. Other applications of constrained equations can be found in [4, 5].

For a general problem (1.1), without imposing any relations between p and q, and/or assuming any structure of Φ and/or P, a natural computational approach to solve this problem consists of iteratively minimizing over P the (squared) residual of the linearized equation in (1.1). This gives the (constrained) Gauss–Newton method. For the current iterate $u \in P$, it defines the next one as $u + v$, where v is a solution of the optimization problem

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

The Frank–Wolfe Theorem [6] implies that (1.2) always has a solution when P is polyhedral, but its solution need not be unique. For a recent discussion of the Gauss–Newton methods and some new results about them, see [7]. One popular approach to overcome the potentially problematic issue of non-uniqueness in (1.2) is regularization. This leads to the Levenberg–Marquardt method, the modern convergence theories for which were recently surveyed in [8]. It consists of adding to the objective function of (1.2) a regularization term given the squared norm of v, multiplied by the regularization parameter defined by a function $\sigma : P \to \mathbb{R}_+$. The resulting subproblem is then given by

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

An alternative possibility is the cubic regularization approach, which in the unconstrained optimization context comes from [9, 10]. Specifically, for the problem

$$\text{minimize} \quad f(u), \quad u \in \mathbb{R}^p,$$

with twice differentiable $f : \mathbb{R}^p \to \mathbb{R}$, the subproblem of the method defining the displacement v at a current iterate $u \in \mathbb{R}^p$ has the form

$$\text{minimize} \quad f(u) + \langle f'(u),\, v\rangle + \frac{1}{2}\langle f''(u)v,\, v\rangle + \frac{1}{3}\sigma(u)\|v\|^3, \quad v \in \mathbb{R}^p. \tag{1.3}$$

Further developments of this approach for optimization problems, with the main emphasis on complexity estimates, include the adaptive cubic regularization algorithm (ARC) [11] and its variants in multiple subsequent publications; see, e.g., [12, 13] for some recent advances.

In this paper, we consider a technique of this kind, but for the constrained equation (1.1): the next iterate is defined as $u + v$, where v solves the optimization problem

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{3}\sigma(u)\|v\|^3 \quad \text{subject to} \quad u + v \in P. \tag{1.4}$$

If $\sigma(u) > 0$, its is clear that the objective function in (1.4) is both coercive and strictly convex, and hence, this subproblem has the unique solution. We note that this is different from cubic regularization in (nonconvex) unconstrained optimization, where the objective functions of the subproblems (1.3) need not be even convex. In fact, it is important to emphasize that the iterative scheme given by (1.4) does not correspond to any cubic regularization method for optimization. In particular, if we take $f(u) = \|\Phi(u)\|^2/2$, the objective function in (1.3) would involve second derivatives of Φ, while (1.4) only employs its first derivative. However, in the unconstrained case, the structure of the subproblem (1.4) is the same as that in (1.3) for unconstrained optimization, and hence the same techniques for (approximately) solving it can be applied; see, e.g., [10, 11, 12]. Perhaps, employing the additional feature that the objective function here is strictly convex might be benefitial as well. In the constrained case, the general-purpose methods for simply-constrained convex optimization problems can be applied, e.g., those discussed in [14, Section 5].

This paper is devoted to local convergence and rate-of-convergence analyses for the Gauss–Newton method with cubic regularization under mild assumptions. Specifically, in Section 2, we present the result on local quadratic convergence of the method in question under the constrained error bound condition allowing, in particular, for nonisolated solutions of (1.1). In Section 3, we study the restrictions on inexactness allowed when solving the subproblems, so that the local quadratic convergence rate is preserved. One particular advantage of the cubic regularization technique is that, as will be shown in Section 2, to guarantee superlinear convergence, unlike in the Levenberg–Marquardt method, there is no need to drive the regularization parameter $\sigma(u)$ in (1.4) to 0 as u approaches the solution set of (1.1). This means that, in a sense, the subproblems (1.4) can be kept "the same" along the iterations, not only qualitatively, but also "quantitatively". In Section 4, we consider the special case of an unconstrained system with the number of equations equal to the number of variables (i.e., $P = \mathbb{R}^p$ and $p = q$), but assuming that the solution in question is singular. In particular, the error bound condition is typically violated in this case. In this setting, we provide conditions ensuring linear convergence of the Gauss–Newton method with cubic regularization, with the exact asymptotic common ratio 1/2.

For unconstrained optimization, local quadratic convergence of the cubic regularization method under the error bound condition (thus allowing for nonisolated solutions) was established in [15]. Some equivalent interpretations and sufficient conditions for the error bound were also proposed in that work. This line of analysis was continued in [16] in the context of optimization on manifolds, for the method with adaptive choice of regularization parameters, which is a variant of ARC algorithm from [11]. In [15], regularization parameters are bounded by their choice in the algorithm, while in [11] and [16] establishing boundedness is an ingredient of the rate-of-convergence analysis for ARC. Therefore, these developments agree with our requirements on σ in Theorem 2.2 below. However, we emphasize yet again that the method considered here is not a cubic regularization method for any optimization problem. Moreover, Theorem 2.2 below deals with the more general constrained case.

Some words are in order about our notation, even though it is quite standard. For a closed convex set $P \subset \mathbb{R}^p$, by $N_P(u)$ we denote the normal cone to P at u, i.e., $N_P(u) = \{v \in \mathbb{R}^p \mid \langle v, \widetilde{u} - u \rangle \leq 0, \forall \widetilde{u} \in P\}$ if $u \in P$ and $N_P(u) = \emptyset$ otherwise, where $\langle u, v \rangle$ is the Euclidean inner product of $u, v \in \mathbb{R}^p$. Let $\|\cdot\|$ stand for the Euclidean norm throughout. For a set $U \subset \mathbb{R}^p$ and a point $u \in \mathbb{R}^p$, by $\mathrm{dist}(u, U) = \inf_{v \in U} \|v - u\|$ we denote the distance from u to U, and by $B(u, \varepsilon) = \{v \in \mathbb{R}^p \mid \|v - u\| \leq \varepsilon\}$ the closed ball of radius $\varepsilon \geq 0$ centered at u. For a linear operator A, by $\ker A$ we denote its null space, and by $\mathrm{im}\, A$ its range space.

2 Superlinear convergence under the error bound condition

The analysis in this section relies on an abstract local convergence framework for constrained equations with possibly nonisolated solutions, developed in [17]. Consider a scalar constrained equation:

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

with $\varphi : \mathbb{R}^p \to \mathbb{R}_+$, and $P \subset \mathbb{R}^p$ is here assumed just nonempty and closed. For a moment, let U be the solution set of (2.1).

Theorem 2.1 below is a simplified version of [17, Theorem 2.1]. It deals with an abstract iterative process intended for solving (2.1), updating the current iterate $u \in P$ to $\Psi(u)$, where $\Psi : P \to P$ is a given mapping.

