

## Critical solutions of nonlinear equations: local attraction for Newton-type methods

A. F. Izmailov · A. S. Kurennoy · M. V. Solodov

Received: March 2016 / Revised: November 2016

**Abstract** We show that if the equation mapping is 2-regular at a solution in some nonzero direction in the null space of its Jacobian (in which case this solution is critical; in particular, the local Lipschitzian error bound does not hold), then this direction defines a star-like domain with nonempty interior from which the iterates generated by a certain class of Newton-type methods necessarily converge to the solution in question. This is despite the solution being degenerate, and possibly non-isolated (so that there are other solutions nearby). In this sense, Newtonian iterates are attracted to the specific (critical) solution. Those results are related to the ones due to A. Griewank for the basic Newton method but are also applicable, for example, to some methods developed specially for tackling the case of potentially non-isolated solutions, including the Levenberg–Marquardt and the LP-Newton methods for equations, and the stabilized sequential quadratic programming for optimization.

**Keywords** Newton method · critical solutions · 2-regularity · Levenberg–Marquardt method · linear-programming-Newton method · stabilized sequential quadratic programming.

**Mathematics Subject Classification (2000)** 90C33 · 65K10 · 49J53.

---

This research is supported in part by the Russian Foundation for Basic Research Grant 17-01-00125, by the Russian Science Foundation Grant 15-11-10021, by the Ministry of Education and Science of the Russian Federation (the Agreement number 02.a03.21.0008), by VolkswagenStiftung Grant 115540, by CNPq Grants PVE 401119/2014-9 and 303724/2015-3, and by FAPERJ.

A. F. Izmailov  
VMK Faculty, OR Department, Lomonosov Moscow State University (MSU), Uchebnyy 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

A. S. Kurennoy  
Department of Mathematics, Physics and Computer Sciences, Derzhavin Tambov State University, TSU, Internationalnaya 33, 392000 Tambov, Russia. E-mail: akurennoy@cs.msu.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

## 1 Introduction

We consider a nonlinear equation

$$\Phi(u) = 0, \quad (1)$$

where the mapping  $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$  is smooth enough (precise smoothness assumptions will be stated later as needed). This paper is concerned with convergence properties of Newton-type methods for solving the equation (1) when it has a singular solution  $\bar{u}$  (i.e., the matrix  $\Phi'(\bar{u})$  is singular). Of particular interest is the difficult case when  $\bar{u}$  may be a non-isolated solution of (1). Note that if  $\bar{u}$  is a non-isolated solution, it is necessarily singular.

To describe the class of methods in question, we define the perturbed Newton method (pNM) framework for equation (1) as follows. For the given iterate  $u^k \in \mathbb{R}^p$ , the next iterate is  $u^{k+1} = u^k + v^k$ , with  $v^k$  satisfying the following linear equation in  $v$ :

$$\Phi(u^k) + (\Phi'(u^k) + \Omega(u^k))v = \omega(u^k). \quad (2)$$

In (2), the mappings  $\Omega : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$  and  $\omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$  are certain perturbation terms which may have different roles, and individually or collectively define specific methods within the general pNM framework. In particular, if  $\Omega \equiv 0$  and  $\omega \equiv 0$ , then (2) reduces to the iteration system of the basic Newton method. The most common setting would be for  $\Omega$  to characterize a perturbation of the iteration matrix of the basic Newton method (i.e., the difference between the matrix a given method actually employs, compared to the exact Jacobian  $\Phi'(u^k)$ ), and for  $\omega$  to account for possible inexactness in solving the corresponding linear system of equations. An example of this setting within this paper, is the stabilized Newton-Lagrange method (stabilized sequential quadratic programming for equality-constrained optimization), considered in Section 3.3 below. However, we emphasize that our framework is not restricted to this situation. In particular, subproblems of a given method need not even be systems of linear equations, as long as they can be related to (2) a posteriori. One example is the linear-programming-Newton (LP-Newton) method discussed in Section 3.2 below, which solves linear programs, and for which the perturbation term  $\omega$  is implicit (i.e., it does not have an explicit analytical formula, but its properties are known). In this respect, we also comment that the way we shall employ the perturbation mappings  $\Omega$  and  $\omega$  is somewhat unusual, in the following sense. They may describe a given method not necessarily on the whole neighborhood of a solution of interest, but possibly only in some relevant star-like domain of convergence; see the discussion of the Levenberg–Marquardt method in Section 3.1 and of the LP-Newton method in Section 3.2. This, however, is exactly what is needed in the presented convergence analysis, as it is shown that the generated iterates do in fact stay within the domain in question and, within this set,  $\Omega$  and  $\omega$  that we construct do adequately represent the given algorithms. Finally, we note that the specific methods that we consider in this paper have been designed to tackle the difficult cases when (1) has degenerate/non-isolated solutions, and in this sense the perturbation terms in (2) that describe these methods can be regarded as “structural”, i.e., introduced *intentionally* for improving convergence properties in the degenerate cases. The assumptions imposed on  $\Omega$  and  $\omega$  are only related to their “size”, which allows  $\omega$  to cover naturally precision control when the subproblems are solved approximately. However, as already commented, the use of  $\omega$  can also be quite different. In the analysis of the LP-Newton method in Section 3.2,  $\omega$  is implicit and is not related to solving subproblems approximately.

Our convergence results assume a certain 2-regularity property of the solution of (1), which implies that this solution is “critical” in the sense of [15]. We next state the relevant definitions, and discuss the relations between those concepts.

Recall first that if  $\Phi$  is differentiable at a solution  $\bar{u}$  of the equation (1), then it holds that

$$T_{\Phi^{-1}(0)}(\bar{u}) \subset \ker \Phi'(\bar{u}),$$

where  $\Phi^{-1}(0)$  is the solution set of (1), and by  $T_S(u)$  we denote the contingent cone to a set  $S \subset \mathbb{R}^p$  at  $u \in S$ , i.e. the tangent cone as defined in [26, Definition 6.1]:

$$T_S(u) = \{v \in \mathbb{R}^p \mid \exists \{t_k\} \subset \mathbb{R}_+ : \{t_k\} \rightarrow 0+, \text{dist}(u + t_k v, S) = o(t_k)\}.$$

Recall that according to [26, Corollary 6.29] (see also [26, Definition 6.4] and the original definition in [3, Definition 2.4.6]), assuming that a set  $S \subset \mathbb{R}^p$  is closed near  $u \in S$ , this set is Clarke-regular at  $u$  if the multifunction  $T_S(\cdot)$  is inner semicontinuous at  $u$  relative to  $S$  (the latter meaning that for any  $v \in T_S(u)$  and any sequence  $\{u^k\} \subset S$  convergent to  $u$ , there exists a sequence  $\{v^k\}$  convergent to  $v$  such that  $v^k \in T_S(u^k)$  for all  $k$ ).

The following notion was introduced in [15].

**Definition 1** Assuming that  $\Phi$  is differentiable at a solution  $\bar{u}$  of the equation (1), this solution is referred to as *noncritical* if the set  $\Phi^{-1}(0)$  is Clarke-regular at  $\bar{u}$ , and

$$T_{\Phi^{-1}(0)}(\bar{u}) = \ker \Phi'(\bar{u}). \quad (3)$$

Otherwise, the solution  $\bar{u}$  is referred to as *critical*.

Observe that according to the above definition, any isolated singular solution is necessarily critical.

As demonstrated in [15], if  $\Phi$  satisfies some mild and natural smoothness assumptions, then noncriticality of  $\bar{u}$  is equivalent to the local Lipschitzian error bound on the distance to the solution set in terms of the natural residual of the equation (1):

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

holds as  $u \in \mathbb{R}^p$  tends to  $\bar{u}$ . Moreover, it is also equivalent to the upper-Lipschitzian stability with respect to right-hand side perturbations of (1): any solution  $u(w)$  of the perturbed equation

$$\Phi(u) = w,$$

close enough to  $\bar{u}$ , satisfies

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

as  $w \in \mathbb{R}^p$  tends to 0. Accordingly, criticality of  $\bar{u}$  means the absence of the properties above.

The interest in critical/noncritical solutions of nonlinear equations originated from the study of special Lagrange multipliers in equality-constrained optimization, also called critical [13, 18, 19, 14, 22], [21, Chapter 7]. For the relations between critical solutions of equations and critical multipliers in optimization, see [15]. It had been demonstrated that critical Lagrange multipliers tend to attract dual sequences generated by a number of Newton-type methods for optimization [18, 19], [21, Chapter 7]. In this paper, we show that critical solutions of nonlinear equations also serve as attractors, in this case for methods described by the pNM framework (2).

For a symmetric bilinear mapping  $B : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^p$  and an element  $v \in \mathbb{R}^p$ , we define the linear operator  $B[v] : \mathbb{R}^p \rightarrow \mathbb{R}^p$  by  $B[v]u = B[v, u]$ . The notion of 2-regularity is a useful tool in nonlinear analysis and optimization theory; see, e.g., the book [1], as well as [2, 16, 17, 9, 10] for some applications. The essence of the construction is that when a mapping  $\Phi$  is irregular at  $\bar{u}$  ( $\Phi'(\bar{u})$  is singular), first-order information is insufficient to adequately represent  $\Phi$  around  $\bar{u}$ , and so second-order information has to come into play. To this end, we have the following.

