A GENERAL PERTURBED NEWTONIAN FRAMEWORK AND CRITICAL SOLUTIONS OF NONLINEAR EQUATIONS*

Alexey F. Izmailov[1, 2] and Mikhail V. Solodov[3]

July 11, 2024

ABSTRACT

We survey the perturbed Newton method framework for smooth nonlinear equations, allowing sharp characterization of local convergence and rate of convergence to singular solutions possessing some 2-regularity properties. The framework covers a wide range of Newton-type methods in a unified manner, including, along with the basic Newton method, the Levenberg-Marquardt method and the LP-Newton method, among the others. We also discuss a linesearch-based globalization of convergence of these methods, and their possible acceleration by means of extrapolation, with asymptotic acceptance of the full step playing the key role for acceleration. These constructions and results are further extended to more general problem settings, such as constrained equations and piecewise smooth equations, allowing for applications to various reformulations of complementarity problems. The 2-regularity property in question is strongly related to the concept of critical solutions of nonlinear equations, which is further naturally tailored to the error bound property (or rather lack of it), and to stability of solutions subject to wide classes of perturbations. Finally, we trace the link of critical solutions of equations to critical Lagrange multipliers in optimization, which was the origin of these developments. Critical Lagrange multipliers have a major effect on behavior of the SQP methods, stabilized SQP, and multiplier (Augmented Lagrangian) methods for optimization and variational problems.

Key words: nonlinear equation; Newton-type method; perturbed Newton method; singular solution; critical solution; 2-regularity; critical Lagrange multipliers.
AMS subject classifications. 65J15, 49M15, 65K15, 90C33.

---

* The first author was funded by the Russian Science Foundation Grant 23-11-20020 (`https://rscf.ru/en/project/23-11-20020/`). The second author was supported in part by CNPq Grant 306775/2023-9, by FAPERJ Grant E-26/200.347/2023, and by PRONEX–Optimization.
[1] VMK Faculty, OR Department, Lomonosov Moscow State University (MSU), Uchebniy Korpus 2, Leninskiye Gory, 119991 Moscow, Russia.
Email: izmaf@cs.msu.ru
[2] Derzhavin Tambov State University (TSU), Internationalnaya 33, 392000 Tambov, Russia.
[3] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: solodov@impa.br

1 Introduction

The primary problem setting in this survey is the equation

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

where the mapping $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ is at least once differentiable at a given solution $\bar{u}$ of (1.1). The case of interest here is when $\bar{u}$ is a singular solution of (1.1), i.e., $\Phi'(\bar{u})$ is a singular matrix; otherwise solution is called nonsingular. In the latter case, everything is classical and long well understood. For the specified problem setting with the number of equations in (1.1) equal to the number of variables, singular solutions include, in particular, all nonisolated solutions.

We shall discuss behavior of various Newton-type methods near singular solutions, not only for (1.1) itself, but also for some problem settings more general. These include the reduced smoothness hypotheses on Φ and the cases when (1.1) has additional constraints, i.e., constrained equations. The main tool is the perturbed Newton method framework that appears to be a convenient paradigm of wide use in this context. For example, it allows to cover various methods, as well as various reformulations of complementarity problems.

The paper is organized as follows. In Section 2, we recall the key notions of 2-regularity and of criticality of solutions of nonlinear equations. Section 3 describes the general perturbed Newtonian framework and its convergence properties. In Section 4, these are applied to the Levenberg-Marquardt and to the LP-Newton methods. The issues of globalization of the local methods and of acceleration techniques are discussed in Scetion 5. Section 6 is devoted to constrained equations and Section 7 to piecewise smooth equations. Applications to complementarity problems are described in Section 8 (both unconstrained and constrained reformulations). Finally, in Section 9, we go back to connecting the notion of critical solutions of nonlinear equations to that of critical Lagrange multipliers in optimization, which is the origin of the developments. Section 10 collects some questions, which we consider as open at the time of wirting this paper.

We complete the introduction by some comments about the notation and conventions we shall employ in the sequel. For simplicity of presentation, let the inner product $\langle \cdot, \cdot \rangle$ and the norm $\| \cdot \|$ be Euclidian, unless some other choice of the norm is explicitly specified. For a linear subspace L, we denote by $L^\perp$ its orthogonal complement in the given space. The identity matrix is denoted by $\mathcal{I}$. By $\operatorname{diag} u$ we mean the diagonal matrix, with the vector u forming the diagonal. The notation u_I stands for the subvector of u with components u_i, $i \in I$, where I is a given index set. We use $\ker A$ for the null space and $\operatorname{im} A$ for the range space of a linear operator A. For a given set $U \subset \mathbb{R}^p$, the distance from $u \in \mathbb{R}^p$ to U is given by $\operatorname{dist}(u, U) = \inf_{v \in U} \|u - v\|$. The interior of U is denoted by $\operatorname{int} U$. Furthermore, $T_U(u)$ is the contingent cone to U at a point $u \in U$ [56, p. 3], $\widehat{T}_U(u)$ is the Clarke tangent cone [10, p. 51] (when U is convex, both coincide with the standard tangent cone, i.e., the closure of the conic hull of $U - u$), and $N_U(u)$ is the Mordukhovich (limiting) normal cone [70, Definition 1.1]. For any $\bar{u}$, $v \in \mathbb{R}^p$, and any given scalars $\varepsilon > 0$ and $\delta > 0$, we define the set

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

which can be thought of as a "conic neighborhood" of u associated to the direction v. Finally, $\partial\Psi(u)$ stands for Clarke's generalized Jacobian of a mapping Ψ at u [10, Definition 2.6.1].

2 Preliminaries: 2-regularity and criticality of solutions

Assume now that Φ is differentiable at every $u \in \mathbb{R}^p$ near the given solution $\bar{u}$ of (1.1). Assume further that its derivative Φ', as a mapping from a neighborhood of $\bar{u}$ to $\mathbb{R}^{p\times p}$, is strongly semismooth at $\bar{u}$, i.e., it is Lipschitz-continuous near $\bar{u}$, directionally differentiable at $\bar{u}$ in every direction, and satisfies the estimate

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

as $u \to \bar{u}$. For the concept of strong semismoothness, see, e.g., [75], and also [56, Section 1.4.2] for a more recent exposition of the related theory.

We emphasize that we do not assume twice differentiability of Φ, even at $\bar{u}$. This, for example, allows us to cover in Section 8 reformulations of complementarity problems, making use of smooth complementarity functions.

Let Π be the orthogonal projector onto $(\operatorname{im}\Phi'(\bar{u}))^\perp$ in $\mathbb{R}^p$, and let $(\Phi')'(\bar{u};\, v)$ stand for the directional derivative of Φ' at $\bar{u}$ in the direction $v \in \mathbb{R}^p$.

Definition 2.1 The mapping Φ is said to be 2-regular at $\bar{u}$ in the direction v if the linear operator $B(v) : \ker\Phi'(\bar{u}) \to (\operatorname{im}\Phi'(\bar{u}))^\perp$ defined as the restriction of $\Pi(\Phi')'(\bar{u};\, v)$ to $\ker\Phi'(\bar{u})$ is nonsingular.

The 2-regularity property can be stated in various equivalent forms. For instance, it is equivalent to saying that $\Phi'(\bar{u}) + \Pi(\Phi')'(\bar{u};\, v)$ is a nonsingular matrix, or to the equality

$$\operatorname{im}\Phi'(\bar{u}) + (\Phi')'(\bar{u};\, v)\ker\Phi'(\bar{u}) = \mathbb{R}^p,$$

the latter form not involving the projector Π.

For mappings with directionally differentiable first derivatives, this notion of 2-regularity was introduced in [50, 51]; see also the discussion there and in [74]. Observe that $(\Phi')'(\bar{u};\, \cdot)$ is positively homogeneous. This implies that 2-regularity is indeed a directional property, i.e., it does not depend on the norm of v. Also, $(\Phi')'(\bar{u};\, \cdot)$ is (Lipschitz-)continuous, which implies that 2-regularity is stable subject to small perturbations of v. See [77], [4] for the origins of this concept in the twice differentiable case, and, e.g., [35], [36], [37], [73], [6], [7] as recent examples of its (and its extensions') use in optimization and variational analysis.

Evidently, if the solution $\bar{u}$ of (1.1) is nonsingular, Φ is 2-regular at $\bar{u}$ in every direction v, including $v = 0$. At the same time, simple examples (like with $p = 2$, $\Phi(u) = (u_1^2,\, u_1u_2)$, $\bar{u} = 0$) show that Φ may be 2-regular at $\bar{u}$ in nonzero directions at singular solutions as well, and even at nonisolated solutions, and even in directions $\bar{v} \in \ker\Phi'(\bar{u})$. The latter property will be referred to below as Key Assumption:

There exists $\bar{v} \in \ker\Phi'(\bar{u})$ such that the mapping Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. (2.1)

Furthermore, combining [48, Definition 1] with considerations in [24], we can say that a solution $\bar{u}$ of (1.1) is refereed to as noncritical if

$$\widehat{T}_{\Phi^{-1}(0)}(\bar{u}) = \ker \Phi'(\bar{u}), \tag{2.2}$$

and critical otherwise. Since it always holds that

$$\widehat{T}_{\Phi^{-1}(0)}(\bar{u}) \subset T_{\Phi^{-1}(0)}(\bar{u}) \subset \ker \Phi'(\bar{u}),$$

noncriticality condition (2.2) can actually be decomposed into the two ingredients, namely

$$\widehat{T}_{\Phi^{-1}(0)}(\bar{u}) = T_{\Phi^{-1}(0)}(\bar{u}), \tag{2.3}$$

which is the Clark-regularity property of the solution set $\Phi^{-1}(0)$ at $\bar{u}$ [10, Definition 2.4.6], and the equality

$$T_{\Phi^{-1}(0)}(\bar{u}) = \ker \Phi'(\bar{u}). \tag{2.4}$$

In fact, in [48, Definition 1] the noncriticality notion for equations was originally introduced precisely in this form.

According to [24], under the stated smoothness assumptions, noncriticaly of $\bar{u}$ is also equivalent to the equality

$$\ker \Phi'(\bar{u}) \cap N_{\Phi^{-1}(0)}(\bar{u}) = \{0\},$$

and according to [48, Theorem 2], it is further equivalent to the local Lipschitzian error bound property

$$\operatorname{dist}(u,\, \Phi^{-1}(0)) = O(\|\Phi(u)\|) \tag{2.5}$$

as $u \in \mathbb{R}^p$ tends to $\bar{u}$, and to the upper Lipschitzan property that consists of saying that for $w \in \mathbb{R}^p$, for any solution $u(w)$ of the perturbed equation

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

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

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

as $w \to 0$. We note in passing that since our smoothness assumptions imply continuous differentiability of Φ near $\bar{u}$, the result obtained in [5] says that the local Lipschitzian error bound property (2.5) (and hence, any of the equivalent properties specified above) implies that $\Phi^{-1}(0)$ is actually a smooth submanifold near $\bar{u}$, and in particular, is indeed automatically Clark regular at $\bar{u}$.

By [48, Theorem 3], every critical solution is necessarily singular, but the converse is not true in general. Finally, as observed in [48, p. 497] (in the twice differentiable case, but those consideration are easily extendable to our current setting), for a singular (e.g., nonisolated) solution $\bar{u}$, Key Assumption (2.1) above may only hold if $\bar{u}$ is critical: the equality (2.4) is necessarily violated in this case. This is the main reason why criticality of solutions appears in this survey, and even in its title.