Theorem 2.1 *Let $\varphi : \mathbb{R}^p \to \mathbb{R}_+$ be a continuous function, $P \subset \mathbb{R}^p$ be a nonempty closed set, $\bar{u} \in U$, and assume that*

$$\varphi(u) = O(\mathrm{dist}(u, U)) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{2.2}$$

Moreover, let $\Psi : P \to P$ be a mapping such that

$$\Psi(u) - u = O(\varphi(u)) \quad \text{as } u \in P \text{ tends to } \bar{u}, \tag{2.3}$$

and

$$\varphi(\Psi(u)) = O((\varphi(u))^2) \quad \text{as } u \in P \text{ tends to } \bar{u}\,. \tag{2.4}$$

Then, for every $\varepsilon > 0$ small enough, and every $u^0 \in P$ close enough to $\bar{u}$, the sequence $\{u^k\}$ defined by $u^{k+1} = \Psi(u^k)$ for all k is contained in $B(\bar{u}, \varepsilon)$ and converges to some $u^* \in U$, with the rate of convergence being quadratic.

Our problem (1.1) can be equivalently stated in the form (2.1) by taking, for example, $\varphi(u) = \|\Phi(u)\|$. From now on, let U be the solution set of (1.1).

The key assumption used in this section is the constrained error bound condition:

$$\mathrm{dist}(u, U) = O(\|\Phi(u)\|) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{2.5}$$

The use of error bounds in convergence analyses of Newton-type methods for equations with potentially nonisolated solutions originated from [18]; see [8] for a survey of further developments of this kind.

For the cubic regularization scheme described by (1.4), we then obtain the following.

Theorem 2.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^q$ be a given mapping, $P \subset \mathbb{R}^p$ be a nonempty closed convex set, and assume that Φ is differentiable near $\bar{u} \in U$, with its derivative being Lipschitz-continuous near $\bar{u}$. Let the constrained error bound condition* (2.5) *be satisfied. Assume finally that the function $\sigma : P \to \mathbb{R}_+$ defining the regularization parameter in* (1.4) *satisfies $\sigma(u) > 0$ for all $u \in P \setminus U$, σ is bounded near $\bar{u}$, and*

$$\Phi(u) = O(\sigma(u)) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{2.6}$$

Then, for every $u^0 \in P$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the solution of (1.4) *with $u = u^k$, and with the additional convention that $u^{k+1} = u^k$ if $u^k \in U$. For any $\varepsilon > 0$ small enough, if $u^0 \in P$ is close enough to $\bar{u}$, then any such sequence is contained in $B(\bar{u}, \varepsilon)$ and converges to some $u^* \in U$, with the rate of convergence being quadratic.*

Proof. The assumption $\sigma(u) > 0$ for all $u \in P \setminus U$ implies that for any such u, the subproblem (1.4) has the unique solution $v(u)$, since its objective function is coercive and strictly convex. Moreover, according to the natural convention specified in the assertion of the theorem, for $u \in U$ the displacement $v(u)$ is 0. This yields the existence and uniqueness of the sequence $\{u^k\}$ in the assertion above.

We shall employ Theorem 2.1 with $\varphi(u) = \|\Phi(u)\|$ and $\Psi(u) = u + v(u)$. The assumption (2.2) in that theorem is satisfied because of the Lipschitz-continuity of Φ near $\bar{u}$. Furthermore, for $u \in U$, the asymptotic relations in (2.3)–(2.4) are obviously satisfied as well.

Consider now $u \in P \setminus U$. Then from (2.5) and (2.6) it follows that

$$\frac{(\mathrm{dist}(u, U))^4}{\sigma(u)} = (\mathrm{dist}(u, U))^3 O\left(\frac{\|\Phi(u)\|}{\sigma(u)}\right) = O((\mathrm{dist}(u, U))^3) \quad \text{as } u \to \bar{u}. \tag{2.7}$$

Let $\widehat{u}$ stand for any metric projection of u onto U:

$$\|u - \widehat{u}\| = \mathrm{dist}(u, U). \tag{2.8}$$

Since $v(u)$ is the global solution of (1.4), it holds that

$$\begin{aligned} & \frac{1}{2}\|\Phi(u) + \Phi'(u)v(u)\|^2 + \frac{1}{3}\sigma(u)\|v(u)\|^3 \\ \leq\ & \frac{1}{2}\|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 + \frac{1}{3}\sigma(u)\|\widehat{u} - u\|^3. \end{aligned} \tag{2.9}$$

Since $\bar{u} \in U$, it evidently holds that $\widehat{u} \to \bar{u}$ as $u \to \bar{u}$, and from (2.8)–(2.9), applying the Mean-Value Theorem (e.g., [19, Theorem A.10]), we derive that

$$\begin{aligned}
\|v(u)\|^3 &\leq \frac{3}{\sigma(u)} \left(\frac{1}{2} \|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 + \frac{1}{3} \sigma(u) \|\widehat{u} - u\|^3 \right) \\
&= \frac{3}{2\sigma(u)} \|\Phi(u) - \Phi(\widehat{u}) - \Phi'(u)(u - \widehat{u})\|^2 + \|u - \widehat{u}\|^3 \\
&\leq \frac{3}{2\sigma(u)} \sup_{t \in [0,\, 1]} \|\Phi'(tu + (1-t)\widehat{u}) - \Phi'(u)\|^2 \|u - \widehat{u}\|^2 + \|u - \widehat{u}\|^3 \\
&= O\left(\frac{\|u - \widehat{u}\|^4}{\sigma(u)} \right) + \|u - \widehat{u}\|^3 \\
&= O((\mathrm{dist}(u,\, U))^3) \quad \text{as } u \to \bar{u}, \qquad (2.10)
\end{aligned}$$

where the last estimate is by (2.7). Hence, employing (2.5), we obtain that

$$v(u) = O(\mathrm{dist}(u,\, U)) = O(\|\Phi(u)\|) \quad \text{as } u \to \bar{u}, \qquad (2.11)$$

which completes the justification of (2.3).