**Definition 2** Assuming that  $\Phi$  is twice differentiable at  $\bar{u} \in \mathbb{R}^p$ ,  $\Phi$  is said to be *2-regular at the point  $\bar{u}$  in the direction  $v \in \mathbb{R}^p$*  if the  $p \times p$ -matrix

$$\Psi(\bar{u}; v) = \Phi'(\bar{u}) + \Pi \Phi''(\bar{u})[v] \quad (4)$$

is nonsingular, where  $\Pi$  is the projector in  $\mathbb{R}^p$  onto an arbitrary fixed complementary subspace of  $\text{im } \Phi'(\bar{u})$ , along this subspace.

In our convergence analysis of Newton-type methods described by (2), we shall assume that  $\Phi$  is 2-regular at a solution  $\bar{u}$  of (1) in some direction  $v \in \ker \Phi'(\bar{u}) \setminus \{0\}$ . It turns out that this implies that the solution  $\bar{u}$  is necessarily critical in the sense of Definition 1. We show this next.

**Proposition 1** *Let  $\Phi$  be twice differentiable at a solution  $\bar{u}$  of the equation (1) and 2-regular at this solution in some direction  $v \in \ker \Phi'(\bar{u}) \setminus \{0\}$ .*

*Then (3) does not hold and, in particular,  $\bar{u}$  is a critical solution of (1).*

*Proof* Let  $v \in \ker \Phi'(\bar{u}) \setminus \{0\}$ . Suppose that (3) holds. Then  $v \in T_{\Phi^{-1}(0)}(\bar{u})$ . Thus, there exist a sequence  $\{t_k\}$  of positive reals and a sequence  $\{r^k\} \subset \mathbb{R}^p$  such that  $\{t_k\} \rightarrow 0$ ,  $r^k = o(t_k)$ , and for all  $k$  it holds that

$$0 = \Phi(\bar{u} + t_k v + r^k) = \Phi'(\bar{u})r^k + \frac{1}{2}t_k^2 \Phi''(\bar{u})[v, v] + o(t_k^2).$$

Therefore,

$$t_k^2 \Phi''(\bar{u})[v, v] + o(t_k^2) \in \text{im } \Phi'(\bar{u}),$$

which implies that

$$\Phi''(\bar{u})[v, v] \in \text{im } \Phi'(\bar{u}).$$

Hence,  $\Pi \Phi''(\bar{u})[v, v] = 0$ , where  $\Pi$  is specified in Definition 2. Since also  $\Phi'(\bar{u})v = 0$ , from (4) we conclude that  $v \in \ker \Psi(\bar{u}; v)$ . As  $v \neq 0$ , this contradicts 2-regularity (specifically, the nonsingularity of  $\Psi(\bar{u}; v)$ ). Thus, (3) cannot hold, and  $\bar{u}$  must be a critical solution.  $\square$

In Section 2, we shall prove that if  $\Phi$  is 2-regular at a (critical) solution  $\bar{u}$  of (1) in some direction  $v \in \ker \Phi'(\bar{u}) \setminus \{0\}$ , then  $v$  defines a domain star-like with respect to  $\bar{u}$ , with nonempty interior, from which the iterates that satisfy the pNM framework (2) necessarily converge to  $\bar{u}$ . In this sense, the iterates are “attracted” specifically to  $\bar{u}$ , even though there may be other nearby solutions. These results are related to [11], where the pure Newton method was considered (i.e., (2) with  $\Omega \equiv 0$  and  $\omega \equiv 0$ ). An interesting extension of the results in [11] to the case when  $\Phi$  is not necessarily twice differentiable, but has a Lipschitz-continuous first derivative, has been proposed in [25]. The latter reference also gives an application to smooth equation reformulations of complementarity problems.

In Section 3, we demonstrate how the general results for the pNM framework (2) apply to some specific Newton-type methods. These include the classical Levenberg–Marquardt method [24, Chapter 10.2] and the LP-Newton method [6] for nonlinear equations, and the stabilized Newton–Lagrange method for optimization (or stabilized sequential quadratic programming) [28, 12, 7, 20]; see also [21, Chapter 7].

We finish this section with some words about our notation. Throughout,  $\langle \cdot, \cdot \rangle$  is the Euclidian inner product, and unless specified otherwise,  $\|\cdot\|$  is the Euclidian norm, where the space is always clear from the context. Then the unit sphere is  $\mathbb{S} = \{u \mid \|u\| = 1\}$ . For a linear operator (a matrix)  $A$ , we denote by  $\ker A$  its null space, and by  $\text{im } A$  its image (range) space. The notation  $I$  stands for the identity matrix. A set  $U$  is called star-like with respect to  $u \in U$  if  $t\hat{u} + (1-t)u \in U$  for all  $\hat{u} \in U$  and all  $t \in [0, 1]$ .

## 2 Local convergence of pNM iterates to a critical solution

Results of this section are related to those in [11], where the basic Newton method was considered (i.e. pNM (2) with  $\Omega \equiv 0$  and  $\omega \equiv 0$ ). To the best of our knowledge, this is the only directly related reference, as it also allows for non-isolated solutions. Other literature on the basic Newton method for singular equations uses assumptions that imply that a solution, though possibly singular, must be isolated.

Let  $\bar{u}$  be a solution of the equation (1). Then every  $u \in \mathbb{R}^p$  is uniquely decomposed into the sum  $u = u_1 + u_2$ , with  $u_1 \in (\ker \Phi'(\bar{u}))^\perp$  and  $u_2 \in \ker \Phi'(\bar{u})$ . The corresponding notation will be used throughout the rest of the paper.

We start with the following counterpart of [11, Lemma 4.1], which establishes that, under appropriate assumptions, the pNM subproblem (2) has the unique solution for all  $u^k$  close enough to  $\bar{u}$ , in a certain star-like domain.

**Lemma 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 solution of the equation (1), and assume that  $\Phi$  is 2-regular at  $\bar{u}$  in a direction  $\bar{v} \in \mathbb{R}^p \cap \mathbb{S}$ . Let  $\Pi$  stand for the orthogonal projector onto  $(\text{im } \Phi'(\bar{u}))^\perp$ . Let  $\Omega : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$  satisfy the following properties:

$$\Omega(u) = O(\|u - \bar{u}\|) \quad (5)$$

as  $u \rightarrow \bar{u}$ , and for every  $\Delta > 0$  there exist  $\varepsilon > 0$  and  $\delta > 0$  such that for every  $u \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying

$$\|u - \bar{u}\| \leq \varepsilon, \quad \left\| \frac{u - \bar{u}}{\|u - \bar{u}\|} - \bar{v} \right\| \leq \delta, \quad (6)$$

it holds that

$$\|\Pi\Omega(u)\| \leq \Delta \|u - \bar{u}\|. \quad (7)$$

Let  $\omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$  satisfy

$$\omega(u) = O(\|u - \bar{u}\|^2) \quad (8)$$

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

Then there exist  $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$  and  $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$  such that for every  $u \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying

$$\|u - \bar{u}\| \leq \bar{\varepsilon}, \quad \left\| \frac{u - \bar{u}}{\|u - \bar{u}\|} - \bar{v} \right\| \leq \bar{\delta}, \quad (9)$$

the equation (2) with  $u^k = u$  has the unique solution  $v$ , satisfying

$$u_1 + v_1 - \bar{u}_1 = O(\|u - \bar{u}\|^2), \quad (10)$$

$$\begin{aligned} u_2 + v_2 - \bar{u}_2 &= \frac{1}{2}(u_2 - \bar{u}_2) + O(\|u_1 - \bar{u}_1\|) \\ &+ O(\|u - \bar{u}\|^{-1} \Pi\omega(u)) + O(\|\Pi\Omega(u)\|) + O(\|u - \bar{u}\|^2) \end{aligned} \quad (11)$$

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

*Proof* Under the stated smoothness assumptions, without loss of generality we can assume that  $\bar{u} = 0$  and

$$\Phi(u) = Au + \frac{1}{2}B[u, u] + R(u), \quad (12)$$

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  $R : \mathbb{R}^p \rightarrow \mathbb{R}^p$  is differentiable near 0, with

$$R(u) = O(\|u\|^3), \quad R'(u) = O(\|u\|^2) \quad (13)$$

as  $u \rightarrow 0$ .

Substituting the form of  $\Phi$  stated in (12) into (2), and multiplying (2) by  $(I - \Pi)$  and by  $\Pi$ , we decompose (2) into the following two equations:

$$\begin{aligned} (A + (I - \Pi)(B[u] + R'(u) + \Omega(u)))v_1 &= -Au_1 - (I - \Pi) \left( \frac{1}{2}B[u, u] + R(u) - \omega(u) \right) \\ &\quad - (I - \Pi)(B[u] + R'(u) + \Omega(u))v_2, \end{aligned} \quad (14)$$

and

$$\Pi(B[u] + R'(u) + \Omega(u))(v_1 + v_2) = -\Pi \left( \frac{1}{2}B[u, u] + R(u) - \omega(u) \right). \quad (15)$$

Let  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$  be arbitrary and fixed for now. From this point on, we consider only those  $u \in \mathbb{R}^p \setminus \{0\}$  that satisfy (9).