We finally mention that criticality plays a crucial role for the stability of a solution, i.e., its potential to "survive" large classes of perturbations. These issues are left out of this survey focusing on Newton-type methods. For stability studies, we address the reader to [48] for twice differentiable unconstrained equations, to [2], [3], and [24] for constrained equations, and to [27] for the case when the second derivatives may not exist, and for the piecewise smooth case.

3 Perturbed Newton method framework

The following perturbed Newton method (pNM) framework for solving (1.1) was proposed and studied in [47], and later in [26]. For a given iterate $u^k \in \mathbb{R}^p$, the next iterate is defined as $u^{k+1} = u^k + v^k$, where v^k is a solution of the linear equation

$$\Phi(u^k) + (\Phi'(u^k) + \Omega(u^k))v = \omega(u^k), \tag{3.1}$$

with mappings $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ serving for characterizing various kinds of perturbations, corresponding to different methods and details of their implementation (e.g., inexactness in solving subproblems of a given method). Specific methods within the pNM framework are defined by the choice of these mappings. In particular, taking $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$ in (3.1) recovers the basic Newton method (NM) with the iteration system

$$\Phi(u^k) + \Phi'(u^k)v = 0. \tag{3.2}$$

Other examples of methods fitting the pNM framework, including those equipped with stabilizing mechanisms specially intended for the cases of singular and even nonisolated solutions, will be discussed in Sections 4 and 9 below.

The key ingredient of the local convergence analysis for the pNM framework is the following sharp characterization of its single step, obtained in [26, Lemma 3.1]. Its somehow less sharp predecessors for the twice differentiable case are [47, Lemma 1], [39, Lemma 4.1] and [32, Lemma 1] for the basic NM.

From this point on, we make use of the unique decomposition of every $u \in \mathbb{R}^p$ into the sum

$$u = u_1 + u_2 \quad \text{with } u_1 \in (\ker \Phi'(\bar{u}))^{\perp},\ u_2 \in \ker \Phi'(\bar{u}).$$

Lemma 3.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be differentiable near $\bar{u} \in \mathbb{R}^p$, and let the derivative of Φ be strongly semismooth at $\bar{u}$. Let $\bar{u}$ be a solution of the equation* (1.1)*, and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \mathbb{R}^p$. Let $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ satisfy the following properties: there exists $\delta > 0$ such that*

$$\Omega(u) = O(\|u - \bar{u}\|), \quad \omega(u) = O(\|u - \bar{u}\|^2) \tag{3.3}$$

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

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

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

Then there exist $\bar{\varepsilon} > 0$ *and* $\bar{\delta} > 0$ *such that, for every* $u \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u};\, \bar{v}) \setminus \{\bar{u}\}$, *the linear operator* $B(u - \bar{u})$ *(given in Definition 2.1) is invertible,*

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

as $u \to \bar{u}$, *the equation* (3.1) *with* $u^k = u$ *has the unique solution* v, *and this solution satisfies*

$$u_1 + v_1 - \bar{u}_1 = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|\|\Omega(u)\|) + O(\|\omega(u)\|) + O(\|u - \bar{u}\|^3),$$

$$\begin{aligned} u_2 + v_2 - \bar{u}_2 &= \frac{1}{2}(u_2 - \bar{u}_2 + (B(u - \bar{u}))^{-1}\Pi(\Phi')'(\bar{u};\, u - \bar{u})(u_1 - \bar{u}_1)) \\ &\quad + O(\|\Pi\Omega(u)\|) + O(\|u - \bar{u}\|^{-1}\|\Pi\omega(u)\|) + O(\|u - \bar{u}\|^2) \end{aligned} \tag{3.4}$$

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

As demonstrated by [26, Example 3.1], the conclusion of this lemma may fail in case of violation of *strong* semismoothness of Φ', even for the basic NM and twice continuously differentiable Φ.

The local convergence result for the pNM framework relying on Lemma 3.1 is coming next. It was established in this form in [26, Theorem 3.1], while in the twice differentiable case it was obtained earlier in [47, Theorem 1]. For the basic NM, this result is [74, Theorem 1] under our current smoothness requirements. Initially, under the twice differentiability, this result corresponds to [39, Lemma 5.1].

Theorem 3.1 *Let* $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ *be differentiable near* $\bar{u} \in \mathbb{R}^p$, *and let the derivative of* Φ *be strongly semismooth at* $\bar{u}$. *Let* $\bar{u}$ *be a solution of equation* (1.1), *and assume that* Φ *is 2-regular at* $\bar{u}$ *in a direction* $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$. *Moreover, let* $\Omega : \mathbb{R}^p \to \mathbb{R}^{p \times p}$ *and* $\omega : \mathbb{R}^p \to \mathbb{R}^p$ *possess the following properties: there exists* $\delta > 0$ *such that, along with* (3.3), *the estimates*

$$\Pi\Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \tag{3.5}$$

and

$$\Pi\omega(u) = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3) \tag{3.6}$$

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

Then, for every $\widehat{\varepsilon} > 0$ *and* $\widehat{\delta} > 0$, *there exist* $\varepsilon = \varepsilon(\bar{v}) > 0$ *and* $\delta = \delta(\bar{v}) > 0$ *such that for any starting point* $u^0 \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$, *there exists the unique sequence* $\{u^k\} \subset \mathbb{R}^p$ *such that for each* k *it holds that* $u^{k+1} = u^k + v^k$, *where* v^k *satisfies* (3.1), *and for this sequence and for each* k, *it holds that* $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u};\, \bar{v})$, $\{u^k\}$ *converges to* $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ *converges to zero monotonically,*

$$\frac{\|u_1^{k+1} - \bar{u}_1\|}{\|u_2^{k+1} - \bar{u}_2\|} = O(\|u^k - \bar{u}\|) \tag{3.7}$$

as $k \to \infty$, *and*

$$\lim_{k\to\infty} \frac{\|u_2^{k+1} - \bar{u}_2\|}{\|u_2^k - \bar{u}_2\|} = \frac{1}{2}. \tag{3.8}$$

As argued in [47, Remark 2], the estimates (3.7)–(3.8) in Theorem 3.1 imply that

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

meaning that $\{u^k\}$ converges to $\bar{u}$ linearly, with the asymptotic ratio exactly equal to 1/2.

Convergence in Theorem 3.1 is established from the staring points in $K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$ that is a convex set with nonempty interior, and hence, is "large" (not asymptotically thin). Moreover, convergence domain can be further enlarged by taking the union of $K_{\varepsilon(\bar{v}),\,\delta(\bar{v})}(\bar{u};\, \bar{v})$ over all $\bar{v} \in \ker \Phi'(\bar{u})$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. For the basic NM in the twice differentiable case, it was further shown in [39, Theorem 6.1] (see also [38] and [40, Theorem 2.1] for further related details and overview of the preceding works) that under Key Assumption (2.1), there exists a domain $U \subset \mathbb{R}^p$ which is starlike with respect to $\bar{u}$ and asymptotically dense at $\bar{u}$, and such that a single NM step from any $u^0 \in U$ is well-defined and produces the iterate $u^1 \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$ with $\bar{v} = \bar{v}(u^0) = \pi(u^0)/\|\pi(u^0)\|$ being such that Φ is necessarily 2-regular at $\bar{u}$ in the direction $\bar{v}$, where

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

(cf. the estimate (3.4) in Lemma 3.1), and with $\varepsilon > 0$ and $\delta > 0$ selected for this $\bar{v}$ according to Theorem 3.1. Then application of Theorem 3.1 yields convergence and rate of convergence estimates for all starting points in U.

Unfortunately, deriving a result of this kind for the pNM framework appears to require assumptions too restrictive on the perturbation terms. I.e., assumptions that do not naturally hold for the specific instances of the pNM framework, to be considered in Sections 4 and 9 below. Moreover, under our current smoothness assumptions, one cannot expect the domain of convergence to be asymptotically dense even for the basic NM, as demonstrated by [26, Examples 5.1–5.3], and also discussed earlier in [74, Section 4.2]. This also concerns the domain of ultimate acceptance of the full step in linesearch globalization techniques, the issue tackled in Theorem 5.1 below.

Nevertheless, we can conclude that critical solutions satisfying Key Assumption (2.1) are specially attractive for the pNM sequences under the specified requirements on perturbation terms: convergence to such solutions is guaranteed from "large" sets of starting points, even when critical solutions form a thin subset of the solution set $\Phi^{-1}(0)$, and the rate of convergence is linear with the asymptotic ratio exactly equal to 1/2.

4 Applications to the Levenberg–Marquardt and the LP-Newton methods

Consider the Levenberg–Marquardt (LM) method dating back to [67, 69], with the subproblem of the form

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u^k) + \Phi'(u^k)v\|^2 + \frac{1}{2}\sigma(u^k)\|v\|^2, \quad v \in \mathbb{R}^p, \tag{4.1}$$

where $\sigma : \mathbb{R}^p \to \mathbb{R}_+$ defines the regularization parameter. In case of nonsingular (hence isolated) solutions, local convergence properties of this method can be found, e.g., in [12, Theorem 10.2.6]. On the other hand, the LM method is a well-established tool for tackling nonisolated solutions. For an overview of modern convergence theories for this method, including local superlinear/quadratic convergence to some solution close to $\bar{u}$, under the local Lipschitzian error bound condition (2.5) (which is the same as noncriticality of $\bar{u}$, and of course allows for this solution to be nonisolated), and for related references, we address the reader to [33].

The LM method subproblem (4.1) is a convex optimization problem, and hence, via its optimality conditions, it is equivalent to the linear equation

$$(\Phi'(u^k))^\top \Phi(u^k) + ((\Phi'(u^k))^\top \Phi'(u^k) + \sigma(u^k)\mathcal{I})v = 0, \tag{4.2}$$

characterizing its stationary points. From from [39, Lemma 3.1] it follows that for $\bar{v} \in \mathbb{R}^p$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$, there exist $\bar{\varepsilon} > 0$ and $\bar{\delta} > 0$ such that, for every $u \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u};\, \bar{v}) \setminus \{\bar{u}\}$, $\Phi'(u)$ is invertible, and $(\Phi'(u))^{-1} = O(\|u - \bar{u}\|^{-1})$ as $u \to \bar{u}$. Multiplying both sides of (4.2) by $(((\Phi'(u^k))^\top)^{-1} = (((\Phi'(u^k))^{-1})^\top$, we then obtain the equation

$$\Phi(u^k) + (\Phi'(u^k) + \sigma(u^k)(((\Phi'(u^k))^{-1})^\top)v = 0,$$

which is the pNM iteration system (3.1) with the perturbation terms that are defined for $u \in K_{\bar{\varepsilon},\bar{\delta}}(\bar{u};\, \bar{v}) \setminus \{\bar{u}\}$ by

$$\Omega(u) = \sigma(u)(((\Phi'(u))^{-1})^\top = O(\sigma(u)\|u - \bar{u}\|^{-1}), \quad \omega(u) = 0. \tag{4.3}$$

The needed requirements (3.3), (3.5), and (3.6) on the perturbations terms do hold if, say, $\sigma(u) = \|\Phi(u)\|^\theta$ with $\theta \geq 2$, since

$$\Phi(u) = \Phi'(\bar{u})(u_1 - \bar{u}) + O(\|u - \bar{u}\|^2) \tag{4.4}$$

as $u \to \bar{u}$. Application of Theorem 3.1 now yields the following result, corresponding in the twice differentiable case to [47, Corollary 1].