Furthermore, any $u \in U$ is a (global) solution of the optimization problem

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

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

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni 0. \qquad (2.12)$$

Employing now [20, Lemma 1] (this lemma states that the constrained error bound implies the upper-Lipschitzian property of the solutions set of the generalized equation (2.12) subject to the right-hand side perturbations), we obtain that under (2.5), for any solution u of the perturbed generalized equation

$$(\Phi'(u))^\top \Phi(u) + N_P(u) \ni \omega, \qquad (2.13)$$

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

$$\mathrm{dist}(u,\, U) = O(\|\omega\|) \quad \text{as } \omega \to 0.$$

The first-order necessary optimality condition for the subproblem (1.4) has the form

$$(\Phi'(u))^\top (\Phi(u) + \Phi'(u)v) + \sigma(u)\|v\|v + N_P(u+v) \ni 0, \qquad (2.14)$$

and (2.12) equivalent to saying that

$$(\Phi'(u+v))^\top \Phi(u+v) + N_P(u+v) \ni 0 \qquad (2.15)$$

holds with $v = 0$. Then (2.14) can be regarded as a perturbation of the generalized equation (2.15). Specifically, if we define

$$\begin{aligned} \omega(u,\, v) &= (\Phi'(u+v))^\top \Phi(u+v) - (\Phi'(u))^\top (\Phi(u) + \Phi'(u)v) - \sigma(u)\|v\|v \\ &= \Big((\Phi'(u+v))^\top - (\Phi'(u))^\top\Big)\,\Phi(u) \\ &\quad + \Big((\Phi'(u+v))^\top - (\Phi'(u))^\top\Big)\,(\Phi(u+v) - \Phi(u)) \\ &\quad + (\Phi'(u))^\top (\Phi(u+v) - \Phi(u) - \Phi'(u)v) - \sigma(u)\|v\|v, \end{aligned} \tag{2.16}$$

then (2.14) takes the form

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

From the Lipschitz-continuity of Φ' (and hence, also of Φ) near $\bar{u}$, from the assumption that σ is bounded near $\bar{u}$, and from (2.11), employing again the Mean-Value Theorem [19, Theorem A.10], we derive that

$$\begin{aligned} \omega(u,\, v(u)) &= O(\|v(u)\|\|\Phi(u)\|) + O(\|v(u)\|^2) + O(\sigma(u)\|v(u)\|^2) \\ &= O(\|\Phi(u)\|^2) \quad \text{as } u \to \bar{u}. \end{aligned} \tag{2.17}$$

Therefore, $u + v(u)$ is a solution of the generalized equation (2.13) with $\omega = \omega(u,\, v(u))$, and it holds that $u + v(u) \to \bar{u}$ and $\omega(u,\, v(u)) \to 0$ as $u \in P \setminus U$ tends to $\bar{u}$. Thus, using [20, Lemma 1] we conclude that

$$\operatorname{dist}(u + v(u),\, U) = O(\omega(u,\, v(u))) = O(\|\Phi(u)\|^2) \quad \text{as } u \to \bar{u},$$

where the second estimate is by (2.17). Therefore, since Φ is Lipschitz-continuous near $\bar{u}$, it holds that

$$\Phi(u + v(u)) = O(\operatorname{dist}(u + v(u),\, U)) = O(\|\Phi(u)\|^2) \quad \text{as } u \to \bar{u}, \tag{2.18}$$

completing the justification of (2.4).

The announced conclusions now follow by applying Theorem 2.1. ∎

Observe that the assumptions on σ in Theorem 2.2 allow to take it as any positive constant. In particular, σ need not tend to zero.

The following example demonstrates that Theorem 2.2 is not valid in the absence of the assumption that σ is bounded near $\bar{u}$.

Example 2.1 Let $p = q = 1$, $\Phi(u) = u$, $P = \mathbb{R}$. Then the only solution of the equation in (1.1) is $\bar{u} = 0$. This solution is nonsingular ($\Phi'(\cdot) \equiv 1$), and the error bound (2.5) holds.

For any $\sigma > 0$, the unique solution of subproblem (1.4) is given by

$$v = \begin{cases} \dfrac{1 - \sqrt{1 + 4\sigma u}}{2\sigma} & \text{if } u > 0, \\ \dfrac{-1 + \sqrt{1 - 4\sigma u}}{2\sigma} & \text{if } u < 0. \end{cases}$$

Let $\sigma(u) = 1/|u|^{\nu}$ for $u \neq 0$, with $\nu > 0$, making $\sigma(\cdot)$ unbounded near $\bar{u}$. For $\nu < 1$, after some elementary computations we obtain

$$u + v = \mathrm{sign}(u)\frac{1}{4}|u|^{2-\nu} + O(|u|^{3-2\nu}) \quad \text{as } u \to \bar{u},$$

thus yielding superlinear but not quadratic convergence of the iterates. For $\nu = 1$,

$$u + v = \frac{3 - \sqrt{5}}{2}u,$$

and convergence is linear. Finally, for $\nu > 1$,

$$u + v = u + O(|u|^{(1+\nu)/2}) \quad \text{as } u \to \bar{u},$$

and convergence is at best sublinear.

The observations above demonstrate that Theorem 2.2 is not valid in the absence of the assumption that σ is bounded near $\bar{u}$. On the other hand, if σ is bounded near $\bar{u}$, then

$$u + v = O(u^2) \quad \text{as } u \to \bar{u},$$

yielding quadratic convergence, which agrees with the assertion of Theorem 2.2, even when (2.6) is violated.

The next example exhibits that Theorem 2.2 is not valid in the absence of the error bound (2.5).

Example 2.2 Let $p = q = 1$, $\Phi(u) = u^2$, $P = \mathbb{R}$. Then the only solution of the equation in (1.1) is $\bar{u} = 0$, this solution is singular ($\Phi'(\bar{u}) = 0$), and the error bound (2.5) does not hold.

For any $\sigma > 0$, the unique solution of subproblem (1.4) is given by

$$v = \begin{cases} \dfrac{4u^2 - \sqrt{16u^4 + 8\sigma u^3}}{2\sigma} & \text{if } u > 0, \\[2ex] \dfrac{-4u^2 + \sqrt{16u^4 - 8\sigma u^3}}{2\sigma} & \text{if } u < 0. \end{cases}$$

If σ is fixed, it holds that

$$u + v = u - \mathrm{sign}(u)\sqrt{\frac{2}{\sigma}}|u|^{3/2} + O(|u|^3) \quad \text{as } u \to \bar{u}.$$

This evidently implies that convergence of the iterates to $\bar{u}$ is at best sublinear, demonstrating, in particular, that Theorem 2.2 is not valid in the absence of the error bound (2.5).

If $\sigma(u) = \chi|\Phi(u)|^{\theta} = \chi|u|^{2\theta}$ with any fixed $\chi > 0$ and $\theta > 1/2$, then

$$u + v = \frac{1}{2}u + O(|u|^{2\theta}) = \frac{1}{2}u + o(|u|) \quad \text{as } u \to \bar{u},$$

yielding linear convergence with asymptotic common ratio 1/2.