Define the family of linear operators  $\mathcal{A}(u) : (\ker A)^\perp \rightarrow \text{im} A$  as the restriction of  $(A + (I - \Pi)(B[u] + R'(u) + \Omega(u)))$  to  $(\ker A)^\perp$ . Let  $\bar{\mathcal{A}} : (\ker A)^\perp \rightarrow \text{im} A$  be the restriction of  $A$  to  $(\ker A)^\perp$ . Then, taking into account (8) and (13), the equality (14) can be written as

$$\mathcal{A}(u)v_1 = -\bar{\mathcal{A}}u_1 - (I - \Pi)(B[u] + R'(u) + \Omega(u))v_2 + O(\|u\|^2) \quad (16)$$

as  $u \rightarrow 0$ . Evidently,  $\bar{\mathcal{A}}$  is invertible, and according to (5) and (13),

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

The latter implies that  $\mathcal{A}(u)$  is invertible, provided  $\bar{\varepsilon} > 0$  is small enough, and

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

as  $u \rightarrow 0$  (see, e.g., [21, Lemma A.6]). Therefore, (16) can be written as

$$v_1 = -u_1 + \mathcal{M}(u)v_2 + O(\|u\|^2), \quad (17)$$

where  $\mathcal{M}(u) : \ker A \rightarrow (\ker A)^\perp$ ,

$$\mathcal{M}(u) = -(\mathcal{A}(u))^{-1}(I - \Pi)(B[u] + R'(u) + \Omega(u)) = O(\|u\|) \quad (18)$$

as  $u \rightarrow 0$ .

Substituting (17) into (15), and taking into account (13), we obtain that

$$\begin{aligned} \Pi(B[u] + R'(u) + \Omega(u))(I + \mathcal{M}(u))v_2 &= -\Pi \left( \frac{1}{2}B[u, u] - \omega(u) \right) + \Pi(B[u] + \Omega(u))u_1 \\ &\quad + O(\|u\|^3). \end{aligned} \quad (19)$$

Define the family of linear operators  $\mathcal{B}(u) : \ker A \rightarrow (\operatorname{im} A)^\perp$  as the restriction of  $\Pi(B[u] + R'(u) + \Omega(u))(I + \mathcal{M}(u))$  to  $\ker A$ . Let  $\tilde{\mathcal{B}}(u) : \ker A \rightarrow (\operatorname{im} A)^\perp$  be the restriction of  $\Pi B[u]$  to  $\ker A$ . Then (19) can be written as

$$\mathcal{B}(u)v_2 = -\frac{1}{2}\tilde{\mathcal{B}}(u)u_2 + \Pi\left(\left(\frac{1}{2}B[u] + \Omega(u)\right)u_1 + \omega(u)\right) + O(\|u\|^3) \quad (20)$$

as  $u \rightarrow 0$ .

The 2-regularity of  $\Phi$  at 0 in the direction  $\bar{v}$  means precisely that  $\tilde{\mathcal{B}}(\bar{v})$  is nonsingular. Then, possibly after reducing  $\bar{\delta} > 0$ , by [21, Lemma A.6] we obtain the existence of  $C > 0$  such that for every  $u \in \mathbb{R}^p \setminus \{0\}$  satisfying the second relation in (9),  $\tilde{\mathcal{B}}(u/\|u\|)$  is invertible, and

$$\left\| \left( \tilde{\mathcal{B}}(\|u\|^{-1}u) \right)^{-1} \right\| \leq C. \quad (21)$$

According to (13), (5), (18), it holds that

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

Choosing  $\Delta > 0$  small enough, and further reducing  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$  if necessary, by (7) and (21), and by [21, Lemma A.6], we now obtain that  $\mathcal{B}(\|u\|^{-1}u)$  is invertible, and

$$\left(\mathcal{B}(\|u\|^{-1}u)\right)^{-1} = \left(\tilde{\mathcal{B}}(\|u\|^{-1}u)\right)^{-1} + O(\|u\|^{-1}\|\Pi\Omega(u)\|) + O(\|u\|).$$

Employing again (5), we further conclude that

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

Therefore, (20) is uniquely solvable, and its unique solution has the form

$$v_2 = -\frac{1}{2}u_2 + \left(\tilde{\mathcal{B}}(u)\right)^{-1}\Pi\left(\frac{1}{2}B[u, u_1] + \omega(u)\right) + O(\|\Pi\Omega(u)\|) + O(\|u\|^2) = O(\|u\|) \quad (22)$$

as  $u \rightarrow 0$ , where the last estimate employs (5) and (8).

Substituting (22) into (17), and employing (18) again, we finally obtain that

$$v_1 = -u_1 + O(\|u\|^2) \quad (23)$$

as  $u \rightarrow 0$ .

From (22) and (23), we have the needed estimates (10) and (11).  $\square$

*Remark 1* From the proof of Lemma 1 it can be seen that under the assumptions of this lemma (removing the assumptions on  $\omega$  which are not needed for the following), the values  $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$  and  $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$  can be chosen in such a way that for every  $u \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying (9), the matrix  $\Phi'(u) + \Omega(u)$  is invertible, and  $(\Phi'(u) + \Omega(u))^{-1} = O(\|u - \bar{u}\|^{-1})$  as  $u \rightarrow \bar{u}$ . When  $\Omega \equiv 0$ , this result is a particular case of [11, Lemma 3.1].

We proceed to establish convergence of the iterates satisfying the pNM framework (2), from any starting point in the relevant domain. This result is a generalization of [11, Lemma 5.1].

**Theorem 1** Let  $\Phi : \mathbb{R}^p \rightarrow \mathbb{R}^p$  be twice differentiable near  $\bar{u} \in \mathbb{R}^p$ , with its second derivative being 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 solution of the equation (1), and assume that  $\Phi$  is 2-regular at  $\bar{u}$  in a direction  $\bar{v} \in \ker \Phi'(\bar{u}) \cap \mathbb{S}$ . Let  $\Omega : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$  and  $\omega : \mathbb{R}^p \rightarrow \mathbb{R}^p$  satisfy the estimates (5), (8), as well as

$$\Pi\Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \quad (24)$$

and

$$\Pi\omega(u) = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3) \quad (25)$$

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

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 any starting point  $u^0 \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying

$$\|u^0 - \bar{u}\| \leq \varepsilon, \quad \left\| \frac{u^0 - \bar{u}}{\|u^0 - \bar{u}\|} - \bar{v} \right\| \leq \delta \quad (26)$$

uniquely defines the sequence  $\{u^k\} \subset \mathbb{R}^p$  such that for each  $k$  it holds that  $v^k = u^{k+1} - u^k$  solves (2),  $u_2^k \neq \bar{u}_2$ , the point  $u = u^k$  satisfies (9), the sequence  $\{u^k\}$  converges to  $\bar{u}$ , the sequence  $\{\|u^k - \bar{u}\|\}$  converges to zero monotonically,

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_2^{k+1} - \bar{u}_2\|} = O(\|u^k - \bar{u}\|) \quad (27)$$

as  $k \rightarrow \infty$ , and

$$\lim_{k \rightarrow \infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|} = \frac{1}{2}. \quad (28)$$

*Proof* We again assume that  $\bar{u} = 0$ , and  $\Phi$  is given by (12) with  $R$  satisfying (13).

Considering that  $\bar{v}_1 = 0$ , observe first that if  $u \in \mathbb{R}^p \setminus \{0\}$  satisfies the second condition in (6) with some  $\delta \in (0, 1)$ , then

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

This implies that

$$\|u_1\| \leq \delta \|u\|, \quad (29)$$

and hence,

$$\|u\| \leq \|u_1\| + \|u_2\| \leq \delta \|u\| + \|u_2\|,$$

so that

$$(1 - \delta)\|u\| \leq \|u_2\|. \quad (30)$$

Then

$$\begin{aligned} \left\| \frac{u_2}{\|u_2\|} - \bar{v} \right\| &\leq \left\| \frac{u_2}{\|u\|} - \bar{v}_2 \right\| + \left\| \frac{u_2}{\|u_2\|} - \frac{u_2}{\|u\|} \right\| \\ &\leq \left\| \frac{u}{\|u\|} - \bar{v} \right\| + \frac{\|u\| - \|u_2\|}{\|u\|} \\ &\leq \delta + 1 - \frac{\|u_2\|}{\|u\|} \\ &\leq 2\delta, \end{aligned} \quad (31)$$



where the second inequality employs the fact that  $u_2/\|u\|$  is the metric projection of  $u/\|u\|$  onto  $\ker A$ .

From (24) and (29) it evidently follows that  $\Omega$  satisfies the corresponding assumptions in Lemma 1. Therefore, according to this lemma, there exist  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$  such that for every  $u \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying (9), the equation (2) with  $u^k = u$  has the unique solution  $v$ , and

$$u_1 + v_1 = O(\|u\|^2)$$

and

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

as  $u \rightarrow 0$ , where (24) and (25) were again taken into account.

In what follows, we use these values  $\bar{\varepsilon}$  and  $\bar{\delta}$ , with the understanding that if we prove the assertion of the theorem for these specific values of  $\bar{\varepsilon}$  and  $\bar{\delta}$ , it will also be valid for any larger values of those constants. At the same time, since Lemma 1 allows for these  $\bar{\varepsilon}$  and  $\bar{\delta}$ , it certainly allows for any smaller values.

