Newton's method for continuous functions ?

Bernd Kummer *

August 15, 2012

Abstract. Recently, a paper [14] on Newton's method for continuous functions in finite dimension appeared. We check and compare the approach with known ones for locally Lipschitz functions. It turns out that [14] contains no non-Lipschitz functions such that the claimed local or global convergence holds true. Moreover, the given sufficient condition based on directional boundedness even prevents local superlinear convergence for real, non-Lipschitz functions. The hypotheses for global convergence imply directly the global Lipschitz property on the crucial set.

Additionally, we present some convergence statements for the Lipschitz case, certain auxiliary results for continuous functions as well as non-Lipschitz examples of different type where Newton's method, indeed, superlinearly converges. Three errors concerning inverse mappings, semismoothness and the proof of global convergence will be corrected, too.

Key words. Newton's method, nonsmooth continuous equations, graphical derivatives, local and global convergence, concrete examples.

Mathematics Subject Classification 2000. 49J53, 49K40, 90C31, 65J05.

1 Introduction

The paper [14] on Newton's method (briefly NM) for continuous functions $f : \mathbb{R}^n \to \mathbb{R}^n$ begins with the following ambitious abstract.

"This paper concerns developing a numerical method of the Newton type to solve systems of nonlinear equations described by nonsmooth continuous functions. We propose and justify a new generalized Newton algorithm based on graphical derivatives, which have never been used to derive a Newton-type method for solving nonsmooth equations. Based on advanced techniques of variational analysis and generalized differentiation, we establish the well-posedness of the algorithm, its local superlinear convergence, and its global convergence of the Kantorovich type. Our convergence results hold with no semismoothness assumption, which is illustrated by examples. The algorithm and main results obtained in the paper are compared with well-recognized semismooth and B-differentiable versions of Newton's method for nonsmooth Lipschitzian equations."

Since relevant papers can be easily overlooked, we consider next mainly papers which are also cited in [14], in particular [20] and [23]. Additional references can be found in the summary. Right, for locally Lipschitz f (briefly $f \in C^{0,1}$), Newton methods have been investigated under several viewpoints in [8], [20], [32], [33], [35], [36], [38] mentioned in [14] and here. Without supposing semismoothness, they are studied in [20, Chapter 10]. In Sect. 10.3, NM is explicitly based on graphical derivatives Cf (notation from [39], we say contingent

*Institut für Mathematik, Humboldt–Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany. E–Mail: kummer@math.hu-berlin.de

derivatives as in [1]), on Thibault derivatives Tf, directional derivatives $f'(x;u)$ and generalized Jacobians ∂f. There, one finds also a detailed analysis of approximate solutions and of the required conditions. Continuous f, multifunctions $f(x)+\Gamma(x)$, and approximations (e.g. by contingent derivatives) are permitted in [23] with auxiliary problems $0 \in F(x_{k+1}, x_k)$, in spite of other comments in [14].

Nevertheless, the present paper is not written because of missed references. It is written for mathematical reasons. Though, for $f \in C^{0,1}$, there is already a developed theory of "graphical derivative based" NM (Section 3.1 contains few main topics), new results for the class $\mathcal{D}$ of *continuous, not locally Lipschitz functions* would be really of interest. Having our comment in [20] for such extensions in mind (here Rem. 3.7), some skepticism is advisable. In particular, the reader of [14] learns nothing for the case of $f \in \mathcal{D}$:

(i) A first observation shows that [14] does not contain any concrete function $f \in \mathcal{D}$ the convergence results can be applied to. Though the contrary is asserted everywhere and examples are added which satisfy (or do not satisfy) *some* imposed conditions, they never fulfill *all* requirements of the local/global convergence-statements [14, Thm. 3.3/3.4], respectively. This is also true for the most complicated example Ex. 3.16 = [14, Ex. 4.10].