Corollary 4.1 *Under the smoothness and 2-regularity assumptions in Theorem 3.1, its conclusion is valid with* (3.1) *replaced by the subproblem* (4.1) *of the LM method, where* $\sigma(u) = \|\Phi(u)\|^\theta$ *with* $\theta \geq 2$.

Observe that in the case of full singularity, i.e., when $\Phi'(\bar{u}) = 0$, the conclusions above are valid for all $\theta \geq 3/2$. Indeed, in this case, (4.3)–(4.4) imply that $\Omega(u) = O(\|u - \bar{u}\|^{2\theta-1})$, and hence, (3.5) holds for $\theta \geq 3/2$.

Another algorithm with similar local convergence properties near noncritical solutions is the LP-Newton (LPN) method introduced in [16] and further studied in [15] (see also [33] for an overview), with the iteration subproblem of the form

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \|\Phi(u^k) + \Phi'(u^k)v\| \leq \gamma\|\Phi(u^k)\|^2, \\ & \|v\| \leq \gamma\|\Phi(u^k)\|, \\ & (v,\, \gamma) \in \mathbb{R}^p \times \mathbb{R}. \end{array} \tag{4.5}$$

If the ∞-norm is used in (4.5), this is a linear programming problem, the reason for the name of the method.

The first constraint in the LPN method subproblem (4.5) can be seen as (3.1) with $\Omega(\cdot) \equiv 0$ and *some* $\omega(\cdot)$ that would satisfy (3.6) if the optimal value $\gamma(u)$ of the subproblem with $u^k = u$ satisfies

$$\gamma(u) = O(\|\Phi(u)\|^{-1}\|u - \bar{u}\|) \tag{4.6}$$

as $u \to \bar{u}$ (recall (4.4)). Since according to Lemma 3.1 applied with $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$, for every $u \in K_{\bar{\varepsilon},\,\bar{\delta}}(\bar{u};\,\bar{v}) \setminus \{\bar{u}\}$, the basic NM step $v(u)$ solving (3.2) with $u^k = u$ is uniquely defined, and $v(u) = O(\|u - \bar{u}\|)$, the pair $(v,\,\gamma) = (v(u),\,\|\Phi(u)\|^{-1}\|v(u)\|)$ is feasible in the LPN method subproblem (4.5). Hence,

$$\gamma(u) \leq \|\Phi(u)\|^{-1}\|v(u)\|,$$

yielding the needed estimate (4.6). This again allows to apply Theorem 3.1 in order to obtain the following version of [47, Corollary 2].

Corollary 4.2 *Under the smoothness and 2-regularity assumptions in Theorem 3.1, for every* $\widehat{\varepsilon} > 0$ *and* $\widehat{\delta} > 0$, *there exist* $\varepsilon = \varepsilon(\bar{v}) > 0$ *and* $\delta = \delta(\bar{v}) > 0$ *such that for any starting point* $u^0 \in K_{\varepsilon,\,\delta}(\bar{u};\,\bar{v})$, *there exists a sequence* $\{u^k\} \subset \mathbb{R}^p$ *such that for each* k, *it holds that* $u^{k+1} = u^k + v^k$, *where the pair* $(v^k,\,\gamma_{k+1})$ *with some* γ_{k+1} *solves* (4.5), *and for any such sequence and for each* k, *it holds that* $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\,\widehat{\delta}}(\bar{u};\,\bar{v})$, $\{u^k\}$ *converges to* $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ *converges to zero monotonically,* (3.7) *holds as* $k \to \infty$, *and* (3.8) *holds as well.*

Observe that unlike in Theorem 3.1 and Corollary 4.1, uniqueness of $\{u^k\}$ is not claimed in Corollary 4.1, as the LPN method subproblem (4.5) may have nonunique solutions and, hence, $\omega(u^k)$ is in general not uniquely defined.

Some illustrations of the behavior described in Corollaries 4.1 and 4.2 can be found in [47, Sections 3.1, 3.2]. In summary, in the presence of critical solutions, attraction to them still occurs for these Newton-type methods from large sets of starting points, despite the stabilizing mechanisms incorporated in these methods. And this attraction phenomenon may be the reason for the lack of superlinear convergence, as it may not allow these methods to enter a sufficiently small neighborhood of any noncritical solution, from where superlinear convergence does occur.

5 Globalization and acceleration issues

Convergence of the local algorithms discussed above can be naturally globalized by means of linesearch techniques for related merit functions fitting specific algorithms; see, e.g., [31], [32] for the basic NM, [79], [80], [65], [18], [34] for the LM method (see also [33] for an overview of these and other proposals), and [23] for the LPN method. Here we shall restrict ourselves to the following prototype algorithm for the general pNM framework.

Algorithm 5.1 Choose $u^0 \in \mathbb{R}^p$, $\varepsilon \in (0,\,1)$, $\theta \in (0,\,1)$, and set $k = 0$.

1. If $\Phi(u^k) = 0$, stop.

2. Compute $v^k \in \mathbb{R}^p$ as a solution of (3.1).

3. Set $\alpha = 1$. If the inequality

$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \varepsilon\alpha)\|\Phi(u^k)\| \tag{5.1}$$

is satisfied, set $\alpha_k = \alpha$. Otherwise, replace α by $\theta\alpha$, check the inequality (5.1) again, etc., until (5.1) becomes valid.

4. Set $u^{k+1} = u^k + \alpha_k v^k$.

5. Increase k by 1 and go to Step 1.

Step 2 of Algorithm 5.1 may fail or produce a direction "of poor quality". In order to have a well-defined algorithm, one may need to supply it with safeguards related to specific instances of the pNM framework. For instance, [31, Algorithm 3.1] suggests such remedies for the basic NM, and develops the corresponding global convergence results for the safeguarded algorithm. For the LM and LPN methods, we again address the reader to [33], and also to [26, Remark 4.1], for discussions of specific globalized algorithms. That said, Theorem 5.1 below is concerned with local properties, and is stated for the prototype Algorithm 5.1.

At this point, and in connection with globalization issues, we briefly discuss a technique for acceleration of convergence to critical solutions. The convergence pattern specified by estimates (3.7)–(3.8) in Theorem 3.1, and some further detail on it for the basic NM in [40, Theorem 2.1], serve as the basis for convergence acceleration developed in [38], [40], and later studied in [74]. One of those techniques is the so-called extrapolation, the simplest variant of which consists of generating an auxiliary sequence $\{\widehat{u}^k\}$ by doubling the (p)NM step: for each k, set

$$\widehat{u}^{k+1} = u^k + 2v^k. \tag{5.2}$$

As demonstrated in [40, Theorem 4.1], for the basic NM, one may expect $\{\widehat{u}^k\}$ to converge linearly with the asymptotic ratio of $1/4$, instead of $1/2$ for $\{u^k\}$. Since (5.2) does not affect the main iteration sequence $\{u^k\}$ in any way, this procedure can be easily combined with any implementations of the algorithms discussed above, and it does not entail any computational overhead except for one extra evaluation of Φ per iteration, needed to assess the quality of the obtained $\widehat{u}^{k+1}$. More sophisticated variants of extrapolation "of higher depth" [38], [40], further increase the convergence rate, yielding the asymptotic ratio $1/8$, $1/16$, etc.

There also exist some alternative acceleration techniques, such as overrelaxation [38], [40], [74], and Anderson acceleration of fixed-point iterations, originating from [1]. A very recent study related to the NM is [11]. To the best of our knowledge, globalization of convergence of these algorithms, preserving their acceleration properties, remains an open question, as well as their comparisons to each other.

Obviously, when combined with Algorithm 5.1, the issue of asymptotic acceptance of the full step at Step 3 of this algorithm becomes crucial for a potential success of the extrapolation technique, as the latter fully relies on the convergence pattern of the full-step pNM, and this pattern needs to be preserved by the globalized algorithm. The issue in question is highly nontrivial in cases of convergence to a singular solution, since unlike in the nonsingular case (for the latter see, e.g., [56, Theorem 5.4]), any neighborhood of a solution may contain points at which the NM step is well defined, but the full step is not accepted. Moreover, for a given $\bar{v} \in \ker \Phi'(\bar{u}) \setminus \{0\}$, such points may exist in $K_{\varepsilon,\delta}(\bar{u}; \bar{v})$ for any choices of $\varepsilon > 0$ and $\delta > 0$; this is demonstrated in [32, Examples 2, 3]. Nevertheless, the following result on the *ultimate* acceptance of the full step was obtained in [26, Theorem 4.1]; it generalizes [32, Proposition 3], both with respect to allowed perturbations of the basic NM and the relaxed smoothness assumptions.

Theorem 5.1 *Under the assumptions of Theorem 3.1, let the estimates* (3.5) *and* (3.6) *hold with removed* Π, *i.e.,*

$$\Omega(u) = O(\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^2) \tag{5.3}$$

and

$$\omega(u) = O(\|u - \bar{u}\|\|u_1 - \bar{u}_1\|) + O(\|u - \bar{u}\|^3) \tag{5.4}$$

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

Then, for every $\widehat{\varepsilon} > 0$ *and* $\widehat{\delta} > 0$, *there exist* $\varepsilon = \varepsilon(\bar{v}) > 0$ *and* $\delta = \delta(\bar{v}) > 0$ *such that for any starting point* $u^0 \in K_{\varepsilon,\delta}(\bar{u}; \bar{v})$, *Algorithm 5.1 with* $\varepsilon \in (0, 3/4)$ *uniquely defines the sequence* $\{u^k\}$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u}, \bar{v})$ *for all* k, *and* $\alpha_k = 1$ *holds for all* k *large enough.*

According to [32, Theorem 1] and the discussion following it, for the basic NM in the twice differentiable case, the set $K_{\varepsilon,\delta}(\bar{u}; \bar{v})$ in Theorem 5.1 can be extended to a set that is starlike with respect to $\bar{u}$ and asymptotically dense at $\bar{u}$. As commented at the end of Section 3, this extension is not possible under our current smoothness requirements, and for pNM without too restrictive assumptions on the perturbation terms.

6 Constrained equations

An important direction of further extensions of the constructions and results in Sections 3–5 is concerned with constrained equations of the form

$$\Phi(u) = 0, \quad u \in P, \tag{6.1}$$

where $P \subset \mathbb{R}^p$ is a given closed convex set. This is a very rich problem class with multiple applications, among which are various reformulations of complementarity systems; see Section 8. Additional constraints can be exogenous by nature, e.g., when solutions of the unconstrained equation make physical sense only if they satisfy these constraints, like nonnegativity restrictions on the components of u representing quantities. They can also be intrinsic ingredients of the problem setting, like, e.g., in smooth constraint reformulations of complementarity conditions; see Section 8.2 below. On the other hand, in some cases relevant constraints are imposed artificially, in order to ensure strong local convergence properties of Newton-type methods [16, 22], as well as for globalization of their convergence; see the discussion at the end of Section 7, and Section 8.2.