Therefore, there exists  $C > 0$  such that

$$\|u_1 + v_1\| \leq C\|u\|^2, \quad (32)$$

$$\left\| u_2 + v_2 - \frac{1}{2}u_2 \right\| \leq \left\| u + v - \frac{1}{2}u \right\| \leq C(\|u_1\| + \|u\|^2), \quad (33)$$

and hence,

$$\frac{1}{2}\|u_2\| - C(\|u_1\| + \|u\|^2) \leq \|u_2 + v_2\| \leq \|u + v\| \leq \frac{1}{2}\|u_2\| + C(\|u_1\| + \|u\|^2). \quad (34)$$

From (9), (29) and (30) with  $\delta = \bar{\delta}$ , and from (34), we further derive that

$$\left( \frac{1 - \bar{\delta}}{2} - C(\bar{\delta} + \bar{\varepsilon}) \right) \|u\| \leq \|u_2 + v_2\| \leq \|u + v\| \leq \left( \frac{1}{2} + C(\bar{\delta} + \bar{\varepsilon}) \right) \|u\|.$$

Reducing  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$  if necessary, so that

$$\frac{\bar{\delta}}{2} + C(\bar{\delta} + \bar{\varepsilon}) < \frac{1}{2},$$

and setting

$$q_- = \frac{1 - \bar{\delta}}{2} - C(\bar{\delta} + \bar{\varepsilon}), \quad q_+ = \frac{1}{2} + C(\bar{\delta} + \bar{\varepsilon}),$$

we then obtain that

$$q_- \|u\| \leq \|u_2 + v_2\| \leq \|u + v\| \leq q_+ \|u\|, \quad (35)$$

where

$$0 < q_- < q_+ < 1. \quad (36)$$

By (33), the right inequality in (34), and the left inequality in (35), we have that

$$\begin{aligned} \left\| \frac{u + v}{\|u + v\|} - \frac{u_2}{\|u_2\|} \right\| &= \frac{\|(u + v)\|u_2\| - \|u_2\|\|u + v\|}{\|u_2\|\|u + v\|} \\ &\leq \frac{2C(\|u_1\| + \|u\|^2)}{q_- \|u\|} \\ &= \frac{2C}{q_-} \left( \frac{\|u_1\|}{\|u\|} + \|u\| \right), \end{aligned} \quad (37)$$

$$\begin{aligned}
\left\| \frac{u_2 + v_2}{\|u_2 + v_2\|} - \frac{u_2}{\|u_2\|} \right\| &= \frac{\|(u_2 + v_2)\|u_2\| - u_2\|u_2 + v_2\|\|}{\|u_2\|\|u_2 + v_2\|} \\
&\leq \frac{2C(\|u_1\| + \|u\|^2)}{q_- \|u\|} \\
&= \frac{2C}{q_-} \left( \frac{\|u_1\|}{\|u\|} + \|u\| \right). \tag{38}
\end{aligned}$$

Now choose  $\varepsilon \in (0, \bar{\varepsilon}]$ ,  $\delta \in (0, \bar{\delta}]$  satisfying

$$2\delta + \frac{2C}{q_-} \left( \delta + \left( \frac{C}{q_-} + 1 \right) \frac{\varepsilon}{1 - q_+} \right) \leq \bar{\delta}, \tag{39}$$

and assume that (26) holds for  $u^0 \in \mathbb{R}^p \setminus \{0\}$ . Suppose that  $u^j \in \mathbb{R}^p \setminus \{0\}$ , satisfy (2) and (9) with  $u = u^j$  for all  $j = 1, \dots, k$ . Then by the choice of  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$ , there exists the unique  $u^{k+1}$  satisfying (2), and by (29), (31), (32), (35), (36), (37), (38), it holds that

$$0 < \|u^{k+1}\| \leq q_+ \|u^k\| \leq q_+^2 \|u^{k-1}\| \leq \dots \leq q_+^{k+1} \|u^0\| \leq q_+^{k+1} \varepsilon \leq \varepsilon \leq \bar{\varepsilon},$$

$$\begin{aligned}
\left\| \frac{u^{k+1}}{\|u^{k+1}\|} - \bar{v} \right\| &\leq \left\| \frac{u_2^k}{\|u_2^k\|} - \bar{v} \right\| + \left\| \frac{u^{k+1}}{\|u^{k+1}\|} - \frac{u_2^k}{\|u_2^k\|} \right\| \\
&\leq \left\| \frac{u_2^{k-1}}{\|u_2^{k-1}\|} - \bar{v} \right\| + \left\| \frac{u_2^k}{\|u_2^k\|} - \frac{u_2^{k-1}}{\|u_2^{k-1}\|} \right\| + \left\| \frac{u^{k+1}}{\|u^{k+1}\|} - \frac{u_2^k}{\|u_2^k\|} \right\| \\
&\leq \dots \\
&\leq \left\| \frac{u_2^0}{\|u_2^0\|} - \bar{v} \right\| + \sum_{j=1}^k \left\| \frac{u_2^j}{\|u_2^j\|} - \frac{u_2^{j-1}}{\|u_2^{j-1}\|} \right\| + \left\| \frac{u^{k+1}}{\|u^{k+1}\|} - \frac{u_2^k}{\|u_2^k\|} \right\| \\
&\leq 2\delta + \sum_{j=1}^k \frac{2C}{q_-} \left( \frac{\|u_1^{j-1}\|}{\|u^{j-1}\|} + \|u^{j-1}\| \right) + \frac{2C}{q_-} \left( \frac{\|u_1^k\|}{\|u^k\|} + \|u^k\| \right) \\
&\leq 2\delta + \frac{2C}{q_-} \sum_{j=0}^k \left( \frac{\|u_1^j\|}{\|u^j\|} + \|u^j\| \right) \\
&\leq 2\delta + \frac{2C}{q_-} \left( \frac{\|u_1^0\|}{\|u^0\|} + \sum_{j=1}^k \frac{C\|u^{j-1}\|}{q_-} + \sum_{j=0}^k \|u^j\| \right) \\
&\leq 2\delta + \frac{2C}{q_-} \left( \delta + \frac{C}{q_-} \sum_{j=1}^k q_+^{j-1} \varepsilon + \sum_{j=0}^k q_+^j \varepsilon \right) \\
&\leq 2\delta + \frac{2C}{q_-} \left( \delta + \frac{C}{q_-} \frac{\varepsilon}{1 - q_+} + \frac{\varepsilon}{1 - q_+} \right) \\
&\leq 2\delta + \frac{2C}{q_-} \left( \delta + \left( \frac{C}{q_-} + 1 \right) \frac{\varepsilon}{1 - q_+} \right) \\
&\leq \bar{\delta},
\end{aligned}$$

where the last inequality is by (39). Therefore, (9) holds with  $u = u^{k+1}$ .

We have thus established that there exists the unique sequence  $\{u^k\} \subset \mathbb{R}^p$  such that for each  $k$  the point  $u^k$  satisfies (2) and (9) with  $u = u^k$ . By (35) and (36), it then follows that  $u^k \neq \bar{u}$  for all  $k$ , and  $\{u^k\}$  converges to 0.

According to (32) and (35), it holds that

$$\frac{\|u_1^{k+1}\|}{\|u_2^{k+1}\|} \leq \frac{C}{q_-} \|u^k\|$$

for all  $k$ . This yields (27).

Furthermore, according to (34),

$$\frac{1}{2} - C \frac{\|u_1^k\| + \|u^k\|^2}{\|u_2^k\|} \leq \frac{\|u_2^{k+1}\|}{\|u_2^k\|} \leq \frac{1}{2} + C \frac{\|u_1^k\| + \|u^k\|^2}{\|u_2^k\|},$$

where by (27) both sides tend to  $1/2$  as  $k \rightarrow \infty$ . This gives (28).  $\square$

*Remark 2* In Theorem 1, by the monotonicity of the sequence  $\{\|u^k - \bar{u}\|\}$ , for every  $k$  large enough it holds that  $\|u_2^{k+1} - \bar{u}_2\| \leq \|u^{k+1} - \bar{u}\| \leq \|u^k - \bar{u}\|$ . Therefore, (27) implies the estimates

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_{k+1} - \bar{u}\|} = O(\|u^k - \bar{u}\|)$$

and

$$\|u_1^{k+1} - \bar{u}_1\| = O(\|u^k - \bar{u}\|^2) \quad (40)$$

as  $k \rightarrow \infty$ .

Furthermore,

$$\frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_1^k - \bar{u}_1\| + \|u_2^k - \bar{u}_2\|} \leq \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u^k - \bar{u}\|} \leq \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|},$$

where by (27) and (28) both sides tend to  $1/2$  as  $k \rightarrow \infty$ . Therefore,

$$\lim_{k \rightarrow \infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u^k - \bar{u}\|} = \frac{1}{2}. \quad (41)$$

Finally,

$$\frac{\|u_2^{k+1} - \bar{u}_2\| - \|u_1^{k+1} - \bar{u}_1\|}{\|u^k - \bar{u}\|} \leq \frac{\|u^{k+1} - \bar{u}\|}{\|u^k - \bar{u}\|} \leq \frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u^k - \bar{u}\|} + \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u^k - \bar{u}\|}$$

where by (40) and (41) both sides tend to  $1/2$  as  $k \rightarrow \infty$ . Therefore,

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

Theorem 1 establishes the existence of a set with nonempty interior, which is star-like with respect to  $\bar{u}$ , and such that any sequence satisfying the pNM relation (2) and initialized from any point of this set, converges linearly to  $\bar{u}$ . Moreover, if  $\Phi$  is 2-regular at  $\bar{u}$  in at least one direction  $\bar{v} \in \ker \Phi'(\bar{u})$ , then the set of such  $\bar{v}$  is open and dense in  $\ker \Phi'(\bar{u}) \cap \mathbb{S}$ : its complement is the null set of the nontrivial homogeneous polynomial  $\det \bar{\mathcal{B}}(\cdot)$  considered on  $\ker \Phi'(\bar{u}) \cap \mathbb{S}$ . The union of convergence domains coming with all such  $\bar{v}$  is also a star-like convergence domain with nonempty interior. In the case when  $\Phi'(\bar{u}) = 0$  (full degeneracy), this domain is quite large. In particular, it is ‘‘asymptotically dense’’: the only excluded directions are those in which  $\Phi$  is not 2-regular at  $\bar{u}$ , which is the null set of a nontrivial homogeneous polynomial. Beyond the case of full degeneracy, the convergence

domain given by Theorem 1 is at least not “asymptotically thin”. Though it is also not “asymptotically dense”.