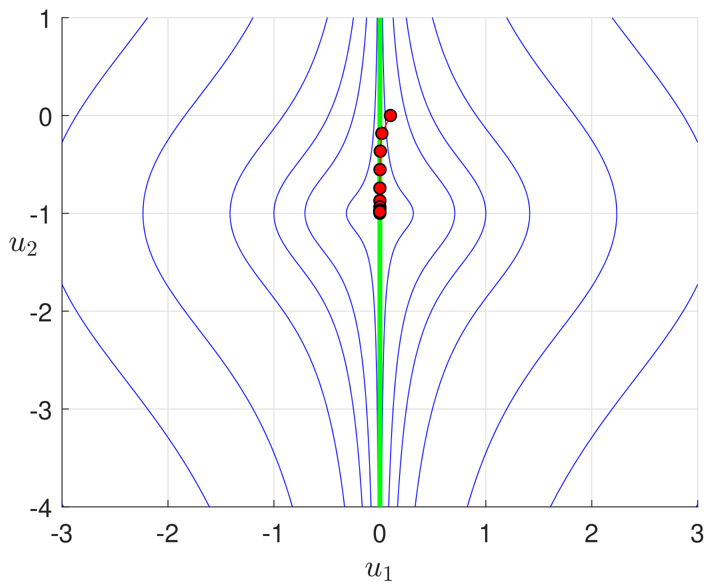


Figure 1: Gauss–Newton method with cubic regularization for Example 2.3

If $\sigma(u) = \chi|\Phi(u)|^{1/2} = |u|$, then

$$u + v = \left(1 + \frac{2}{\chi}\left(1 - \sqrt{1 + \frac{\chi}{2}}\right)\right) u,$$

also yielding linear convergence, but with asymptotic common ratio $\rho(\chi) = 1 + 2(1 - \sqrt{1 + \chi/2})/\chi \in (1/2,\, 1)$ for any $\chi > 0$. This function ρ is monotonically increasing on $(0,\, +\infty)$, with

$$\lim_{\chi \to 0+} \rho(\chi) = \frac{1}{2}, \quad \lim_{\chi \to +\infty} \rho(\chi) = 1.$$

We complete this section with a discussion of the restriction (2.6) on the regularization parameter $\sigma(u)$, not allowing it to go to 0 too fast as u tends to $\bar{u}$. In particular, if $\sigma(u) = \chi|\Phi(u)|^{\theta}$ with some fixed $\chi > 0$ and θ, this restriction does not allow the values $\theta > 1$. This is similar to the situation for the Levenberg–Marquardt method: the local convergence theory presented in [8, Section 3] does not allow to take $\theta > 2$, and moreover, [8, Example 3.1] demonstrates that at least $\theta \geq 4$ cannot be taken indeed for the local convergence result in [8, Theorem 3.2] to hold. (The question regarding the values $\theta \in (2,\, 4)$ remains open.) For the Gauss–Newton method with cubic regularization, the cited example provides numerical (though not analytical) evidence that the claim of Theorem 2.2 may not hold at least for $\theta \geq 3.5$; see Example 2.3 below. (We cannot say anything definite about the values of $\theta \in (1,\, 3.5)$.)

Example 2.3 Let $p = q = 2$, $\Phi(u) = (2u_1(1 + u_2),\, u_1^2)$, $P = \mathbb{R}^2$. Then the solution set of the equation in (1.1) is $U = \{0\} \times \mathbb{R}$, and the error bound (2.5) holds near any point of this set except for the solution $\bar{u} = (0,\, -1)$.

Numerical runs demonstrate that for $\theta = 3.5$ and larger, the Gauss–Newton method with cubic regularization has a tendency to converge to this $\bar{u}$, even when started far away from this solution, but close to some other $\widetilde{u} \in U$. One such run starting from the point $u^0 = (0.1,\, 0)$ is shown in Figure 1, where the thick vertical line is the solution set U, and the thin lines are the level curves of the equation's residual. The iterates do not stay close to $\widetilde{u}$, as is claimed in Theorem 2.2 under its assumptions (here violated). Moreover, the rate of convergence is only linear, with the asymptotic common ratio 1/2, and this agrees with Theorem 4.2 below.

3 Inexact Solution of Subproblems

We next characterize the level of inexactness that might be allowed when solving the subproblems (1.4), so that the local convergence and rate of convergence properties of the Gauss–Newton method with cubic regularization, established in Theorem 2.2, are preserved.

Let the process of solving the subproblem (1.4) be terminated once its optimality condition (2.14) is satisfied approximately, in the following sense:

$$(\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) + \sigma(u)\|v\|v + N_P(u+v) \ni w\,, \tag{3.1}$$

with the error term $w \in \mathbb{R}^p$ smaller by norm than some given tolerance. One natural condition for this tolerance is to conform with $\sigma(u)$ and with the residual of the equation in (1.1), as follows:

$$w = O(\sigma(u)\|\Phi(u)\|^2) \quad \text{as } u \in P \text{ tends to } \bar{u}. \tag{3.2}$$

For one practical procedure for computing a "small by norm" w satisfying (3.1) along iterations of any convergent algorithm for solving the (variational inequality) subproblems like (2.14), see [21, Section 2.2].

Observe that (3.1) is a necessary and sufficient optimality condition for the following perturbation of the subproblem (1.4):

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{3}\sigma(u)\|v\|^3 - \langle w,\, v\rangle \quad \text{subject to} \quad u + v \in P. \tag{3.3}$$

Whatever is w, if $\sigma(u) > 0$, the objective function of this problem is still strictly convex and coercive, and hence, (3.3) has the unique solution $v(u)$.

Under the assumptions of Theorem 2.2, using again any metric projection $\widehat{u}$ of $u \in P \setminus U$ onto U, similarly to (2.9) we obtain that

$$\begin{aligned}\frac{1}{2}\|\Phi(u) + \Phi'(u)v(u)\|^2 + \frac{1}{3}\sigma(u)\|v(u)\|^3 - \langle w,\, v(u)\rangle \quad &\le \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)(\widehat{u} - u)\|^2 \\ &\quad + \frac{1}{3}\sigma(u)\|\widehat{u} - u\|^3 - \langle w,\, \widehat{u} - u\rangle.\end{aligned}$$

Then, similarly to (2.10) (and in particular, making use of (2.7)–(2.8)), we derive the following estimate:

$$
\begin{aligned}
\|v(u)\|^3 &\leq \frac{3}{\sigma(u)}\left(\frac{1}{2}\|\Phi(u)+\Phi'(u)(\widehat{u}-u)\|^2+\frac{1}{3}\sigma(u)\|\widehat{u}-u\|^3+\langle w,\, v(u)\rangle-\langle w,\, \widehat{u}-u\rangle\right)\\
&\leq \frac{3\|w\|}{\sigma(u)}(\|v(u)\|+\operatorname{dist}(u,\, U))+O((\operatorname{dist}(u,\, U))^3)\\
&= O(\|\Phi(u)\|^2(\|v(u)\|+\operatorname{dist}(u,\, U)))+O((\operatorname{dist}(u,\, U))^3)\\
&= O((\operatorname{dist}(u,\, U))^2\|v(u)\|)+O((\operatorname{dist}(u,\, U))^3) \quad \text{as } u\to\bar{u}, \qquad (3.4)
\end{aligned}
$$

where the next-to-last estimate is by (3.2), while the last one is by the Lipschitz-continuity of Φ near $\bar{u}$. Evidently, (3.4) implies (2.11). Indeed, if we were to assume that

$$
\frac{\|v(u^k)\|}{\operatorname{dist}(u^k,\, U)}\to\infty \quad \text{as } k\to\infty
$$

for some sequence $\{u^k\}\subset P$ converging to $\bar{u}$, then dividing by $\|v(u^k)\|^3$ both sides of (3.4) with $u=u^k$, we obtain

$$
1=O\left(\frac{(\operatorname{dist}(u^k,\, U))^2}{\|v(u^k)\|^2}\right)+O\left(\frac{(\operatorname{dist}(u^k,\, U))^3}{\|v(u^k)\|^3}\right)\to 0 \quad \text{as } k\to\infty,
$$

yielding a contradiction.