A natural generalization of the basic NM to the constrained setting is the constrained Gauss–Newton (GN) that instead of solving the linearized equation (3.2) (whose solution may result in an infeasible next iterate), computes the step v^k by minimizing the (squared) residual of the linearized equation (3.2) over $P - u^k$, thus producing $u^{k+1} = u^k + v^k \in P$. Therefore, the subproblem of the method has the form

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u^k) + \Phi'(u^k)v\|^2 \quad \text{subject to} \quad u^k + v \in P. \tag{6.2}$$

The regularized version of the subproblem (6.2) has the form

$$\text{minimize} \quad \frac{1}{2}\|\Phi(u^k) + \Phi'(u^k)v\|^2 + \frac{1}{2}\sigma(u^k)\|v\|^2 \quad \text{subject to} \quad u^k + v \in P, \tag{6.3}$$

with a function $\sigma : P \to \mathbb{R}_+$ defining the values of the regularization parameter, as in (4.1). The method with the subproblem (6.3) is the constrained Levenberg–Marquardt method; see [33] and references therein.

Finally, the version of the LPN method subproblem (4.5) can also incorporate the additional constraint:

$$\begin{array}{ll} \text{minimize} & \gamma \\ \text{subject to} & \|\Phi(u^k) + \Phi'(u^k)v\| \leq \gamma\|\Phi(u^k)\|^2, \\ & \|v\| \leq \gamma\|\Phi(u^k)\|, \\ & u^k + v \in P. \end{array} \tag{6.4}$$

In fact, the subpropblem of the LPN method was originally introduced in [16] precisely in the form (6.4).

The contemporary local superlinear/quadratic convergence results for the constrained LM and LPN methods under the local constrained Lipschitzian error bound condition

$$\operatorname{dist}(u,\, \Phi^{-1}(0) \cap P) = O(\|\Phi(u)\|) \text{ as } u \in P \text{ tends to } \bar{u}, \tag{6.5}$$

and related references, are discussed in [33]. Here, we are concerned with the case when (6.5) need not to hold.

As demonstrated in [30, Section 3] (see also [26, Remark 3.2]), if the assumptions of Theorem 3.1 are complemented with the additional requirement that $\bar{v}$ belongs to the interior of the tangent cone to P at $\bar{u}$, the iterates u^k in that theorem can be additionally claimed to stay feasible, i.e., to belong to P for all k. In other words, being initialized within $K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$ with appropriate $\varepsilon > 0$ and $\delta > 0$, minimizing the residual of the pNM subproblem (3.1) over $P - u^k$ will produce exactly the same iterate $u^{k+1} = u^k + v^k$ as solving (3.1) itself without any additional constraints. The following result extends [30, Theorem 3.1] to the case of our current smoothness requirements.

Theorem 6.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be differentiable near $\bar{u} \in \mathbb{R}^p$, and let the derivative of Φ be strongly semismooth at $\bar{u}$. Let $\bar{u}$ be a solution of* (6.1)*, and assume that Φ is 2-regular at $\bar{u}$ in a direction $\bar{v} \in \ker \Phi'(\bar{u}) \cap \operatorname{int} T_P(\bar{u})$. Moreover, let $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ possess the following properties: there exists $\delta > 0$ such that, along with* (3.3)*, the estimates* (3.5) *and* (3.6) *hold for $u \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v}) \cap P$ as $\varepsilon \to 0+$.*

Then, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$, there exists the unique sequence $\{u^k\} \subset \mathbb{R}^p$ such that for *each k it holds that $u^{k+1} = u^k + v^k$, where v^k satisfies* (3.1), *and for this sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\widehat{\delta}}(\bar{u};\, \bar{v}) \cap P$, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically,* (3.7) *holds as $k \to \infty$, and* (3.8) *holds as well.*

Theorem 6.1 readily covers the constrained GN method (by taking $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$), and allows to obtain generalizations of Corollaries 4.1 and 4.2 for the constrained LM and LPN methods respectively, employing the reasoning from Section 4. This development is rather straightforward, and we skip the details for brevity.

We complete this section with a discussion of the case when the direction $\bar{v}$ is not in the interior of the tangent cone to P at $\bar{u}$. At any singular solution $\bar{u}$ of the equation in (6.1), it holds that $\ker \Phi'(\bar{u}) \neq \{0\}$, and the behavior of Newton type methods near $\bar{u}$ is much defined by the relative position of $\ker \Phi'(\bar{u})$ and $T_P(\bar{u})$, with respect to each other.

One principal case is when $\bar{u} + \ker \Phi'(\bar{u})$ intersects P "transversally" at $\bar{u}$, i.e.,

$$\ker \Phi'(\bar{u}) \cap T_P(\bar{u}) = \{0\},$$

(the term "transversally" is used here in a somehow loose meaning, as the opposite to "tangentially"). In this case, $\bar{u}$ is necessarily an isolated solution of (6.1), and the local constrained Lipshitzian error bound condition (6.5) is valid, now reducing to the form

$$\|u - \bar{u}\| = O(\|\Phi(u)\|) \;\text{ as } u \in P \text{ tends to } \bar{u}.$$

As discussed above, this implies local superlinear/quadratic convergence of the constrained LM and LPN methods, in this case to $\bar{u}$ itself, but this is not the case of interest in this survey.

Theorem 6.1 above deals with another extreme case when

$$\ker \Phi'(\bar{u}) \cap \operatorname{int} T_P(\bar{u}) \neq \emptyset. \tag{6.6}$$

The intermediate case when (6.6) may be violated, but

$$\ker \Phi'(\bar{u}) \cap T_P(\bar{u}) \neq \{0\},$$

was partially addressed in [59], assuming that there exists $\bar{v} \in \ker \Phi'(\bar{u}) \cap T_P(\bar{u})$ such that Φ is 2-regular at $\bar{u}$ in the direction $\bar{v}$. As discussed in [3, p. 624], such $\bar{v}$ may only exist when the constrained error bound (6.5) is violated, actually no matter whether $\bar{v} \in \operatorname{int} T_P(\bar{u})$ or not.

Clearly, in this intermediate case, one cannot expect the constrained Newton-type methods initialized within $K_{\varepsilon,\delta}(\bar{u};\, \bar{v})$ with arbitrarily small $\varepsilon > 0$ and $\delta > 0$ to work as their unconstrained versions. The idea adopted in [59] is to interpret, however, the constrained GN method as an (unconstrained) pNM, with the appropriate estimates on the perturbation terms, and to obtain the local convergence and rate of convergence result by applying Theorem 3.1 to this instance of pNM. This is done in [59, Theorem 3.1] under twice differentiability, and most importantly, under some additional requirements on P and $\bar{v}$, including the assumption that $\bar{v}$ is a feasible direction for P at $\bar{u}$. With this result for the constrained GN method at hand, the analysis is further extended in [59, Section 4] to the constrained LM and LPN methods. Moreover, [59, Example 3.1] demonstrates that without those additional additional requirements on P and $\bar{v}$, this line of analysis would not be possible, while [59, Example 3.2] demonstrates the need to assume Lipschitz-continuity with respect to $\bar{u}$ for the second derivative of Φ, even for simplest polyhedral P. (Actually, the mapping Φ in this example is the same as in [26, Example 3.1] that was cited above as an evidence for the need of strong semismoothness of the derivative in Theorem 3.1.)

7 Piecewise smooth equations

Another important direction of extensions is the case when the mapping Φ in (1.1) (or in (6.1), though we do not consider this here) is not necessarily differentiable (even once) but only piecewise smooth near a solution $\bar{u}$ in question. By this we mean that Φ is continuous near $\bar{u}$, and there exists a finite collection of smooth (say, continuously differentiable near $\bar{u}$) selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^p$, such that

$$\Phi(u) \in \{\Phi^1(u), \ldots, \Phi^s(u)\} \quad \forall\, u \in \mathbb{R}^p.$$

According to [41, Theorem 2.1], any mapping that is piecewise smooth near $\bar{u}$ is Lipschitz-continuous near $\bar{u}$. For natural reformulations of complementarity systems that can be cast as piecewise smooth equations, see Section 8.

For each $u \in \mathbb{R}^p$, define the set

$$\mathcal{A}(u) = \{j \in \{1, \ldots, s\} \mid \Phi(u) = \Phi^j(u)\} \tag{7.1}$$

of indices of smooth selection mappings active at u. Then the set-valued mapping $\mathcal{A}(\cdot)$ is evidently outer semicontinuous, and in particular, $\mathcal{A}(u) \subset \mathcal{A}(\bar{u})$ holds for any $\bar{u} \in \mathbb{R}^p$ and all $u \in \mathbb{R}^p$ close enough to $\bar{u}$.

Furthermore, according to [17, Lemma 4.6.1] we also have that Φ is directionally differentiable at $\bar{u}$ in any direction $v \in \mathbb{R}^p$, with the directional derivative

$$\Phi'(\bar{u};\, v) \in \{(\Phi^j)'(\bar{u})v \mid j \in \mathcal{A}(\bar{u})\} \quad \forall\, v \in \mathbb{R}^p. \tag{7.2}$$

"Piecewise Newton-type methods" is a general name for a class of algorithms with iteration at a current $u^k \in \mathbb{R}^p$ being the iteration of a corresponding Newton-type method for the smooth equation

$$\Phi^j(u) = 0 \tag{7.3}$$

with *some* $j \in \mathcal{A}(u^k)$. Taking $s = 1$ recovers the case when Φ is continuously differentiable near $\bar{u}$. More generally, if $\mathcal{A}(\bar{u})$ is a singleton $\{\widehat{\jmath}\}$, then by the outer semicontinuity of $\mathcal{A}(\cdot)$, any piecewise Newton-type method reduces locally to the the corresponding method for the smooth equation (7.3) with $j = \widehat{\jmath}$, and in this case, the results from preceding sections can be readily extended to the piecewise smooth setting. However, if $\mathcal{A}(\bar{u})$ is not a singleton, index j used in (7.3) may vary from one iteration to another, no matter how close the iterates are to $\bar{u}$, and a piecewise Newton-type method cannot be interpreted as a Newton-type method for a single smooth equation.

In particular, the piecewise NM, the piecewise LM method, and the piecewise LPN method have the subproblems of the form (3.2), (4.1), and (4.5), respectively, but with $\Phi'(u^k)$ in these subproblems replaced by $(\Phi^j)'(u^k)$ with some $j \in \mathcal{A}(u^k)$.

For every $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^p$, we further define the index set

$$\mathcal{A}(u;\, v) = \{j \in \mathcal{A}(u) \mid v \in \ker(\Phi^j)'(u)\}.$$

The key observation in this analysis is the following

Lemma 7.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be piecewise smooth near $\bar{u} \in \mathbb{R}^p$, with smooth selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^p$. Let $\bar{u}$ be a solution of* (1.1)*, and let $\bar{v} \in \mathbb{R}^p$ be such that*

$$\Phi'(\bar{u};\, \bar{v}) = 0. \tag{7.4}$$

Then there exist $\varepsilon > 0$ and $\delta > 0$ such that

$$\mathcal{A}(u) \subset \mathcal{A}(\bar{u};\, \bar{v}) \quad \forall\, u \in K_{\varepsilon,\,\delta}(\bar{u};\, \bar{v}).$$

In particular, if $\mathcal{A}(\bar{u};\, \bar{v})$ consist of a single index $\widehat{\jmath}$, then there exist $\varepsilon > 0$ and $\delta > 0$ such that $\mathcal{A}(u) = \{\widehat{\jmath}\}$ for all $u \in K_{\varepsilon,\,\delta}(\bar{u};\, \bar{v})$.