For the (unperturbed) Newton method, the existence of “asymptotically dense” star-like domain of convergence was established in [11, Theorem 6.1]. Specifically, it was demonstrated that one Newton step from a point  $u^0$  in this domain leads to the convergence domain coming with the appropriate  $\bar{v} = \pi(u^0)/\|\pi(u^0)\|$ , where

$$\pi(u^0) = \frac{1}{2}(u_2^0 - \bar{u}_2) + \frac{1}{2}(\bar{\mathcal{B}}(u^0 - \bar{u}))^{-1}PB[u^0 - \bar{u}, u_1^0 - \bar{u}_1];$$

see (22) above. Deriving a result like this for the pNM scheme (2) is hardly possible in general, at least without rather restrictive assumptions on perturbation terms. Perhaps results along those lines can be derived for specific methods, rather than for the general pNM framework, but such developments are not known at this time.

### 3 Applications to some specific algorithms

In this section we show how our general results for the pNM framework (2) can be applied to some specific methods. In particular, we consider the following algorithms, all developed for tackling the difficult case of non-isolated solutions: the Levenberg–Marquardt method and the LP-Newton method for equations, and the stabilized Newton–Lagrange method for optimization.

We start with observing that the assumptions (5), (8), (24) and (25) on the perturbations terms in Theorem 1 hold automatically if

$$\Omega(u) = O(\|\Phi(u)\|), \quad \omega(u) = O(\|u - \bar{u}\|\|\Phi(u)\|) \quad (42)$$

as  $u \rightarrow \bar{u}$ . Indeed, this readily follows from the relation

$$\Phi(u) = \Phi'(\bar{u})(u - \bar{u}) + O(\|u - \bar{u}\|^2)$$

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

Thus, to apply Theorem 1, in the part of perturbations properties it is sufficient to verify (42).

#### 3.1 Levenberg–Marquardt method

An iteration of the classical Levenberg–Marquardt method [24, Chapter 10.2] consists in solving the following subproblem:

$$\text{minimize } \frac{1}{2}\|\Phi(u^k) + \Phi'(u^k)v\|^2 + \frac{1}{2}\sigma(u^k)\|v\|^2, \quad v \in \mathbb{R}^p, \quad (43)$$

where  $u^k \in \mathbb{R}^p$  is the current iterate and  $\sigma : \mathbb{R}^p \rightarrow \mathbb{R}_+$  defines the regularization parameter.

In [29], it was established that under the Lipschitzian error bound condition (i.e., being initialized near a noncritical solution  $\bar{u}$  of (1)), the method described by (43) with  $\sigma(u) = \|\Phi(u)\|^2$  generates a sequence which is quadratically convergent to a (nearby) solution of (1). For analysis under the Lipschitzian error bound condition of a rather general framework that includes the Levenberg–Marquardt method, see [5, 8]. Our interest here is the case of critical solutions, when the error bound does *not* hold.

First, note that the unique (if  $\sigma(u^k) > 0$ ) minimizer of the convex quadratic function in (43) is characterized by the linear system

$$(\Phi'(u^k))^T \Phi(u^k) + ((\Phi'(u^k))^T \Phi'(u^k) + \sigma(u^k)I)v = 0. \quad (44)$$

We next show how (44) can be embedded into the pNM framework (2). In particular, we construct the perturbation terms for which the conditions (42) hold, and hence Theorem 1 is applicable, and which correspond to (44) on the relevant domain of convergence. As a result, we obtain the following convergence assertions.

**Corollary 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  $\bar{u}$ . Let  $\bar{u}$  be a solution of the equation (1), and assume that  $\Phi$  is 2-regular at  $\bar{u}$  in a direction  $\bar{v} \in \ker \Phi'(\bar{u}) \cap \mathbb{S}$ .*

*Then for any  $\tau \geq 2$ , there exist  $\varepsilon = \varepsilon(\bar{v}) > 0$  and  $\delta = \delta(\bar{v}) > 0$  such that any starting point  $u^0 \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying (26) uniquely defines the sequence  $\{u^k\} \subset \mathbb{R}^p$  such that for each  $k$  it holds that  $v^k = u^{k+1} - u^k$  solves (43) with  $\sigma(u) = \|\Phi(u)\|^\tau$ ,  $u_2^k \neq \bar{u}_2$ , the sequence  $\{u^k\}$  converges to  $\bar{u}$ , the sequence  $\{\|u^k - \bar{u}\|\}$  converges to zero monotonically, and (27) and (28) hold. Moreover, if  $\Phi'(\bar{u}) = 0$ , then the same assertion is valid with any  $\tau \geq 3/2$ .*

*Proof* Define  $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$  and  $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$  according to Remark 1, where we set  $\Omega \equiv 0$ . Define the set

$$K = K(\bar{v}) = \{u \in \mathbb{R}^p \setminus \{\bar{u}\} \mid (9) \text{ holds}\}. \quad (45)$$

Then  $\Phi'(u)$  is invertible for all  $u \in K$ , and  $(\Phi'(u))^{-1} = O(\|u - \bar{u}\|^{-1})$  as  $u \rightarrow \bar{u}$  (this can also be concluded directly from [11, Lemma 3.1], with an appropriate choice of  $\bar{\varepsilon} > 0$  and  $\bar{\delta} > 0$ ).

We next define the mappings  $\Omega$  and  $\omega$ . First, we set  $\Omega(u) = 0$  and  $\omega(u) = 0$  for all  $u \in \mathbb{R}^p \setminus K$ . Of course, those vanishing perturbation terms for  $u \notin K$  have no relation to (44). The point is that we next show that with the appropriate definitions for  $u \in K$ , we obtain  $\Omega$  and  $\omega$  satisfying (42). Then Theorem 1 ensures that if the starting point satisfies (26) with appropriate  $\varepsilon > 0$  and  $\delta > 0$ , it follows that the subsequent iterates are well defined and remain in the set  $K$ . Finally, in  $K$ , the constructed  $\Omega$  and  $\omega$  do correspond to (44).

Let  $u \in K$ . Considering (44) with  $u^k = u$ , and multiplying both sides of this relations by the matrix  $((\Phi'(u^k))^T)^{-1} = (((\Phi'(u^k))^{-1})^T)$ , we obtain that

$$\Phi(u^k) + (\Phi'(u^k) + \sigma(u^k)((\Phi'(u^k))^{-1})^T)v = 0,$$

which is the pNM iteration system (2) with the perturbation terms given by

$$\Omega(u) = \sigma(u)((\Phi'(u))^{-1})^T, \quad \omega(u) = 0, \quad u \in K.$$

Then

$$\Omega(u) = O(\|u - \bar{u}\|^{-1} \sigma(u)).$$

Therefore, the needed first estimate in (42) would hold if

$$\sigma(u) = O(\|u - \bar{u}\| \|\Phi(u)\|) \quad (46)$$

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

Let  $\sigma(u) = \|\Phi(u)\|^\tau$  with  $\tau > 0$ . Then (46) takes the form

$$\|\Phi(u)\|^{\tau-1} = O(\|u - \bar{u}\|),$$

and since  $\Phi(u) = O(\|u - \bar{u}\|)$ , the last estimate is satisfied as  $u \rightarrow \bar{u}$  if  $\tau \geq 2$ .

Moreover, in the case when  $\Phi'(\bar{u}) = 0$  (full singularity) it holds that  $\Phi(u) = O(\|u - \bar{u}\|^2)$  as  $u \rightarrow \bar{u}$ , and hence, in this case, the appropriate values are all  $\tau \geq 3/2$ .

The construction is complete. As the exhibited  $\Omega$  and  $\omega$  satisfy (42) as  $u \rightarrow \bar{u}$  (regardless of whether  $u$  stays in  $K$  or not), Theorem 1 is applicable. In particular, it guarantees that for appropriate starting points all the iterates stay in  $K$ . In this set, the perturbation terms define the Levenberg–Marquardt iterations (44). Thus, the assertions follow from Theorem 1.  $\square$

The following example is taken from DEGEN test collection [4].