(ii) The *given sufficient conditions* for the hypotheses of Thm. 3.11 = [14, Thm. 3.3], namely (H1), (H2) and metric regularity (MR) together (in particular for (H2) alone) concern only semismooth functions $f \in C^{0,1}$. Thus the extension [14, Thm. 3.3] to continuous f may concern the empty set. This is not surprising since directional boundedness (3.5) along with (MR) just implies, at least for real $f \in \mathcal{D}$, that NM *cannot* superlinearly converge, cf. Thm. 4.1. Note that (3.5) plus *strong* regularity was the only sufficient condition for (H1) presented in [14], cf. Prop. 3.13 =[14, Prop. 4.4]

(iii) By Rem. 3.18, the hypotheses of Prop. 3.17 = [14, Thm. 3.4] automatically imply that the continuous function f is globally Lipschitz on the crucial set Ω. In addition, our Rem. 2.8 shows why the authors proof is wrong and works only in the trivial situation when also f^{-1} is Lipschitz of $f(\Omega)$. More comments are contained in Sect. 3.2.3.

(iv) The example [23, Sect.2.3] = [20, BE1] = [8, Ex. 7.4.1] shows, in contrary to the assertion at the end of [14, Sect. 5]: Nonsingularity of the generalized Jacobian *does not imply semismoothness* of a Lipschitzian transformation $H : \mathbb{R}^n \to \mathbb{R}^n$. Rem. 2.1 indicates another error related to the inverse.

(v) Speaking (as in the abstract) about a "numerical method" is not justified. For piecewise C^1 functions, the auxiliary problems (2.3) are linear complementarity problems. For continuous f, they are highly nontrivial without any tool for computing the crucial sets $Cf(x)(u)$. In view of using "advanced techniques", the authors are not stringent in doing this. On the one hand, they apply $Cf(x)$ for NM, on the other hand they ignore that (MR) guarantees immediately solvability of the auxiliary problems by the well-known openness condition (2.16). Instead, they emphasize the (MR)-characterization $\ker D^*f(\bar{x}) = \{0\}$ as being an essential tool. This may be true, but not in the present context where all essential statements depend on Cf and $\tilde{C}f$ only.

(vi) Metric regularity, throughout required, excludes not only the abs-value-function from all considerations. It is too strong for [14, Thm. 3.3] and too weak at least for the proof of [14, Thm. 3.4].

In consequence, one has even to ask:

(vii) Does there exist at least one function $f \in \mathcal{D}$ such that NM, based on contingent (= graphical) derivatives, converges locally superlinear ?

We give a positive answer by the examples 4.3 and 4.4 which do not satisfy the basic hypothesis (H1) of [14]. The first one does neither satisfy the approximation condition (CA)*, known from the $C^{0,1}$-case, nor (H2). The second one satisfies (H2). Hence also the hypothesis (H2), which is (CA)* for $f \in C^{0,1}$, is no longer crucial for $f \in \mathcal{D}$.

In what follows, we discuss the needed assumptions for the convergence-statements and their consequences in a detailed manner for $f \in C^{0,1}$ and $f \in \mathcal{D}$ and justify the assertions (i), (ii), (iii). We do not comment the selected auxiliary statements of nonsmooth analysis in [14]. In Section 2, we summarize general facts for generalized NM and necessary tools concerning stability and solvability. In Section 3, we compare the approaches of [14] and [20]. The reader, interested in $f \in \mathcal{D}$ only, may omit section 3.1 where we add, for $f \in C^{0,1} \cup \mathcal{D}$, only statements of [20] which are needed for comparisons. Hence we omit the study of *Newton-maps or locPC1 functions* of [20, § 6.4.2] as well as *applications* to complementarity or KKT- systems of chapter 11. Section 4 presents (perhaps indeed new) helpful statements and examples for the real case.

Throughout the paper, we suppose that $f : \mathbb{R}^n \to \mathbb{R}^n$ is *continuous* if nothing else is explicitly said. We write $f \in C^{0,1}$ to say that f is locally Lipschitz near the reference point $\bar{x}$ and $f \in \mathcal{D}$ otherwise. All x, y, u, v, with or without an index, belong to $\mathbb{R}^n$. Our notations are standard in nonsmooth analysis and coincide with [14] where, however, Df stands for Cf. Here, Df denotes the Fréchet derivative. By x *near* $\bar{x}$ we abbreviate *for all* x *in some neighborhood of* $\bar{x}$. For f being directionally differentiable at x, we write $f'(x; u) = \lim_{t \downarrow 0} t^{-1}(f(x + tu) - f(x))$.