The local convergence result for the perturbed piecewise NM method is obtained by showing, through Lemma 7.1, that this method fits the pNM framework of Theorem 3.1 applied to the smooth equations (7.3) corresponding to an appropriate collection of active smooth selections. The original version of this result in [25, Theorem 1] requires twice differentiability of selection mappings.

We will be considering $\bar{v} \in \mathbb{R}^p$ such that $\mathcal{N} = \ker(\Phi^j)'(\bar{u})$ is the same for all $j \in \mathcal{A}(\bar{u},\, \bar{v})$, and we will now use the decomposition of every $u \in \mathbb{R}^p$ into the sum $u = u_1 + u_2$, with $u_1 \in \mathcal{N}^{\perp}$ and $u_2 \in \mathcal{N}$. The assumption that the null spaces of $(\Phi^j)'(\bar{u})$ coincide for all $j \in \mathcal{A}(\bar{u},\, \bar{v})$ may seem restrictive, but in fact it is reasonable, in the following sense. Suppose that for a given $\bar{v}$, there exist $j_1,\, j_2 \in \mathcal{A}(\bar{u},\, \bar{v})$ such that $\ker(\Phi^{j_1})'(\bar{u}) \neq \ker(\Phi^{j_2})'(\bar{u})$. This means that there exists $\widehat{v} \in \mathbb{R}^p$ such that it belongs, say, to the first of these null spaces but not to the second. Then, for any real t close enough to 0, it holds that $\bar{v} + t\widehat{v} \in \ker(\Phi^{j_1})'(\bar{u})$, $\bar{v} + t\widehat{v} \notin \ker(\Phi^{j_2})'(\bar{u})$, while the evident outer semicontinuity of $\mathcal{A}(\bar{u},\, \cdot)$ implies that $\mathcal{A}(\bar{u},\, \bar{v} + t\widehat{v}) \subset \mathcal{A}(\bar{u},\, \bar{v})$. Inclusion in (7.2) then implies that $\mathcal{A}(\bar{u},\, \bar{v} + t\widehat{v})$ cannot contain both indices j_1 and j_2 simultaneously. Repeating this procedure with $\bar{v}$ replaced by $\bar{v} + t\widehat{v}$, we eventually end up with $\bar{v}$ such that either $\mathcal{A}(\bar{u},\, \bar{v})$ is a singleton, or the null spaces of $(\Phi^j)'(\bar{u})$ coincide for all $j \in \mathcal{A}(\bar{u},\, \bar{v})$. Moreover, this $\bar{v}$ can be taken arbitrarily close to the original one, and therefore, the 2-regularity properties in the original direction will be preserved for this $\bar{v}$, since 2-regularity is stable with respect to small perturbations of the direction.

Theorem 7.1 *Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be piecewise smooth near $\bar{u} \in \mathbb{R}^p$, with smooth selection mappings $\Phi^1, \ldots, \Phi^s : \mathbb{R}^p \to \mathbb{R}^p$. Let $\bar{u}$ be a solution of* (1.1)*, and let $\bar{v} \in \mathbb{R}^p \setminus \{0\}$ be such that* (7.4) *holds, and for every $j \in \mathcal{A}(\bar{u};\, \bar{v})$, the derivative of Φ^j is strongly semismooth at $\bar{u}$, $\ker(\Phi^j)'(\bar{u}) = \mathcal{N}$, where the linear subspace $\mathcal{N}$ does not depend on j, and Φ^j is 2-regular at $\bar{u}$*

in the direction $\bar{v}$. Moreover, let $\Omega : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ and $\omega : \mathbb{R}^p \to \mathbb{R}^p$ satisfy the estimates (3.3) *and* (5.3)–(5.4) *as $u \to \bar{u}$. Assume finally that $G : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ is a fixed mapping satisfying*

$$G(u) \in \{(\Phi^j)'(u) \mid j \in \mathcal{A}(u)\} \quad \forall\, u \in \mathbb{R}^p. \tag{7.5}$$

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

$$\Phi(u^k) + (G(u^k) + \Omega(u^k))(u - u^k) = \omega(u^k), \tag{7.6}$$

and for this sequence and for each k, it holds that $u_2^k \neq \bar{u}_2$, $u^k \in K_{\widehat{\varepsilon},\,\widehat{\delta}}(\bar{u};\, \bar{v}) \cap P$, $\{u^k\}$ converges to $\bar{u}$, $\{\|u^k - \bar{u}\|\}$ converges to zero monotonically, (3.7) *holds as $k \to \infty$, and* (3.8) *holds as well.*

Theorem 7.1 applied with $\Omega(\cdot) \equiv 0$ and $\omega(\cdot) \equiv 0$ covers the piecewise NM, and moreover, employing the reasoning from Section 4, it also allows to obtain generalizations of Corollaries 4.1 and 4.2 for the piecewise LM and LPN methods respectively.

As for asymptotic acceptance of the full step by the piecewise counterpart of Algorithm 5.1, this was established in [25, Theorem 2] in the case of twice differentiability of selection mappings. We now present this result as a generalization of Theorem 5.1. Yet again, the key role is played by Lemma 7.1 allowing to show that the algorithm fits Theorem 5.1 applied to proper active smooth selections.

Theorem 7.2 *Under the assumptions of Theorem 7.1, for every $\widehat{\varepsilon} > 0$ and $\widehat{\delta} > 0$, there exist $\varepsilon = \varepsilon(\bar{v}) > 0$ and $\delta = \delta(\bar{v}) > 0$ such that for any starting point $u^0 \in K_{\varepsilon,\,\delta}(\bar{u};\, \bar{v})$, Algorithm 5.1 with $\varepsilon \in (0,\, 3/4)$, and with* (3.1) *on Step 2 replaced by* (7.6)*, uniquely defines the sequence $\{u^k\}$, $u^k \in K_{\widehat{\varepsilon},\,\widehat{\delta}}(\bar{u},\, \bar{v})$ for all k, and $\alpha_k = 1$ holds for all k large enough.*

We now briefly comment on the possibility to adapt the piecewise Newton-type methods of this section to the constrained setting (6.1), and we start with the following observation highlighting the special role of additional constraints in the piecewise smooth setting. Even though Theorem 7.2 is valid and characterizes an important local feature of a piecewise version of Algorithm 5.1, the linesearch procedure in this algorithm actually does not make much sense in the piecewise context without further assumptions, as the direction v^k of, say, the piecewise NM, does not need to be a direction of descent for $\|\Phi(\cdot)\|$ at u^k. However, the situation changes under the additional restriction on the character of piecewise smoothness that has already appeared before, e.g., in [23, (4.8)]. Specifically, this happens if $P \subset \mathbb{R}^p$ is chosen in such a way that

$$\|\Phi(u)\| \leq \|\Phi^j(u)\| \quad \forall\, j \in \{1,\, \ldots,\, s\},\ \forall\, u \in P, \tag{7.7}$$

and if the iterates of the method are forced to stay in such P. Indeed, if v^k is the NM direction for the equation (7.3) for some $j \in \mathcal{A}(u^k)$, and assuming that $\|\Phi^j(u^k)\| = \|\Phi(u^k)\| \neq 0$, one obtains in a standard way that v^k is a direction of descent for $\|\Phi^j(\cdot)\|$ at u^k, where this function is differentiable with its gradient at u^k equal to $((\Phi^j)'(u^k))^\top \Phi^j(u^k)/\|\Phi^j(u^k)\|$. Assuming now that $u^k \in P$, and v^k is a feasible direction for P at u^k (which is of course not automatic, and has to be ensured by appropriate modifications of Algorithm 5.1, like those employing the constrained Newton-type methods in Section 6), from (7.7) we have that

$$\|\Phi(u^k + \alpha v^k)\| \leq \|\Phi^j(u^k + \alpha v^k)\| < \|\Phi^j(u^k)\| = \|\Phi(u^k)\|$$

for all $\alpha > 0$ small enough.

The local convergence results for the constrained piecewise Newton-type methods, in the spirit of Theorem 6.1, can be obtained by using Lemma 3.1 and Theorems 3.1 and 7.1. See [25, Section 3] for the details of this reasoning, and in particular, [25, Theorem 3]. One subtle point is that the requirement $\bar{v} \in \operatorname{int} T_P(\bar{u})$ in Theorem 6.1 appears too restrictive for the choices of P relevant in natural constrained piecewise smooth reformulations of complementarity problems, in the absence of strict complementarity; see the next section for a discussion of such reformulations. The trick proposed in [25, Section 3] to resolve this difficulty is to split the constraints into two parts, with the interiority requirement on $\bar{v}$ assumed to hold only for one part of constrains, while the other constraints are automatically satisfied by the iteration of the piecewise NM itself.

8 Applications to complementarity problems

We shall restrict the following exposition to the nonlinear complementarity problem (NCP)

$$x \geq 0, \quad F(x) \geq 0, \quad \langle x,\, F(x)\rangle = 0, \tag{8.1}$$

with a smooth mapping $F : \mathbb{R}^n \to \mathbb{R}^n$. The material can certainly be extended to more general problem settings, like the mixed complementarity problem (MCP) [17, Definition 1.1.6], [56, (1.3.7)], including the Karush–Kuhn–Tucker systems for optimization and variational problems [17, Section 1.3.2], [56, (1.3.9)].

8.1 Unconstrained reformulation of complementarity

Employing a complementarity function $\psi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, i.e., any function satisfying $\psi^{-1}(0) = \mathbb{R}_+ \times \mathbb{R}_+$ [17, Definition 1.5.1], [56, Section 3.2.1], and setting $u = x$ (and $p = n$), problem (8.1) can be equivalently reformulated as the equation (1.1) with the mapping Φ defined by

$$\Phi(u) = \psi(u,\, F(u)), \tag{8.2}$$

where the complementarity function is applied componentwise. Different complementarity functions lead to equations with different smoothness and regularity properties, and as a consequence, to different methods for solving complementarity problems.

One immediate and widely used choice of a complementarity function is the so-called natural residual function

$$\psi(a,\, b) = \min\{a,\, b\}. \tag{8.3}$$

With this choice, Φ defined in (8.2) is piecewise smooth. Indeed, set $s = 2^n$, and fix any one-to-one mapping $j \mapsto I(j)$ from $\{1,\, \ldots,\, s\}$ to the set of all different subsets of $\{1,\, \ldots,\, n\}$

(including $\emptyset$ and the entire $\{1, \ldots, n\}$). Then the corresponding smooth selection mappings $\Phi^j : \mathbb{R}^n \to \mathbb{R}^n$ have the components

$$\Phi_i^j(u) := \begin{cases} u_i, & \text{if } i \in I(j), \\ F_i(u) & \text{otherwise,} \end{cases} \quad i = 1, \ldots, n, \ j = 1, \ldots, s.$$

Deciphering the assumptions of Theorem 7.1 for the specified piecewise smooth reformulation of the NCP, and examples demonstrating the corresponding behavior of the piecewise Newton-type methods from Section 7 for this reformulation, can be found in [25, Section 4.1]. In particular, for a given $u \in \mathbb{R}^n$, the set of indices of active selection mappings defined according to (7.1) takes the form

$$\mathcal{A}(u) = \{j \in \{1, \ldots, s\} \mid I(j) = J \cup I_<(u), \ J \subset I_=(u)\}, \tag{8.4}$$

with

$$\begin{array}{rcl} I_>(u) & = & \{i \in \{1, \ldots, n\} \mid u_i > F_i(u)\}, \\ I_=(u) & = & \{i \in \{1, \ldots, n\} \mid u_i = F_i(u)\}, \\ I_<(u) & = & \{i \in \{1, \ldots, n\} \mid u_i < F_i(u)\} \end{array} \tag{8.5}$$

forming the natural partitioning of the index set $\{1, \ldots, n\}$. Hence, the requirement (7.5) on the choice of a mapping $G : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ in Theorem 7.1 can be written for the rows $G_i(u)$ of $G(u)$ as follows:

$$\text{for some } J \subset I_=(u), \quad G_i(u) = \begin{cases} e^i, & \text{if } i \in J \cup I_<(u), \\ F_i'(u) & \text{otherwise,} \end{cases} \quad i = 1, \ldots, n,$$

where $e^1, \ldots, e^n$ is the canonical basis of $\mathbb{R}^n$. We also mention that this analysis employs [56, Proposition 3.21] deciphering condition (7.4) in this context.