Following the remaining part of the proof of Theorem 2.2, but with $\omega(u,\, v)$ defined in (2.16) substituted by $\omega(u,\, v)+w$, and employing (3.2) and the assumption that σ is bounded near $\bar{u}$, we arrive to the following result.

Theorem 3.1 *Under the assumptions of Theorem 2.2, let the function $\psi : P\to\mathbb{R}_+$ satisfy $\psi(u)=O(\sigma(u)\|\Phi(u)\|^2)$ as $u\in P$ tends to $\bar{u}$.*

Then, for every $u^0\in P$, there exists a sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1}-u^k$ is the solution of (3.1) with $u=u^k$, with some $w\in\mathbb{R}^p$ satisfying $\|w\|\leq\psi(u^k)$, and with the additional convention that $u^{k+1}=u^k$ if $u^k\in U$. For any $\varepsilon>0$, if $u^0\in P$ is close enough to $\bar{u}$, such sequence is contained in $B(\bar{u},\, \varepsilon)$ and converges to some $u^*\in U$, with the rate of convergence being quadratic.

Example 2.1 continued For any $\sigma>0$, the unique solution of problem (3.1) is given by

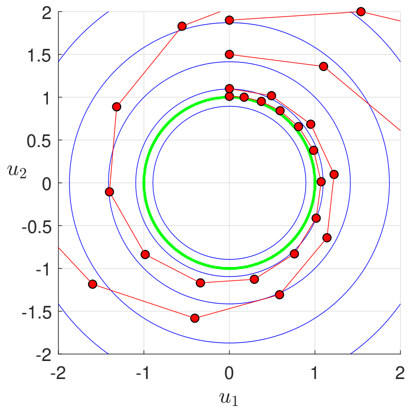
$$
v=\begin{cases}
\dfrac{1-\sqrt{1+4\sigma(u-w)}}{2\sigma} & \text{if } u\geq w,\\[2ex]
\dfrac{-1+\sqrt{1-4\sigma(u-w)}}{2\sigma} & \text{if } u<w.
\end{cases}
$$

If $\sigma(\cdot)$ is bounded near $\bar{u}$, then

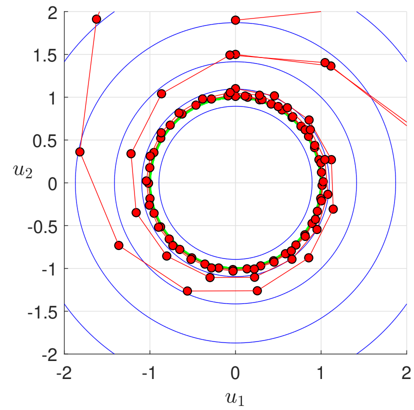
$$
u+v=w+O(\sigma(u)(u-w)^2) \quad \text{as } u\to\bar{u},\ w\to 0,
$$

and for quadratic convergence it is necessary and sufficient that

$$
w=O(u^2) \quad \text{as } u\to\bar{u}.
$$



(a) $\nu = 0.9$



(b) $\nu = 1$

Figure 2: Gauss–Newton method with cubic regularization for Example 3.1

This agrees with (3.2) in the case when σ is separated from zero near $\bar{u}$.

The following example is taken from [20, Section 5]; here it illustrates performance of the inexact Gauss–Newton method with cubic regularization in the case when solutions are nonisolated.

Example 3.1 Let $p = 2$, $q = 1$, $\Phi(u) = u_1^2 + u_2^2 - 1$, $P = \mathbb{R}^2$. Then the solution set U is the unit circle, and for every $\bar{u} \in U$ it holds that $\Phi'(\bar{u}) \neq 0$, implying the error bound (2.5).

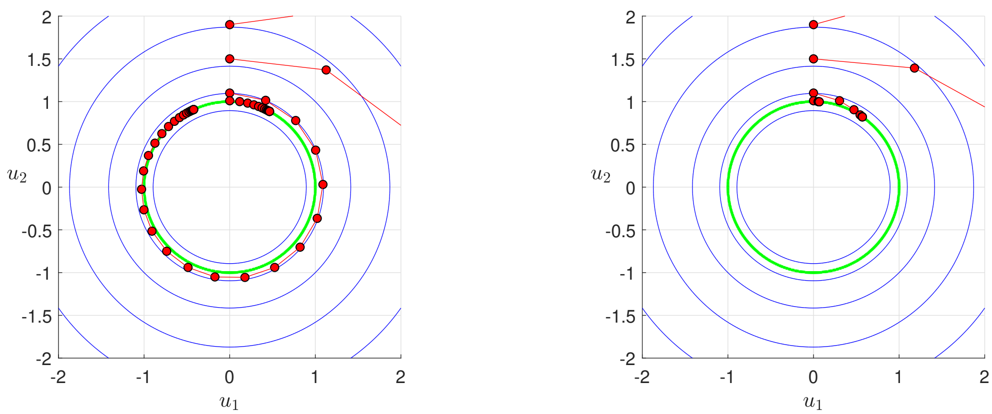
Let $\sigma(u) = |\Phi(u)|^\theta$ with a fixed $\theta \geq 0$. In our experiments, we implemented inexact solution of subproblems in the sense of (3.1) by solving (with high precision) subproblems (3.3) with the inexactness/perturbation term w constructed according to [20, Section 5]:

$$w(u) = |\Phi(u)|^{\theta+\nu}(u + u^\perp),$$

where $u^\perp = (u_2, -u_1)/\|u\|$ is orthogonal to u. In particular, for the requirement (3.2) on inexactness to be satisfied, it must hold that $\nu \geq 2$.

Figures 2 and 3 show some runs of the Gauss–Newton method with cubic regularization from four starting points of the form $u^0 = (0, 1 + t)$, $t = 0.9, 0.5, 0.1, 0.01$, for $\theta = 1$ (the maximum value satisfying (2.6)) and for various values of ν. The algorithm was implemented in Matlab, with the perturbed subproblems (3.3) solved by `fminunc` with the "Optimality-Tolerance" parameter set to 10^{-15}. The thick circle in the figures is the solution set, and thin circles are the level curves of the equation's residual.