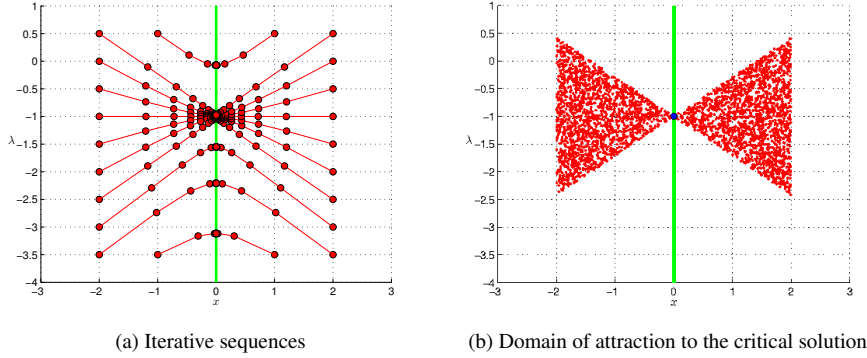


Fig. 1: Levenberg–Marquardt method with  $\tau = 1$  for Example 1.

*Example 1 (DEGEN 20101)* Consider the equality-constrained optimization problem

$$\begin{aligned} &\text{minimize } x^2 \\ &\text{subject to } x^2 = 0. \end{aligned}$$

Stationary points and associated Lagrange multipliers of this problem are characterized by the Lagrange optimality system which has the form of the nonlinear equation (1) with

$$\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \Phi(u) = (2x(1 + \lambda), x^2), \quad u = (x, \lambda).$$

The unique feasible point (hence, the unique solution, and the unique stationary point) of this problem is  $\bar{x} = 0$ , and the set of associated Lagrange multipliers is the entire  $\mathbb{R}$ . Therefore, the solution set of the Lagrange system (i.e., the primal-dual solution set) is  $\{\bar{x}\} \times \mathbb{R}$ . The unique critical solution is  $\bar{u} = (\bar{x}, \bar{\lambda})$  with  $\bar{\lambda} = -1$ , the one for which  $\Phi'(\bar{u}) = 0$  (full singularity).

In Figures 1 and 2, the vertical line corresponds to the primal-dual solution set. These figures show some iterative sequences generated by the Levenberg–Marquardt method, and the domains from which convergence to the critical solution was detected. Using zoom in, and taking smaller areas for starting points, does not significantly change the picture in Figure 1, corresponding to  $\tau = 1$ . At the same time, such manipulations with Figures 2a and 2b put in evidence that for  $\tau = 3/2$ , the domain of convergence is in fact asymptotically dense (see Figures 2c and 2d). These observations are in agreement with Theorem 1.

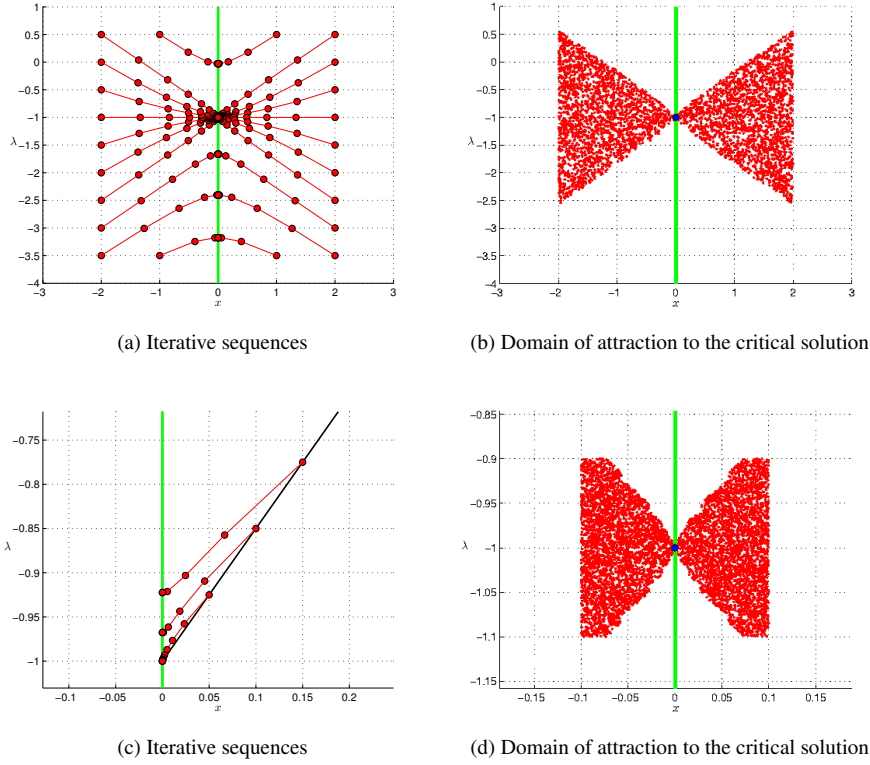


Fig. 2: Levenberg–Marquardt method with  $\tau = 3/2$  for Example 1.

### 3.2 LP-Newton method

The LP-Newton method was introduced in [6]. For the equation (1), the iteration subproblem of this method has the form

$$\begin{aligned}
 & \text{minimize } \gamma \\
 & \text{subject to } \|\Phi(u^k) + \Phi'(u^k)v\| \leq \gamma \|\Phi(u^k)\|^2, \\
 & \|v\| \leq \gamma \|\Phi(u^k)\|, \\
 & (v, \gamma) \in \mathbb{R}^p \times \mathbb{R}.
 \end{aligned} \tag{47}$$

The subproblem (47) always has a solution if  $\Phi(u^k) \neq 0$  (naturally, if  $\Phi(u^k) = 0$  the method stops). If the  $l_\infty$ -norm is used, this is a linear programming problem (hence the name). As demonstrated in [5, 6] (see also [8]), local convergence properties of the LP-Newton method (under the error bound condition, i.e., near noncritical solutions) are the same as for the Levenberg–Marquardt algorithm. Again, our setting is rather that of *critical* solutions.

The proof of the following result is again by placing (47) within the pNM framework (2). It is interesting that in this case  $\Omega \equiv 0$ , while  $\omega(\cdot)$  is defined *implicitly*: there is no analytic expression for it.

**Corollary 2** *Under the assumptions of Corollary 1, there exist  $\varepsilon = \varepsilon(\bar{v}) > 0$  and  $\delta = \delta(\bar{v}) > 0$  such that for any starting point  $u^0 \in \mathbb{R}^p \setminus \{\bar{u}\}$  satisfying (26) the following assertions are valid:*

- (a) *There exists a sequence  $\{u^k\} \subset \mathbb{R}^p$  such that for each  $k$  the pair  $(v^k, \gamma_{k+1})$  with  $v^k = u^{k+1} - u^k$  and some  $\gamma_{k+1}$  solves (47).*
- (b) *For any such sequence,  $u_2^k \neq \bar{u}_2$  for each  $k$ , the sequence  $\{u^k\}$  converges to  $\bar{u}$ , the sequence  $\{\|u^k - \bar{u}\|\}$  converges to zero monotonically, and (27) and (28) hold.*

*Proof* By the second constraint in (47), the equality (2) holds for  $\Omega \equiv 0$  and some  $\omega(\cdot)$  satisfying

$$\|\omega(u)\| \leq \gamma(u) \|\Phi(u)\|^2, \quad (48)$$

where  $\gamma(u)$  is the optimal value of the subproblem (47) with  $u^k = u$ . Note that  $\omega(\cdot)$  would satisfy (42), and thus the assumptions of Theorem 1, if

$$\gamma(u) = O(\|\Phi(u)\|^{-1} \|u - \bar{u}\|) \quad (49)$$

as  $u \rightarrow \bar{u}$ . Indeed, from the first constraint in (47) we then obtain that

$$\|\Phi(u) + \Phi'(u)v\| \leq \gamma(u) \|\Phi(u)\|^2 = O(\|u - \bar{u}\| \|\Phi(u)\|),$$

which implies the second estimate in (42).

We thus have to establish (49) on the relevant set. In this respect, the construction is similar to what had been done in Section 3.1 above.

Define  $\bar{\varepsilon} = \bar{\varepsilon}(\bar{v}) > 0$  and  $\bar{\delta} = \bar{\delta}(\bar{v}) > 0$  according to Lemma 1 applied with  $\Omega \equiv 0$  and  $\omega \equiv 0$ , and define the set  $K$  according to (45). Then the step  $v(u)$  of the (unperturbed) Newton method from any point  $u \in K$  exists, it is uniquely defined, and by (10) and (11), it holds that  $v(u) = O(\|u - \bar{u}\|)$ .

We now define the mappings  $\Omega$  and  $\omega$ . Similarly to the case of the Levenberg–Marquardt method, we first set them identically equal to zero on  $\mathbb{R}^p \setminus K$ . Let  $u \in K$ . Then the point  $(v, \gamma) = (v(u), \|v(u)\|/\|\Phi(u)\|)$  is feasible in (47), and hence,

$$\gamma(u) \leq \gamma = \|\Phi(u)\|^{-1} \|v(u)\| = O(\|\Phi(u)\|^{-1} \|u - \bar{u}\|)$$

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

As in the proof of Corollary 1, we note that the constructed  $\Omega$  and  $\omega$  satisfy (42) as  $u \rightarrow \bar{u}$ . Therefore, Theorem 1 ensures that for appropriate starting points all the iterates stay in  $K$ , and in this set the perturbation terms defined hereby correspond to (47). In particular, the assertions then follow from Theorem 1.  $\square$

Observe that for the LP-Newton method, the values of  $\omega(\cdot)$  are defined in a posteriori manner, *after*  $v^k$  is computed. For this reason, Theorem 1 cannot yield uniqueness of the iterative sequence: the next iterate can be defined by any  $\omega(\cdot)$  satisfying (48) for  $u = u^k$ , and different choices of appropriate  $\omega(\cdot)$  may give rise to different next iterates.