We note that there also exist smooth complementarity functions, e.g.,

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

originally introduced in [14]. See [66] and [64] for other examples of this kind. If F is differentiable at $u \in \mathbb{R}^n$, the corresponding mapping defined in (8.2) is differentiable at u as well, and $\Phi'(u)$ has the rows

$$\Phi_i'(u) = 2u_i F_i'(u) + 2F_i(u)e^i - 2\min\{0, u_i + F_i(u)\}(F_i'(u) + e^i), \quad i = 1, \ldots, n. \tag{8.6}$$

Moreover, assuming that F' is strongly semismooth at $\bar{u}$ (which is automatic if F is twice differentiable near $\bar{u}$, and its second derivative is Lipschitz-continuous near $\bar{u}$), from [56, Proposition 1.75] it follows that Φ' is strongly semismooth at $\bar{u}$.

At a solution $\bar{u} = \bar{x}$ of the NCP (8.1), the index sets in (8.5) take the form

$$\begin{array}{rcl} I_>(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{u}_i > 0 = F_i(\bar{u})\}, \\ I_=(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{u}_i = 0 = F_i(\bar{u})\}, \\ I_<(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{u}_i = 0 < F_i(\bar{u})\}, \end{array}$$

and (8.6) then yields

$$\Phi_i'(\bar{u}) = \begin{cases} 2\bar{u}_i F_i'(\bar{u}) & \text{if } i \in I_>(\bar{u}), \\ 0 & \text{if } i \in I_=(\bar{u}), \\ 2F_i(\bar{u})e^i & \text{if } i \in I_<(\bar{u}). \end{cases} \tag{8.7}$$

Therefore, if the strict complementarity condition is violated at $\bar{u}$, i.e., $I_=(\bar{u}) \neq \emptyset$, then $\bar{u}$ is necessarily a singular solution of equation (1.1) with the specified Φ. Moreover, this is also true for any other choice of a smooth complementarity function, and hence, such reformulations of the NCP serve as a natural source of applications of the results presented in Sections 3–5.

We next discuss the assumptions of Theorem 3.1 in the current context, implying that the conclusions of that theorem, of Corollaries 4.1 and 4.2, and of Theorem 5.1, are valid for the considered smooth reformulation of the NCP. From (8.6) it can be readily derived that for every $v \in \mathbb{R}^p$ and $i \in I_=(\bar{u})$,

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

Taking into account (8.7), we then conclude that our Key Assumption (2.1) of 2-regularity of Φ at $\bar{u}$ in some direction $\bar{v} \in \ker \Phi'(\bar{u})$ automatically holds with any $\bar{v} \in \mathbb{R}^n$ such that

$$\langle F_i'(\bar{u}),\, \bar{v}\rangle = 0,\ i \in I_>(\bar{u}), \quad \bar{v}_i = 0,\ i \in I_<(\bar{u}),$$

meaning precisely that $\bar{v} \in \ker \Phi'(\bar{u})$, and the matrix with the rows

$$\begin{gathered} F_i'(\bar{u}),\ i \in I_>(\bar{u}), \\ \max\{\bar{v}_i,\, -\langle F_i'(\bar{u}),\, \bar{v}\rangle\}F_i'(\bar{u}) - \min\{\bar{v}_i,\, -\langle F_i'(\bar{u}),\, \bar{v}\rangle\})e^i,\ i \in I_=(\bar{u}), \\ e^i,\ i \in I_<(\bar{u}), \end{gathered}$$

is nonsingular, which is a sufficient condition for 2-regularity in the direction $\bar{v}$. The latter evidently implies that

$$F_i'(\bar{u}),\ i \in I_>(\bar{u}), \quad e^i,\ i \in I_<(\bar{u}), \quad \text{are linearly independent,} \tag{8.8}$$

and moreover, this sufficient condition for 2-regularity also becomes necessary under (8.8). Condition (8.8) means that singularity of a solution $\bar{u}$ is imposed in a natural way, i.e., solely by violation of strict complementarity. An exact characterization of 2-regularity for the specified Φ, not relying on (8.8), can be found in [74].

For examples demonstrating the behavior of the NM and the LM method from Section 7 near singular solutions of the considered smooth reformulation of the NCP, as well as for some numerical testing, we address the reader to [26, Section 5], [63].

8.2 Constrained reformulation of complementarity

Reasonable constrained reformulations of NCP (8.1) require introducing slack variable $y \in \mathbb{R}^n$, which leads to the MCP

$$F(x) - y = 0, \quad x \geq 0, \quad y \geq 0, \quad \langle x, y \rangle = 0. \tag{8.9}$$

Evidently, $\bar{x}$ is a solution of the NCP (8.1) if and only if $(\bar{x}, F(\bar{x}))$ is a solution of the MCP (8.9). With this in mind, setting $u = (x, y)$ (and $p = 2n$), we reformulate the NCP (8.1) as (6.1) with Φ defined by

$$\Phi(u) = (F(x) - y, \psi(x, y)) \tag{8.10}$$

and with

$$P = \mathbb{R}^n_+ \times \mathbb{R}^n_+, \tag{8.11}$$

where ψ is a complementarity function.

With ψ taken according to (8.3), the mapping Φ in (8.10) is piecewise smooth, with the corresponding smooth selection mappings $\Phi^j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$,

$$\Phi^j(u) = (F(x) - y, \Psi^j(u)), \quad j = 1, \ldots, s, \tag{8.12}$$

where the mappings $\Psi^j : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ have the components

$$\Psi^j_i(u) := \begin{cases} x_i & \text{if } i \in I(j), \\ y_i & \text{otherwise}, \end{cases} \quad i = 1, \ldots, n, \; j = 1, \ldots, s, \tag{8.13}$$

with $I(j)$ defined as in Section 8.1.

One important observation is in order. Since for any complementarity function ψ, the equality $\psi(x, y) = 0$ implies that $x \geq 0$ and $y \geq 0$, introducing the constraint $u \in P$ with P defined in (8.11) in the reformulation in question may seem redundant, and it is indeed unnecessary if only the equivalence of reformulations is of concern. However, say, for ψ defined according to (8.3), condition (7.7) is satisfied for the specified Φ and P, and for the smooth selections of Φ defined in (8.12)–(8.13). At the same time, simple examples show that (7.7) may not hold if one takes $P = \mathbb{R}^n \times \mathbb{R}^n$, i.e., in the unconstrained case. The importance of (7.7) for globalization of convergence was discussed in Section 7, and this observation highlights one of the roles played by the constraints in the piecewise smooth reformulations of complementarity.

For a given $u \in \mathbb{R}^n \times \mathbb{R}^n$, the set of indices of active selection mappings has the same form (8.4) as above, but now with

$$\begin{aligned} I_>(u) &= \{i \in \{1, \ldots, n\} \mid x_i > y_i\}, \\ I_=(u) &= \{i \in \{1, \ldots, n\} \mid x_i = y_i\}, \\ I_<(u) &= \{i \in \{1, \ldots, n\} \mid x_i < y_i\}. \end{aligned} \tag{8.14}$$

Hence, a mapping $G : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{2n \times 2n}$ satisfying (7.5) is defined by

$$G(u) = \begin{pmatrix} F'(x) & -\mathcal{I} \\ \Gamma^1(u) & \Gamma^2(u) \end{pmatrix}, \tag{8.15}$$

where the rows of $\Gamma^1 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ and $\Gamma^2 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n\times n}$ are such that

$$\text{for some } J \subset I_=(u),$$
$$\Gamma_i^1(u) = \begin{cases} e^i & \text{if } i \in J \cup I_<(u), \\ 0 & \text{otherwise,} \end{cases} \quad \Gamma_i^2(u) = \begin{cases} 0 & \text{if } i \in J \cup I_<(u), \\ e^i & \text{otherwise,} \end{cases} \quad i = 1, \ldots, n. \tag{8.16}$$

Further deciphering the assumptions of [25, Theorem 3] in the current context, as well as applications of that theorem combined with the results for the unconstrained case to the constrained piecewise Newton-type methods, are discussed in [25, Section 4.2].

Observe finally that the MCP (8.9) can be equivalently reformulated as the constrained equation (6.1) with P defined in (8.11) without making use of any complementarity function at all, e.g., by taking

$$\Phi(u) = (F(x) - y,\, x \circ y), \tag{8.17}$$

where $x \circ y = (x_1y_1, \ldots, x_ny_n)$ is the Hadamard product of two vectors. Evidently, the constraint $u \in P$ is essential for such reformulation to be equivalent to the MCP (8.9).

At a solution $\bar{x}$ of the NCP (8.1), and for $\bar{u} = (\bar{x},\, F(\bar{x}))$, the index sets in (8.14) take the form

$$\begin{array}{rcccl} I_> & = & I_>(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{x}_i > 0 = F_i(\bar{x})\}, \\ I_= & = & I_=(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{x}_i = 0 = F_i(\bar{x})\}, \\ I_< & = & I_<(\bar{u}) & = & \{i \in \{1, \ldots, n\} \mid \bar{x}_i = 0 < F_i(\bar{x})\}. \end{array}$$

From (8.17) it follows that

$$\Phi'(\bar{u}) = \begin{pmatrix} F'(\bar{x}) & -\mathcal{I} \\ \operatorname{diag} F(\bar{x}) & \operatorname{diag} \bar{x} \end{pmatrix},$$

and hence,

$$\ker \Phi'(\bar{u}) = \left\{ v = (\xi,\, \eta) \;\middle|\; \begin{array}{c} \dfrac{\partial F_{I_>}}{\partial x_{I_> \cup I_=}}(\bar{x})\xi_{I_> \cup I_=} = 0,\; \dfrac{\partial F_{I_= \cup I_<}}{\partial x_{I_> \cup I_=}}(\bar{x})\xi_{I_> \cup I_=} = \eta_{I_= \cup I_<}, \\ \xi_{I_<} = 0,\; \eta_{I_>} = 0 \end{array} \right\}. \tag{8.18}$$

This evidently implies that $\bar{u}$ is a singular solution of the equation in (6.1) if and only if $I_= \neq \emptyset$, or $I_> \neq \emptyset$ and $\frac{\partial F_{I_>}}{\partial x_{I_>}}(\bar{x})$ is a singular matrix.