The lack of convergence is observed numerically for $\nu = 0.9$ (Figure 2a) and $\nu = 1$ (Figure 2b). For larger ν, like $\nu = 1.1$ (Figure 3a) and $\nu = 1.5$ (Figure 3b), still violating (3.2), convergence from starting points close to the solution set appears superlinear (which for $\nu = 1.1$ only shows up after a long sequence of steps with slow decrease of the convergence measures). Convergence is not quadratic. The limit of the iterates approaches the starting point as the latter is getting closer to the solution set. For $\nu \geq 2$ (see Figure 4), satisfying



(a) $\nu = 1.1$

(b) $\nu = 1.5$

Figure 3: Gauss–Newton method with cubic regularization for Example 3.1.

(3.2), convergence from starting points close to the solution set appears quadratic, in agreement with Theorem 3.1. That said, we would caution that those conclusions regarding the convergence rates may be not fully reliable, as they are derived by considering some indirect measures of convergence (distance to the solution set, residual, length of the displacement), because solutions are nonisolated and the limit point of the iterates is not known. Also, some numerical (accuracy) issues were encountered in our runs when close to the solution set. In particular, the runs are often terminated earlier than the required stopping tolerance (10^{-12} for the residual) is achieved, because of a failure of `fminunc`.

We complete this section by mentioning the possibility to extend, under the appropriate assumptions, the analysis above to the case when Φ is only piecewise smooth, along the lines of how this is done for the Levenberg–Marquardt method in [22]. Another direction of possible developments is concerned with the projected (instead of constrained) Gauss-Newton method with cubic regularization, generating the next iterate as the projection of $u + v$ onto P, where v is the solution of the unconstrained version of (1.4), i.e., with P substituted therein by $\mathbb{R}^p$. However, the related considerations for the projected Levenberg–Marquardt method in [23, 24] suggest that superlinear convergence should not be expected in this case, at least under the assumptions that are not unnaturally strong.

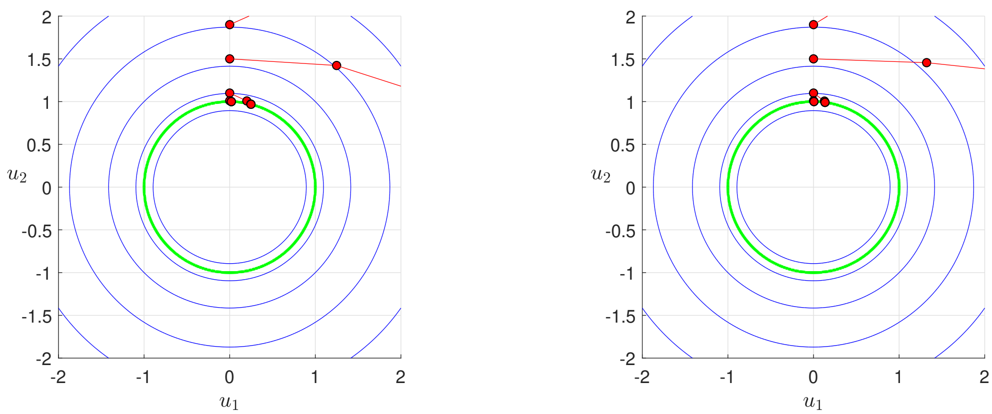
4 Linear convergence to singular solutions

In this section, we consider the unconstrained nonlinear equation

$$\Phi(u) = 0, \tag{4.1}$$

with smooth $\Phi : \mathbb{R}^p \to \mathbb{R}^p$, i.e., the special case of (1.1) with $p = q$ and $P = \mathbb{R}^p$.

Here, we are concerned with the case when the solution $\bar{u}$ in question is singular. That is, $\Phi'(\bar{u})$ is a singular matrix, and in particular, the error bound (2.5) may not hold.



(a) $\nu = 2$

(b) $\nu = 2.5$

Figure 4: Gauss–Newton method with cubic regularization for Example 3.1.

Assuming that Φ is twice differentiable at $\bar{u}$, the key assumption used in this section consists of the existence of some $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. The latter condition means that the linear operator $B(\bar{v}) : \ker \Phi'(\bar{u}) \to (\operatorname{im} \Phi'(\bar{u}))^{\perp}$ defined as the restriction of $\Pi \Phi''(\bar{u})[v]$ to $\ker \Phi'(\bar{u})$ is nonsingular, where Π is the orthogonal projector onto $(\operatorname{im} \Phi'(\bar{u}))^{\perp}$ in $\mathbb{R}^p$. In the case of a nonsingular solution $\bar{u}$, 2-regularity holds automatically with any $\bar{v}$, including $\bar{v} = 0$. What is important is that it may hold naturally for singular solutions as well (but with $\bar{v} \neq 0$). Note that 2-regularity is indeed a directional property, i.e., it does not depend on the norm of $\bar{v} \neq 0$. We refer the reader to [1] for a recent discussion of 2-regularity and related references.

The result of our Theorem 4.2 below relies on the perturbed Newton method framework discussed in [1]; some earlier related references are [25] and [26]. Within (a simplified version of) this framework, for a given iterate $u \in \mathbb{R}^p$, the next iterate is $u + v$, where v is a solution of the linear equation

$$\Phi(u) + \Phi'(u)v = \omega(u), \tag{4.2}$$

with a mapping $\omega : \mathbb{R}^p \to \mathbb{R}^p$ characterizing perturbations.

For any given scalars $\varepsilon > 0$ and $\delta > 0$, define the set

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

which can be thought of as a conic-like neighborhood of $\bar{u}$ associated with the direction $\bar{v}$.

In what follows, we shall make use of the unique decomposition of every $u \in \mathbb{R}^p$ into the sum $u = u_1 + u_2$ with $u_1 \in (\ker \Phi'(\bar{u}))^{\perp}$, $u_2 \in \ker \Phi'(\bar{u})$.

The following is a particular case of [1, Theorem 1].