Figure 3 shows the same information for the LP-Newton method as Figure 2 for the Levenberg–Marquardt algorithm, with the same conclusions.



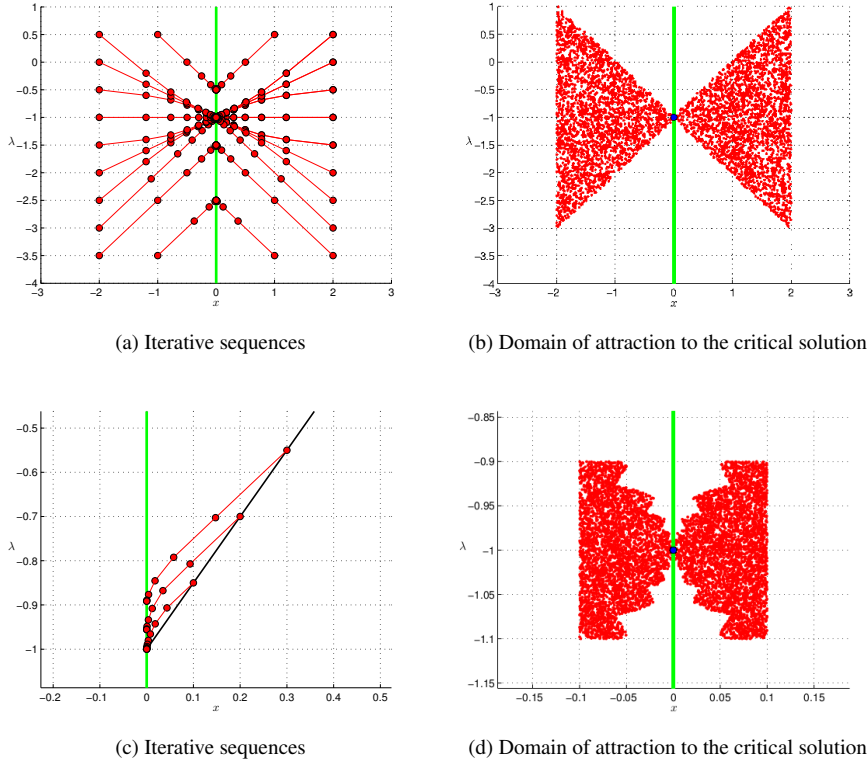


Fig. 3: LP-Newton method for Example 1.

### 3.3 Equality-constrained optimization and the stabilized Newton–Lagrange method

We next turn our attention to the origin of the critical solutions issues, namely, to the equality-constrained optimization problem

$$\text{minimize } f(x) \text{ subject to } h(x) = 0, \quad (50)$$

where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$  are smooth. The Lagrangian  $L : \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}$  of this problem is given by

$$L(x, \lambda) = f(x) + \langle \lambda, h(x) \rangle.$$

Then stationary points and associated Lagrange multipliers of (50) are characterized by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x, \lambda) = 0, \quad h(x) = 0,$$

with respect to  $x \in \mathbb{R}^n$  and  $\lambda \in \mathbb{R}^l$ .

The Lagrange optimality system is a special case of nonlinear equation (1), corresponding to setting  $p = q = n + l$ ,  $u = (x, \lambda)$ ,

$$\Phi(u) = \left( \frac{\partial L}{\partial x}(x, \lambda), h(x) \right). \quad (51)$$

The stabilized Newton–Lagrange method (or stabilized sequential quadratic programming) was developed for solving the Lagrange optimality system (or optimization problem) when the multipliers associated to a stationary point need not be unique [28, 12, 7, 20]; see also [21, Chapter 7]. For the current iterate  $u^k = (x^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$ , the iteration subproblem of this method is given by

$$\begin{aligned} & \text{minimize } \langle f'(x^k), \xi \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \xi, \xi \right\rangle + \frac{\sigma(u^k)}{2} \|\eta\|^2 \\ & \text{subject to } h(x^k) + h'(x^k)\xi - \sigma(u^k)\eta = 0, \end{aligned}$$

where the minimization is in the variables  $(\xi, \eta) \in \mathbb{R}^n \times \mathbb{R}^l$ , and  $\sigma : \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}_+$  now defines the stabilization parameter. Equivalently, the following linear system (characterizing stationary points and associated Lagrange multipliers of this subproblem) is solved:

$$\frac{\partial L}{\partial x}(x^k, \lambda^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi + (h'(x^k))^T \eta = 0, \quad h(x^k) + h'(x^k)\xi - \sigma(u^k)\eta = 0. \quad (52)$$

With a solution  $(\xi^k, \eta^k)$  of (52) at hand, the next iterate is given by  $u^{k+1} = (x^k + \xi^k, \lambda^k + \eta^k)$ . Note that if  $\sigma \equiv 0$ , then (52) becomes the usual Newton–Lagrange method, i.e., the basic Newton method applied to the Lagrange optimality system.

For a given Lagrange multiplier  $\bar{\lambda}$  associated with a stationary point  $\bar{x}$  of problem (50), define the linear subspace

$$\mathcal{Q}(\bar{x}, \bar{\lambda}) = \left\{ \xi \in \ker h'(\bar{x}) \mid \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \text{im}(h'(\bar{x}))^T \right\}.$$

Recall that the multiplier  $\bar{\lambda}$  is called critical if  $\mathcal{Q}(\bar{x}, \bar{\lambda}) \neq \{0\}$ ; see [13, 18]. Otherwise  $\bar{\lambda}$  is noncritical.

As demonstrated in [20], if initialized near a primal-dual solution with a noncritical dual part, the stabilized Newton–Lagrange method with  $\sigma(u) = \|\Phi(u)\|$ , where  $\Phi$  is given by (51), generates a sequence which is superlinearly convergent to a (nearby) solution. Again, of current interest is the *critical* case.

Evidently, the iteration (52) fits the pNM framework (2) for  $\Phi$  defined in (51), taking

$$\Omega(u) = \begin{pmatrix} 0 & 0 \\ 0 & -\sigma(u)I \end{pmatrix}, \quad \omega \equiv 0.$$

These perturbations satisfy the assumptions in Theorem 1 if, e.g.,  $\sigma(u) = \|\Phi(u)\|^\tau$ ,  $\tau \geq 1$ . We thus conclude the following.

**Corollary 3** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$  be three times differentiable near a stationary point  $\bar{x} \in \mathbb{R}^n$  of problem (50), with their third derivatives Lipschitz-continuous with respect to  $\bar{x}$ , and let  $\bar{\lambda} \in \mathbb{R}^l$  be a Lagrange multiplier associated to  $\bar{x}$ . Assume that the mapping  $\Phi$  defined in (51) is 2-regular at  $\bar{u} = (\bar{x}, \bar{\lambda})$  in a direction  $\bar{v} \in \ker \Phi'(\bar{u}) \cap \mathbb{S}$ .*

*Then for any  $\tau \geq 1$ , there exist  $\varepsilon = \varepsilon(\bar{v}) > 0$  and  $\delta = \delta(\bar{v}) > 0$  such that any starting point  $u^0 = (x^0, \lambda^0) \in (\mathbb{R}^n \times \mathbb{R}^l) \setminus \{\bar{u}\}$  satisfying (26) uniquely defines a sequence  $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$  such that for each  $k$  it holds that  $(\xi^k, \eta^k) = (x^{k+1} - x^k, \lambda^{k+1} - \lambda^k)$  solves (52) with  $\sigma(u) = \|\Phi(u)\|^\tau$ ,  $u_2^k \neq \bar{u}_2$ , the sequence  $\{u^k\}$  converges to  $\bar{u}$ , the sequence  $\{\|u^k - \bar{u}\|\}$  converges to zero monotonically, and (27) and (28) hold.*

Table 1: Cases of convergence to  $\bar{\lambda} = -1$  in Example 1 (%)

$\varepsilon$	L-M $\tau = 1$	L-M $\tau = 3/2$	L-M $\tau = 2$	LP-N	sN-L	N-L
1	38	44	56	54	42	83
0.5	34	52	68	55	57	86
0.25	36	57	79	58	73	86
0.1	38	68	86	73	83	90
0.01	38	84	86	86	88	86

Table 2: Cases of convergence to  $\bar{\lambda} = (1, \sqrt{3})$  in Example 2 (%)

$\varepsilon$	L-M $\tau = 1$	L-M $\tau = 3/2$	L-M $\tau = 2$	LP-N	sN-L	N-L
1	21	26	31	42	4	65
0.5	21	27	37	45	7	75
0.25	23	29	47	48	15	82
0.1	23	33	63	55	26	91
0.01	24	60	89	82	59	97

It is worth to mention that the above results and discussion can be extended to variational problems (of which optimization is a special case), see [7], [21, Chapter 7].

We proceed with some further considerations. A multiplier is said to be critical of order one if  $\dim Q(\bar{x}, \bar{\lambda}) = 1$ . The following was established in [15].