As demonstrated in [30, Section 4.2], for any $\bar{v} = (\bar{\xi},\, \bar{\eta}) \in \ker \Phi'(\bar{u})$, 2-regularity of Φ at $\bar{u}$ in the direction $\bar{v}$ means that there exists no nonzero $v = (\xi,\, \eta) \in \ker \Phi'(\bar{u})$ satisfying

$$\frac{\partial^2 F_{I_>}}{\partial x^2_{I_> \cup I_=}}(\bar{x})[\bar{\xi}_{I_> \cup I_=},\, \xi_{I_> \cup I_=}] \in \operatorname{im} \frac{\partial F_{I_>}}{\partial x_{I_> \cup I_=}}(\bar{x}), \tag{8.19}$$

$$\bar{\eta}_{I_=} \circ \xi_{I_=} + \bar{\xi}_{I_=} \circ \eta_{I_=} = 0.$$

By (8.18), this is further equivalent to saying that there exists no $\xi \neq 0$ satisfying

$$\frac{\partial F_{I_>}}{\partial x_{I_> \cup I_=}}(\bar{x})\xi_{I_> \cup I_=} = 0, \quad \xi_{I_<} = 0,$$

and such that (8.19) holds, and

$$\frac{\partial F_{I_=}}{\partial x_{I_> \cup I_=}}(\bar{x})\bar{\xi}_{I_> \cup I_=} \circ \xi_{I_=} + \bar{\xi}_{I_=} \circ \frac{\partial F_{I_=}}{\partial x_{I_> \cup I_=}}(\bar{x})\xi_{I_> \cup I_=} = 0.$$

We address the reader to [30, Section 4.2] and [59, Section 5] for further interpretations of Theorem 6.1 and the results in [59, Theorem 3.1, Corolaries 4.1, 4.2] for the constrained GN, the constrained LM, and the LPN methods applied to (6.1) with Φ and P defined in (8.17) and (8.11), respectively, and for illustrative examples.

9 Critical Lagrange multipliers

The notion of a critical solution of a nonlinear equation originates from the concept of a critical Lagrange multiplier in equality-constrained optimization, first introduced in [42], and further developed in [53]. Here, we discuss the relations between the two concepts and the impact of critical multipliers on the performance of algorithms only briefly, addressing the reader to the expositions of these and other related issues in [57], [58], and [56, Section 7.1].

Consider the equality-constrained optimization problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad h(x) = 0, \tag{9.1}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint mapping $h : \mathbb{R}^n \to \mathbb{R}^l$ are twice differentiable.

Stationary points and associated Lagrange multipliers of problem (9.1) are characterized by the Lagrange optimality system

$$\frac{\partial L}{\partial x}(x,\, \lambda) = 0, \quad h(x) = 0, \tag{9.2}$$

with respect to $(x,\, \lambda) \in \mathbb{R}^n \times \mathbb{R}^l$, where $L : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ is the Lagrangian of problem (9.1), defined as

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

Classical theory says that if $\bar{x}$ is a local solution of problem (9.1), and the constraints regularity condition

$$\operatorname{rank} h'(\bar{x}) = l \tag{9.3}$$

holds, then $\bar{x}$ is a stationary point of problem (9.1), i.e., that the set

$$\Lambda(\bar{x}) = \left\{ \lambda \in \mathbb{R}^l \;\middle|\; \frac{\partial L}{\partial x}(\bar{x},\, \lambda) = 0 \right\}$$

of Lagrange multipliers associated to $\bar{x}$ is nonempty. Moreover, for every stationary point $\bar{x}$, condition (9.3) is equivalent to saying that $\Lambda(\bar{x})$ is a singleton. In particular, if (9.3) is violated at a stationary point $\bar{x}$, then $\Lambda(\bar{x})$ is an affine manifold of a positive dimension, necessarily leading to nonisolated solutions of the system (9.2).

The Newton–Lagrange method (NLM) is the Newton method applied to the Lagrange system (9.2): for a current primal-dual iterate $(x^k,\, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^l$, the next iterate is

$(x^{k+1}, \lambda^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k)$, with a primal-dual displacement (ξ^k, η^k) defined by the linear system

$$\begin{gathered} \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi + (h'(x^k))^\top \eta = -\frac{\partial L}{\partial x}(x^k, \lambda^k), \\ h'(x^k)\xi = -h(x^k). \end{gathered} \tag{9.4}$$

Guarantees of local superlinear convergence of the NLM method to a solution $(\bar{x}, \bar{\lambda})$ of the Lagrange system (9.2) include at least the constraints regularity condition (9.3), this condition being necessary for the the Jacobian of the system (9.2) at $(\bar{x}, \bar{\lambda})$ to be nonsingular; see [56, Theorem 4.3]. The cases when (9.3) is not expected to hold can be tackled by the LM or the LPN methods, but there also exist special dual stabilization techniques, employing the primal-dual structure of (9.2). Among such algorithms, a prominent one is the stabilized NLM (originating from [78]; see also [58, Section 4] and [56, Section 7.2.2], and extensive bibliography therein). The stabilized NLM has the iteration linear system

$$\begin{gathered} \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)\xi + (h'(x^k))^\top \eta = -\frac{\partial L}{\partial x}(x^k, \lambda^k), \\ h'(x^k)\xi - \sigma_k \eta = -h(x^k), \end{gathered} \tag{9.5}$$

where $\sigma_k \geq 0$ is the dual stabilization parameter. If this parameter equals 0, the stabilized NLM iteration system (9.5) coincides with the NLM iteration system (9.4). The point is that managing appropriately positive values of σ_k entails a stabilizing effect for the dual iterates. In [20], the stabilized NLM method was shown to converge locally superlinearly when the dual starting point is close enough to $\bar{\lambda} \in \Lambda(\bar{x})$ such that the second-order sufficient optimality condition (SOSC)

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\} \tag{9.6}$$

is satisfied, even if the constraints regularity condition (9.3) is violated. Somewhat surprisingly, the same local convergence properties were established in [21] for the classical Augmented Lagrangian method (method of multipliers) [9], assuming an appropriate control of the penalty parameter.

At this point we recall that a Lagrange multiplier $\bar{\lambda}$ associated to a stationary $\bar{x}$ of the problem (9.1) is called critical if there exists $\xi \in \ker h'(\bar{x}) \setminus \{0\}$ such that

$$\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^\top,$$

and noncritical otherwise. In other words, $\bar{\lambda}$ is critical if the corresponding reduced Hessian of the Lagrangian is singular, where the reduced Hessian is understood as the symmetric matrix $H(\bar{x}, \bar{\lambda})$ of the quadratic form

$$\xi \mapsto \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle : \ker h'(\bar{x}) \to \mathbb{R}.$$

It then becomes evident that the multiplier $\bar{\lambda}$ is necessarily noncritical if it satisfies the SOSC (9.6), and moreover, noncriticality is equivalent to SOSC if the condition

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle \geq 0 \quad \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}$$

holds with this $\bar{\lambda}$. However, in general, a multiplier can be noncritical when it violates SOSC (9.6), and moreover, the subset of critical multipliers $\lambda \in \Lambda(\bar{x})$ is characterized by the algebraic equation $\det H(\bar{x}, \lambda) = 0$, and hence, it is typically thin within $\Lambda(\bar{x})$.

Local superlinear convergence of the stabilized NLM with the SOSC (9.6) replaced by a weaker assumption that $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical was established in [55], while for the augmented Lagrangian method, such result was derived in [46]. Along with that, the behavior of various methods near critical multipliers was also actively studied in the literature; see [53], [52], [43], [54], [60], [61], [62], [47], [44], and an overview in [57]. The overall understanding emerging from these works is that methods supplied with stabilization mechanisms may still be attracted to critical multipliers from large domains of starting points, and this is the reason for the lack of superlinear rate of convergence. A technique of generating and employing special dual iterates (instead of those naturally computed by methods in question), and allowing to reduce the negative effect of attraction was proposed in [28].

In order to place these developments in the perspective of the current survey, we now discuss the relations between criticality of a multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ and of the related solution $(\bar{x}, \bar{\lambda})$ of the Lagrange system (9.2) that can be written as the equation (1.1) if we set $u = (x, \lambda)$ (and $p = n + l$), and define Φ as

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda), h(x) \right). \tag{9.7}$$

The following result was obtained in [48, Proposition 2].

Proposition 9.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable at $\bar{x} \in \mathbb{R}^n$. Let $\bar{x}$ be a stationary point of the optimization problem* (9.1)*, and let $\bar{\lambda} \in \mathbb{R}^l$ be an associated Lagrange multiplier.*

If $\bar{\lambda}$ is a noncritical Lagrange multiplier, then $\bar{u} = (\bar{x}, \bar{\lambda})$ is a noncritical solution of the equation (1.1) *with Φ defined in* (9.7)*.*

Moreover, if $\bar{x}$ is an isolated stationary point, then $\bar{u} = (\bar{x}, \bar{\lambda})$ is a critical solution of (1.1) *if and only if $\bar{\lambda}$ is a critical Lagrange multiplier.*

The next question to be addressed is the fulfilment of Key Assumption (2.1) of this paper, stated in Section 2. In order to do this, define a linear subspace

$$Q(\bar{x}, \bar{\lambda}) = \left\{ \xi \in \ker h'(\bar{x}) \;\middle|\; \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^\top \right\}. \tag{9.8}$$

Criticality of a multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ means that this linear subspace is nontrivial, and $\dim Q(\bar{x}, \bar{\lambda})$ is refereed to as the order of criticality. The next result is [48, Proposition 4].

Proposition 9.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be three times differentiable at $\bar{x} \in \mathbb{R}^n$. Let $\bar{x}$ be a stationary point of the optimization problem* (9.1)*, and let $\bar{\lambda} \in \mathbb{R}^l$ be an associated Lagrange multiplier. 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 1.*

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

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

We conclude this discussion by mentioning that the case when $\bar{\lambda}$ is critical of order higher than 1 opens wide possibilities for Key Assumption (2.1) to hold, and hence, for the results based on it, surveyed in this paper, to be valid.

The concept of a critical multiplier and related results extend (at least partially) in various directions, one of them being optimization and variational problems with inequality constraints; see [52], [55], and [56, Definitions 1.41, 7.8]. In [19], conditions for primal superlinear convergence of the variants of sequential quadratic programming methods to noncritical multipliers were investigated, while techniques for improving the performance of Newton-type methods in the presence of critical multipliers were addressed in [49], [29]. Further extensions are concerned with reducing the smoothness assumptions [45], and with recent development and applications of this concept to extended optimization problems, including semidefinite and second-order cone problems, composite optimization problems, etc.; see [68], [71], [13], [72], [81], [76], [8].

10 Open questions

We conclude with stating some issues that require further investigation:

- The two phenomena being observed for critical solutions, namely, the attraction of iterative sequences generated by Newton-type methods, and special stability properties of such solutions subject to perturbations, have been studied so far somehow separately. Establishing more clear relations between them would be of interest, and might provide some further insight into the nature of criticality. For instance, one might expect that the attraction phenomenon to critical solutions can be somehow explained by their special stability properties. One reason for this point of view is that the Newtonian subproblems can be naturally considered as some kind of perturbations of the problem being solved. This line of analysis has not been pursued, up to now.

- Detailed numerical comparison of (globalized) methods fitting the pNM framework (LM, LPN, primal-dual methods for optimization and variational problems) for equations with singular solutions, under various smoothness assumptions.