Theorem 4.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be twice differentiable near $\bar{u} \in \mathbb{R}^p$, and assume that*

$$\Phi''(u) - \Phi''(\bar{u}) = O(\|u - \bar{u}\|) \quad as\ u \to \bar{u}. \tag{4.3}$$

Let $\bar{u}$ be a solution of equation (4.1), and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$. Let $\omega : \mathbb{R}^p \to \mathbb{R}^p$ possess the following properties: there exists $\delta > 0$ such that

$$\omega(u) = O(\|u - \bar{u}\|^2), \tag{4.4}$$

$$\Pi\omega(u) = O(\|u_1 - \bar{u}_1\| \|u - \bar{u}\|) + O(\|u - \bar{u}\|^3) \tag{4.5}$$

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

Then, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$, there exists the unique sequence $\{u^k\} \subset \mathbb{R}^p$ such that for each k, the displacement $u^{k+1} - u^k$ is the solution of (4.2) with $u = u^k$, and for this sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u};\, \bar{v})$. Furthermore, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically,

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_2^{k+1} - \bar{u}_2\|} = O(\|u^k - \bar{u}\|) \quad \text{as } k \to \infty, \tag{4.6}$$

and

$$\lim_{k\to\infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|} = \frac{1}{2}. \tag{4.7}$$

We now get back to the Gauss–Newton method with cubic regularization, whose subproblem (1.4) for the unconstrained equation (4.1) becomes

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{3}\sigma(u)\|v\|^3, \quad v \in \mathbb{R}^p. \tag{4.8}$$

The first-order optimality condition (2.14) for this problem takes the form

$$(\Phi'(u))^\top(\Phi(u) + \Phi'(u)v) + \sigma(u)\|v\|v = 0. \tag{4.9}$$

Assuming that $\sigma(u) > 0$, for the solution v of (4.8) we have that

$$\frac{1}{2}\|\Phi(u) + \Phi'(u)v\|^2 + \frac{1}{3}\sigma(u)\|v\|^3 \leq \frac{1}{2}\|\Phi(u) + \Phi'(u)\widetilde{v}\|^2 + \frac{1}{3}\sigma(u)\|\widetilde{v}\|^3 \quad \forall\, \widetilde{v} \in \mathbb{R}^p,$$

and therefore,

$$\|v\|^3 \leq \frac{3}{2\sigma(u)}\|\Phi(u) + \Phi'(u)\widetilde{v}\|^2 + \|\widetilde{v}\|^3 \quad \forall\, \widetilde{v} \in \mathbb{R}^p. \tag{4.10}$$

From [27, Lemma 3.1], we obtain the existence of $\varepsilon > 0$ and $\delta > 0$ such that for any $u \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v}) \setminus \{\bar{u}\}$, the matrix $\Phi'(u)$ is invertible, and

$$(\Phi'(u))^{-1} = O(\|u - \bar{u}\|^{-1}) \quad \text{as } \varepsilon \to 0+. \tag{4.11}$$

Therefore, for such u, there exists the unique displacement $\widetilde{v}$ of the basic Newton method for (4.1), defined by (4.2) with $\omega \equiv 0$. Moreover, according to [1, Lemma 1], it holds that

$$\widetilde{v} = O(\|u - \bar{u}\|) \quad \text{as } \varepsilon \to 0+.$$

From (4.10) applied with this $\widetilde{v}$, we then obtain that

$$\|v\| \leq \|\widetilde{v}\| = O(\|u - \bar{u}\|) \quad \text{as } \varepsilon \to 0+\,. \tag{4.12}$$

Then, setting

$$\omega(u) = \sigma(u)\|v\|((\Phi'(u))^\top)^{-1}v = \sigma(u)\|v\|((\Phi'(u))^{-1})^\top v,$$

we have that v satisfies (4.2). Next, from (4.11) and (4.12) we obtain that

$$\omega(u) = O(\sigma(u)\|u - \bar{u}\|) \quad \text{as } \varepsilon \to 0+\,. \tag{4.13}$$

Let $\sigma(u) = \chi\|\Phi(u)\|^\theta$ with fixed $\chi > 0$ and $\theta \geq 0$. Since

$$\Phi(u) = \Phi'(\bar{u})(u_1 - \bar{u}) + O(\|u - \bar{u}\|^2) \quad \text{as } \varepsilon \to 0+,$$

in this case we have from (4.13) that for $\theta \geq 1$ it holds that

$$\omega(u) = O(\|\Phi(u)\|^\theta\|u - \bar{u}\|) = O(\|\Phi(u)\|\|u - \bar{u}\|) = O(\|u_1 - \bar{u}\|\|u - \bar{u}\|) + O(\|u - \bar{u}\|^3),$$

yielding the needed properties (4.4)–(4.5), for $u \in K_{\varepsilon,\,\delta}(\bar{u};\, \bar{v})$ as $\varepsilon \to 0+$.

Applying Theorem 4.1, we now obtain the following result.

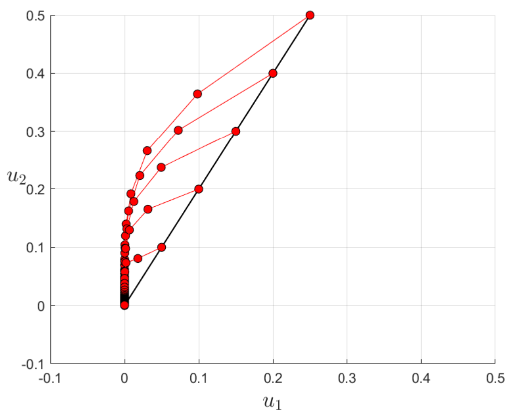
Theorem 4.2 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be twice differentiable near $\bar{u} \in \mathbb{R}^p$, and assume that* (4.3) *holds. Let $\bar{u}$ be a solution of equation* (4.1)*, and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$. Let $\sigma(u) = \chi\|\Phi(u)\|^\theta$ with fixed $\chi > 0$ and $\theta \geq 1$.*

Then, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\,\delta}(\bar{u};\, \bar{v})$, there exists the unique sequence $\{u^k\}$ such that for every k, the displacement $u^{k+1} - u^k$ is the solution of (4.8) *with $u = u^k$, and for this sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\,\widehat{\delta}}(\bar{u};\, \bar{v})$, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically, and* (4.6) *and* (4.7) *hold.*

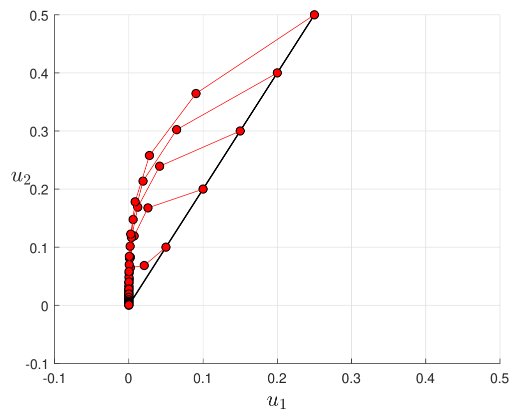
The following example illustrates the importance of the choice of the parameter $\theta \geq 1$ in Theorem 4.2.

Example 4.1 Let $p = q = 2$, $\Phi(u) = (u_1^2 + u_2^3,\, u_1u_2)$, $P = \mathbb{R}^2$. Then the only solution of the equation in (1.1) is $\bar{u} = 0$, this solution is singular ($\Phi'(\bar{u}) = 0$), and Φ is 2-regular at $\bar{u}$ in every direction $v \in \mathbb{R}^2$ with $v_1 \neq 0$.