**Proposition 2** *Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  and  $h : \mathbb{R}^n \rightarrow \mathbb{R}^l$  be three times differentiable at a stationary point  $\bar{x} \in \mathbb{R}^n$  of problem (50), and let  $\bar{\lambda} \in \mathbb{R}^l$  be a Lagrange multiplier associated to  $\bar{x}$ . Let  $Q(\bar{x}, \bar{\lambda})$  be spanned by some  $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$ , i.e.,  $\bar{\lambda}$  is a critical multiplier of order one.*

*If  $\text{rank } h'(\bar{x}) = l - 1$ , then  $\ker \Phi'(\bar{u})$  contains elements of the form  $v = (\bar{\xi}, \eta)$  with some  $\eta \in \mathbb{R}^l$ , and  $\Phi$  is 2-regular at  $\bar{u}$  in every such direction if and only if  $h''(\bar{x})[\bar{\xi}, \bar{\xi}] \notin \text{im } h'(\bar{x})$ .*

*If  $\text{rank } h'(\bar{x}) \leq l - 2$ , then  $\Phi$  cannot be 2-regular at  $\bar{u}$  in any direction  $v \in \ker \Phi'(\bar{u})$ .*

*If  $h'(\bar{x}) = 0$ , and  $l \geq 2$  or  $h''(\bar{x})[\bar{\xi}, \bar{\xi}] = 0$ , then  $\Phi$  cannot be 2-regular at  $\bar{u}$  in any direction  $v \in \ker \Phi'(\bar{u})$ .*

In the last two cases specified in the proposition above, Theorem 1 is not applicable; these cases require special investigation. In the last case, when  $l \geq 2$  but  $h''(\bar{x})[\bar{\xi}, \bar{\xi}] \neq 0$ , allowing for non-isolated critical multipliers, the effect of attraction of the basic Newton-Lagrange method to such multipliers had been studied in [23] for fully quadratic problems.

On the other hand, if, e.g.,  $l = 1$ ,  $h'(\bar{x}) = 0$ , and  $h''(\bar{x})[\bar{\xi}, \bar{\xi}] \neq 0$ , then Theorem 1 is applicable with  $\bar{v} = (\bar{\xi}, \eta)$  for every  $\eta \in \mathbb{R}$ . Taking here  $\eta = 0$  recovers the results in [27] for the basic Newton-Lagrange method and the fully quadratic case. Moreover, this is exactly the situation that we have in Example 1. Table 1 reports on the percentage of detected cases of dual convergence to the unique critical multiplier  $\bar{\lambda} = -1$  in Example 1, for the algorithms discussed above, depending on the size  $\varepsilon$  of the region for starting points around  $(\bar{x}, \bar{\lambda})$ . In the table, L-M refers to the Levenberg–Marquardt method (with different values for the power  $\tau$  that defines the regularization parameter), LP-N refers to the LP-Newton method, N-L to the Newton-Lagrange method (i.e., (52) with  $\sigma \equiv 0$ ), and sN-L to the stabilized Newton-Lagrange method.

The case when  $\dim Q(\bar{x}, \bar{\lambda}) \geq 2$  (i.e., when  $\bar{\lambda}$  is critical of order higher than 1) opens wide possibilities for 2-regularity in the needed directions, and such solutions are often specially attractive for Newton-type iterates.

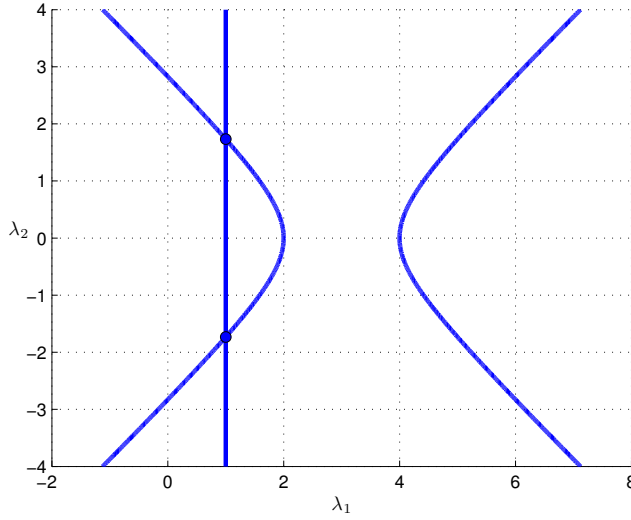


Fig. 4: Critical multipliers in Example 2.

*Example 2 (DEGEN 20302)* Consider the equality-constrained optimization problem

$$\begin{aligned} & \text{minimize } x_1^2 - x_2^2 + 2x_3^2 \\ & \text{subject to } -\frac{1}{2}x_1^2 + x_2^2 - \frac{1}{2}x_3^2 = 0, \quad x_1x_3 = 0. \end{aligned}$$

Here,  $\bar{x} = 0$  is the unique solution,  $h'(\bar{x}) = 0$ , and the set of associated Lagrange multipliers is the entire  $\mathbb{R}^2$ . Critical multipliers are those satisfying  $\lambda_1 = 1$  or  $(\lambda_1 - 3)^2 - \lambda_2^2 = 1$ . In Figure 4, critical multipliers are those forming the vertical straight line and two branches of the hyperbola.

According to Proposition 2, the 2-regularity property cannot hold for  $\Phi$  defined in (51), at  $(\bar{x}, \lambda)$  for any direction  $v \in \ker \Phi'(\bar{u})$ , for all critical multipliers  $\lambda$ , except for  $\bar{\lambda} = (1, \pm\sqrt{3})$ , which are the two intersection points of the vertical line and the hyperbola. One can directly check that the mapping  $\Phi$  is indeed 2-regular at  $\bar{u} = (\bar{x}, \bar{\lambda})$  in some directions  $v \in \ker \Phi'(\bar{u})$ .

Table 2 reports on the percentage of detected cases of dual convergence to  $\bar{\lambda} = (1, \sqrt{3})$ , for the algorithms discussed above, depending on the size  $\varepsilon$  of the region for starting points around  $(\bar{x}, \bar{\lambda})$ .

**Acknowledgment.** The authors thank the two anonymous referees for their comments on the original version of this article.

## References

1. A.V. Arutyunov. *Optimality Conditions: Abnormal and Degenerate Problems*. Kluwer Academic Publishers, Dordrecht, 2000.

2. E.R. Avakov. Extremum conditions for smooth problems with equality-type constraints. *USSR Comput. Math. Math. Phys.* 25 (1985), 24–32.
3. F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley, New York, 1983.
4. DEGEN. [http://w3.impa.br/~optim/DEGEN\\_collection.zip](http://w3.impa.br/~optim/DEGEN_collection.zip).
5. F. Facchinei, A. Fischer, and M. Herrich. A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. *Math. Methods Oper. Res.* 77 (2013), 433–443.
6. F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions. *Math. Program.* 146 (2014), 1–36.
7. 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.
8. 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.
9. H. Gfrerer and B.S. Mordukhovich. Complete characterizations of tilt stability in nonlinear programming under weakest qualification conditions. *SIAM J. Optim.* 25 (2015), 2081–2119.
10. H. Gfrerer and J.V. Outrata. On computation of limiting coderivatives of the normal-cone mapping to inequality systems and their applications. *Optim.* 65 (2016), 671–700.
11. A. Griewank. Starlike domains of convergence for Newton’s method at singularities. *Numer. Math.* 35 (1980), 95–111.
12. W.W. Hager. Stabilized sequential quadratic programming. *Comput. Optim. Appl.* 12 (1999), 253–273.
13. A.F. Izmailov. On the analytical and numerical stability of critical Lagrange multipliers. *Comput. Math. Math. Phys.* 45 (2005), 930–946.
14. A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. *Math. Program.* 142 (2013), 591–604.
15. A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Critical solutions of nonlinear equations: stability issues. *Math. Program.*, 2016. DOI 10.1007/s10107-016-1047-x.
16. A.F. Izmailov and M.V. Solodov. Error bounds for 2-regular mappings with Lipschitzian derivatives and their applications. *Math. Program.* 89 (2001), 413–435.
17. A.F. Izmailov and M.V. Solodov. The theory of 2-regularity for mappings with Lipschitzian derivatives and its applications to optimality conditions. *Math. Oper. Res.* 27 (2002), 614–635.
18. A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. *Math. Program.* 117 (2009), 271–304.
19. A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. *Math. Program.* 126 (2011), 231–257.
20. A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. *Math. Program.* 122 (2012), 93–120.
21. A.F. Izmailov and M.V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering, 2014.
22. A.F. Izmailov and M.V. Solodov. Critical Lagrange multipliers: what we currently know about them, how they spoil our lives, and what we can do about it. *TOP* 23 (2015), 1–26. Rejoinder of the discussion: *TOP* 23 (2015), 48–52.
23. A.F. Izmailov and E.I. Uskov. Attraction of Newton method to critical Lagrange multipliers: fully quadratic case. *Math. Program.* 152 (2015), 33–73.
24. J. Nocedal and S.J. Wright. *Numerical Optimization*. 2nd edition. Springer-Verlag, New York, Berlin, Heidelberg, 2006.
25. C. Oberlin and S.J. Wright. An accelerated Newton method for equations with semismooth Jacobians and nonlinear complementarity problems. *Math. Program.* 117 (2009), 355–386.
26. R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*. Springer-Verlag, Berlin, 1998.
27. E.I. Uskov. On the attraction of Newton method to critical Lagrange multipliers. *Comp. Math. Math. Phys.* 53 (2013), 1099–1112.
28. S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.* 11 (1998), 253–275.
29. N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg–Marquardt method. *Computing. Suppl.* 15 (2001), 237–249.