- When convergence is to a critical solution, the progress in the distance to the solution is not necessarily properly reflected by the decrease of the residual of the equation. Therefore, "catching" (or recognizing) the effect of acceleration techniques by the residual remains an issue.

- Globalization of convergence for methods with overrelaxation and Anderson acceleration, preserving their accelerating properties. Comparison of the resulting algorithms.

- For optimizations problems, globalization of convergence of methods with the modification of dual iterates proposed in [28]; its use with methods other than sSQP. Possible development of similar techniques for avoiding convergence to critical solutions for equations that are not related to optimization.

- Further insights into criticality of solutions and its roles under the reduced smoothness assumptions (piecewise smoothness, general semismoothness).

- Full understanding of critical solutions of constrained equations, without the interiority assumption, at least for polyhedral constraints.

Declarations

Conflicts of Interest. The author A.F. Izmailov is Associate Editor of the Journal. 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] D.G. Anderson. Iterative procedures for nonlinear integral equations. J. Assoc. Comput. Mach. 12 (1965), 547–560.

[2] A.V. Arutyunov and A.F. Izmailov. Stability of possibly nonisolated solutions of constrained equations, with applications to complementarity and equilibrium problems. Set-Valued Var. Analys. 26 (2018), 327–352.

[3] A.V. Arutyunov and A.F. Izmailov. Covering on a convex set in the absence of Robinson's regularity. SIAM J. Optim. 30 (2020), 604–629.

[4] E.R. Avakov. Extremum conditions for smooth problems with equality-type constraints. USSR Comput. Math. Math. Phys. 25 (1985), 24–32.

[5] R. Behling and A. Iusem. The effect of calmness on the solution set of systems of nonlinear equations. Math. Program. (2013) 137, 155–165.

[6] M. Benko and P. Mehltiz. On the directional asymptotic approach in optimization theory. Part A: approximate, M-, and mixed-order stationarity. 2022. arXiv.2204.13932.

[7] M. Benko and P. Mehltiz. On the directional asymptotic approach in optimization theory. Part B: constraint qualifications. 2022. arXiv.2205.00775.

[8] M. Benko and P. Mehltiz. Isolated calmness of perturbation mappings in generalized nonlinear programming and local superlinear convergence of Newton-type methods. 2024. arXiv.2401.08163v1.

[9] D.P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, 1982.

[10] F.H. Clarke. Optimization and Nonsmooth Analysis. John Wiley, New York, 1983.

[11] M. Dallas and S. Pollock. Newton-Anderson at singular points. Int. J. Numer. Anal. Mod. 20 (2023), 667–692.

[12] J.E. Dennis and R.B. Schnabel. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Inc., Englewood Cliffs, 1983.

[13] H. Do, B.S. Mordukhovich, and M.E. Sarabi. Criticality of Lagrange multipliers in extended nonlinear optimization. Optimizat. 79 (2019), 511–544.

[14] Y.G. Evtushenko and V.A. Purtov. Sufficient conditions for a minimum for nonlinear programming problems. Soviet Math. Dokl. 30 (1984), 313–316.

[15] 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.

[16] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. 146 (2014), 1–36.

[17] F. Facchinei and J.-S. Pang. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York, 2003.

[18] J.-Y. Fan and Y.-X. Yuan. On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption. Computing. 74 (2005), 23–39.

[19] D. Fernández, A.F. Izmailov, and M.V. Solodov. Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. SIAM J. Optim. 20 (2010), 3312–3334.

[20] D. Fernández and M.V. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. 125 (2010), 47–73.

[21] D. Fernández and M.V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. 22 (2012), 384–407.

[22] 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.

[23] A. Fischer, M. Herrich, A.F. Izmailov, and M.V. Solodov. A globally convergent LP-Newton method. SIAM J. Optim. 26 (2016), 2012–2033.

[24] A. Fischer, A.F. Izmailov, and M. Jelitte. Constrained Lipschitzian error bounds and noncritical solutions of constrained equations. Set-Valued Var. Analys. 29 (2021), 745–765.

[25] A. Fischer, A.F. Izmailov, and M. Jelitte. Newton-type methods near critical solutions of piecewise smooth nonlinear equations. Comput. Optim. Appl. 80 (2021), 587–615.

[26] 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. 2023. https://doi.org/10.1007/s10957-023-02350-w.

[27] A. Fischer, A.F. Izmailov, and M. Jelitte M. Stability of singular solutions of nonlinear equations with restricted smoothness assumptions. J. Optim. Theory Appl. 196 (2023), 1008–1035.

[28] A. Fischer, A.F. Izmailov, and W. Scheck. Adjusting dual iterates in the presence of critical Lagrange multipliers. SIAM J. Optim. 30 (2020), 1555–1581.

[29] A. Fischer, A.F. Izmailov, and W. Scheck. Techniques for adjusting dual iterates in the presence of critical Lagrange multipliers for optimization problems with inequality constraints Optimization. 2023. https://doi.org/10.1080/02331934.2023.2277715.

[30] A. Fischer, A.F. Izmailov, and M.V. Solodov. Local attractors of Newton-type methods for constrained equations and complementarity problems with nonisolated solutions. J. Optim. Theory Appl. 180 (2019), 140–169.

[31] A. Fischer, A.F. Izmailov, and M.V. Solodov. Accelerating convergence of the globalized Newton method to critical solutions of nonlinear equations. Comput. Optim. Appl. 78 (2021), 273–286.

[32] A. Fischer, A.F. Izmailov, and M.V. Solodov. Unit stepsize for the Newton method close to critical solutions. Math. Program. 187 (2021). 697–721.

[33] A. Fischer, A.F. Izmailov, and M.V. Solodov. The Levenberg–Marquardt method: An overview of modern convergence theories and more. Comput. Optim. Appl. 2024. https://doi.org/10.1007/s10589-024-00589-1.

[34] A. Fischer and P.K. Shukla. A Levenberg–Marquardt algorithm for unconstrained multicriteria optimization. Oper. Res. Lett. 36 (2008), 643–646.

[35] H. Gfrerer. On metric pseudo-(sub)regularity of multifunctions and optimality conditions for degenerated mathematical programs. Set-Valued Var. Anal. 22 (2013), 79–115.

[36] 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.

[37] H. Gfrerer and J.V. Outrata. On computation of limiting coderivatives of the normal-cone mapping to inequality systems and their applications. Optimization 65 (2016), 671–700.

[38] A. Griewank. Analysis and modification of Newton's method at singularities. PhD Thesis. Australian National University, Canberra (1980)

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

[40] A. Griewank. On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. 27 (1985), 537–563.

[41] W.W. Hager. Lipschitz continuity for constrained processes. SIAM J. Control Optim. 17 (1979), 321–338.

[42] A.F. Izmailov. On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. 45 (2005), 930–946.

[43] A.F. Izmailov. On the limiting properties of dual trajectories in the Lagrange multipliers method. Comput. Math. Math. Phys. 51 (2011), 1–20.

[44] A.F. Izmailov. Accelerating convergence of a globalized sequential quadratic programming method to critical Lagrange multipliers. Comput. Optim. Appl. 80 (2021), 943–978.

[45] 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.

[46] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Local convergence of the method of multipliers for variational and optimization problems under the noncriticality assumption. Comput. Optim. Appl. 60 (2015), 111–140.

[47] 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.

[48] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Critical solutions of nonlinear equations: Stability issues. Math. Program. 168 (2018), 475–507.

[49] A.F. Izmailov and I.S. Rodin. Accelerating sequential quadratic programming for inequality-constrained optimization near critical Lagrange multipliers. Adv. Syst. Sci. Appl. 22 (2022), 73–84.

[50] 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.

[51] 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.

[52] A.F. Izmailov and M.V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl. 42 (2009), 231–264.

[53] 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.

[54] 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. Prog. 126 (2011), 231–257.

[55] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. Math. Program. 122 (2012), 93–120.

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

[57] 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.

[58] A.F. Izmailov and M.V. Solodov. Rejoinder on: Critical Lagrange multipliers: What we currently know about them, how they spoil our lives, and what we can do about it. TOP 23 (2015), 48–52.

[59] A.F. Izmailov and M.V. Solodov. On local behavior of Newton-type methods near critical solutions of constrained equations. J. Optim. Theory Appl. 2024. https://doi.org/10.1007/s10957-023-02367-1.

[60] A. Izmailov and E. Uskov. On the influence of the critical Lagrange multipliers on the convergence rate of the multiplier method. Comput. Math. Math. Phys. 52 (2012), 1504–1519.

[61] A. Izmailov and E. Uskov. Attraction of Newton method to critical Lagrange multipliers: fully quadratic case. Math. Program. 152 (2015), 33–73.

[62] A. Izmailov and E. Uskov. Subspace-stabilized sequential quadratic programming. Comput. Optim. Appl. 67 (2017), 129–154.

[63] A. Izmailov, E. Uskov, and Y. Zhibai. Newton method vs. semismooth Newton method for singular solutions of nonlinear complementarity problems. Adv. Syst. Sci. Appl. 23 (2023). 16–26.

[64] C. Kanzow. Some equation-based methods for the nonlinear complementarity problem. Optim. Methods Softw. 3 (1994), 327–340.

[65] 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.

[66] O.L. Mangasarian. Equivalence of the complementarity problem to a system of nonlinear equations. SIAM J. Appl. Math. 31 (1976), 89–92.

[67] K. Levenberg. A method for the solution of certain non-linear problems in least squares. Quarterly of Appl. Math. 2 (1944), 164–168.

[68] Y. Liu and S. Pan. Strong calmness of perturbed KKT system for a class of conic programming with degenerate solutions. Optimization 68 (2018), 1131–1156.

[69] D.W. Marquardt. An algorithm for least squares estimation of non-linear parameters. SIAM J. 11 (1963), 431–441.

[70] B.S. Mordukhovich. Variational Analysis and Generalized Differentiation. Springer, Berlin, 2006.

[71] B.S. Mordukhovich and M.E. Sarabi. Critical multipliers in variational systems via second-order generalized differentiation. Math. Program. 169 (2018), 605–648.

[72] B.S. Mordukhovich and M.E. Sarabi. Criticality of Lagrange multipliers in variational systems. SIAM J. Optim. 29 (2019), 1524–1557.

[73] H.V. Ngai, N.H. Tron, N.V. Vu, and M. Théra. Directional metric pseudo subregularity of set-valued mappings: a general model. Set-Valued Var. Anal. 28 (2020), 61–87.

[74] 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.

[75] L. Qi and J. Sun. A nonsmooth version of Newton's method. Math. Program. 58 (1993), 353–367.

[76] M.E. Sarabi. Primal superlinear convergence of SQP methods in piecewise linear-quadratic composite optimization. Set-Valued and Variational Analysis, 30 (2022), 1–37.

[77] A.A. Tretyakov. Necessary and sufficient conditions for optimality of p-th order. USSR Comput. Math. Math. Phys. 24 (1984), 123–127.

[78] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. 11 (1998), 253–275.

[79] 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.

[80] J.-L. Zhang. On the convergence properties of the Levenberg–Marquardt method. Optimization 52 (2003), 739–756.

[81] T.Y. Zhang and L.W. Zhang. Critical multipliers in semidefinite programming. Asia-Pacific J. Oper. Res.(2020). Art. number 2040012.