Numerical experiments show that for $\theta < 1$, when starting from the points of the form $t(1/2,\, 2)$ (with 2-regularity satisfied in the direction $(1/2,\, 1)$), no matter how small $t > 0$ is, the Newton method with cubic regularization converges along the ray in the direction $(0,\, 1)$ violating 2-regularity, and the rate of convergence is either sublinear, or linear with the asymptotic common ratio greater than 1/2 (close to 1). An example of this behavior can be seen in Fig. 5. At the same time, for $\theta > 1$, we observe that when $t > 0$ is taken small enough, convergence is along some direction satisfying 2-regularity, and the rate of convergence is linear with the asymptotic common ratio 1/2, thus agreeing with the theory; see Fig. 6.

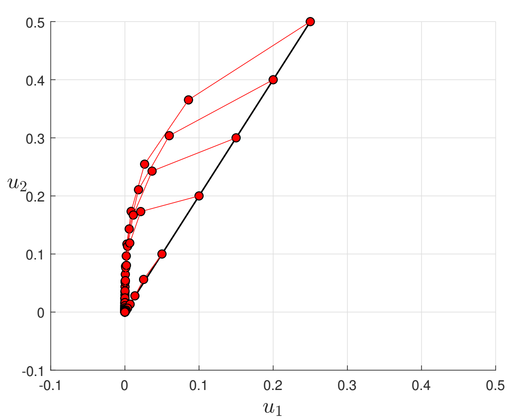


(a) $\theta = 0.6$

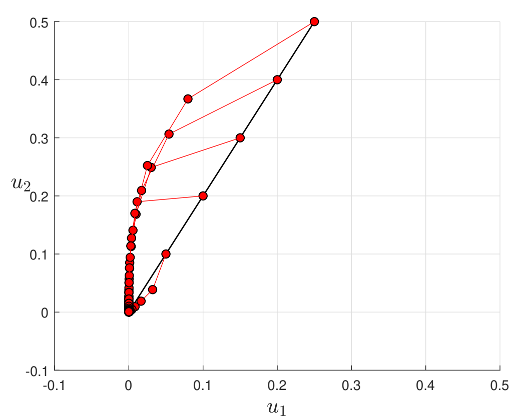


(b) $\theta = 0.9$

Figure 5: Gauss–Newton method with cubic regularization for Example 4.1; $\chi = 1$.



(a) $\theta = 1.1$



(b) $\theta = 1.4$

Figure 6: Gauss–Newton method with cubic regularization for Example 4.1; $\chi = 1$.

Declarations

Conflicts of Interest. The authors declare that they have no conflict of interest of any kind related to the manuscript.

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

References

[1] A.F. Izmailov and M.V. Solodov. A general perturbed Newtonian framework and critical solutions of nonlinear equations. Set-Valued Var. Anal. 33 (2025), Article 3.

[2] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov. Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. Comput. Optim. Appl. 63 (2016), 425–459.

[3] A. Fischer, A.F. Izmailov, M. Punke, and Z. Yan. Error bounds and Newton-type methods for reformulations of Karush-Kuhn-Tucker systems of mathematical programs with complementarity constraints. Comput. Optim. Appl. 93 (2025), 567–599.

[4] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. 172 (2004), 375–397.

[5] M. Macconi, B. Morini, and M. Porcelli. A Gauss-Newton method for solving bound-constrained underdetermined nonlinear systems. Optim. Meth. Software 24 (2009), 219–235.

[6] M. Frank and P. Wolfe. An algorithm for quadratic programming. Naval Research Logistics Quarterly 3 (1956), 95–110.

[7] A.F. Izmailov and M.V. Solodov. Local convergence of the Gauss–Newton methods for constrained nonlinear equations". Comput. Optim. Appl., 2026. https://doi.org/10.1007/s10589-026-00801-4.

[8] A. Fischer, A.F. Izmailov, and M.V. Solodov. The Levenberg–Marquardt method: an overview of modern convergence theories and more. Comput. Optim. Appl. 89 (2024), 33–67.

[9] A. Griewank. The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical report NA/12 (1981).

[10] Y. Nesterov and B.T. Polyak. Cubic regularization of Newton method and its global performance. Math. Program. 108 (2006), 177–205.

[11] C. Cartis, N.I.M. Gould, and P.L. Toint. Adaptive cubuc regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Math. Program. 127 (2011), 245–295.

[12] J.-P. Dussault, T. Migot, and D. Orban. Scalable adaptive cubic regularization methods. Math. Program. 207 (2024), 191–225.

[13] Y. Pei, Y. Lin, S. Shao, M. Silva Louzeiro, and D. Zhu. A scalable sequential adaptive cubic regularization algorithm for optimization with general equality constraints. Numerical Algorithms (2026), https://doi.org/10.1007/s11075-026-02401-z.

[14] C.T. Kelley. Iterative Methods for Optimization. SIAM, Philadelphia, 1999.

[15] M.-C. Yue, Z. Zhou, and A.M.-C. So. On the quadratic convergence of the cubic regularization method under a local error bound condition. SIAM J. Optim. 29 (2019), 904–932.

[16] Q. Rebjock and N. Boumal. Fast convergence to non-isolated minima: four equivalent conditions for C^2 functions. Math. Program. 213 (2025), 151–199.

[17] F. Fischer and N. Strasdat. An extended convergence framework applied to complementarity systems with degenerate and nonisolated solutions. Pure Appl. Func. Analys. 8 (2023), 1039–1054.

[18] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg–Marquardt method. In: Alefeld, G., Chen, X. (eds) Topics in Numerical Analysis. Computing Supplementa, vol. 15, Springer, Vienna, 2001, pp. 239–249.

[19] A.F. Izmailov and M.V. Solodov. Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, Cham, 2014.

[20] R. Behling and A. Fischer. A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. 6 (2012), 927–940.

[21] J.P. Luna, C. Sagastizábal, and M.V. Solodov. A class of Danzig–Wolfe type decomposition methods for variational inequality problems. Math. Program. 143 (2014), 177–209.

[22] A.F. Izmailov, E.I. Uskov, and Z. Yan. Globalization of convergence of the constrained piecewise Levenberg-Marquardt method Optim. Meth. Software 40 (2025), 243–265.

[23] R. Behling, A. Fischer, M. Herrich, A. Iusem, and Y. Ye. A Levenberg-Marquardt method with approximate projections. Comput. Optim. Appl. 59 (2014), 2–26.

[24] R. Behling, A. Fischer, G. Haeser, A. Ramos, and K. Schonefeld. On the constrained error bound condition and the projected Levenberg-Marquardt method. Optimization 66 (2017), 1397–1411.

[25] A. Fischer, A.F. Izmailov, and M. Jelitte. Behavior of Newton-type methods near critical solutions of nonlinear equations with semismooth derivatives. J. Optim. Theory Appl. 203 (2024), 2179–2205.

[26] 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.

[27] A. Griewank. Starlike domains of convergence for Newton's method at singularities. Numer. Math. 35 (1980), 95–111.