2 Generalized NM and superlinear convergence

Contingent derivative: Given any multifunction $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $y \in F(x)$ the contingent (= graphical or Bouligand-) derivative of F at (x, y) is defined by

$$v \in CF(x, y)(u) \ \text{ if } \ \exists\, t_k \downarrow 0, (u_k, v_k) \to (u, v) : \ (x + t_k u_k, \ y + t_k v_k) \in \operatorname{gph} F. \tag{2.1}$$

The symmetric form yields for the (multivalued) inverse F^{-1} the well-known [1] formula

$$u \in CF^{-1}(y, x)(v) \ \Leftrightarrow \ v \in CF(x, y)(u). \tag{2.2}$$

Setting $u_k = u$ in (2.1), one obtains subsets $\tilde{C}F(x, y)(u) \subset CF(x, y)(u)$.

Remark 2.1. *In spite of [14, Prop. 2.2],* $\tilde{C}F$ *does not satisfy (2.2); take* $F(x) = \{x^3\}$ *with* $0 \in \tilde{C}F(0, 0)(1)$ *and* $1 \notin \tilde{C}F^{-1}(0, 0)(0)$.

For functions f, $y = f(x)$ is unique, and one writes $Cf(x, f(x))(u) = Cf(x)(u)$. Clearly, the inclusions $0 \in \tilde{C}f(x)(0) \cap Cf(x)(0)$ and $\tilde{C}f(x)(0) = \{0\}$ are always true.

2.1 Newton methods and related generalized derivatives

To describe different Newton methods for $f : \mathbb{R}^n \to \mathbb{R}^n$, let $x \in \mathbb{R}^n$ and $Gf(x) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be any multifunction. Newton iterations may depend on the "generalized derivative" Gf. Given x (near a zero $\bar{x}$ of f) the next (Newton-) iterate is, by definition, any x' such that

$$-f(x) \ \in \ Gf(x)\ (x' - x). \tag{2.3}$$

Let $\sigma(x)$ denote the set of such x'. *Local superlinear convergence of NM* then means

$$\begin{array}{c} \sigma(x) \neq \emptyset \ \ \forall x \text{ near } \bar{x} \quad \text{and} \quad \forall x' \in \sigma(x) \ \text{ it holds} \\ \frac{x' - \bar{x}}{\|x - \bar{x}\|} \to 0 \text{ as } x \to \bar{x}, \ x \neq \bar{x}, \qquad i.e., \quad \|x' - \bar{x}\| = o(x - \bar{x}). \end{array} \tag{2.4}$$

Evidently, for x_0 sufficiently close to $\bar{x}$, then the procedure

$$\text{find } x_{k+1} \text{ such that } -f(x_k) \in Gf(x_k)\,(x_{k+1}-x_k); \quad k=0,1,2,... \tag{2.5}$$

is well defined and generates a sequence with $\|x_{k+1}-\bar{x}\| = o(x_k-\bar{x})$. Using (2.2) for $Gf = Cf$, the iterations (2.5) can be also written as

$$\text{find } x_{k+1} \in x_k + Cf^{-1}(f(x_k),x_k)\,(-f(x_k)); \quad k=0,1,2,... \tag{2.6}$$

Possible settings
In the classical case, we have $Gf(x)(x'-x) = \{Df(x)(x'-x)\}$ and (2.3) is the usual Newton equation $f(x)+Df(x)(x'-x)=0$. Standard non-smooth Newton methods use non-empty sets $M(x)$ of regular matrices,

$$Gf(x)(u) = \{Au \mid A \in M(x)\} \text{ and solve } f(x)+A(x'-x)=0 \text{ with any } A \in M(x). \tag{2.7}$$

Regularity of all $A \in M(x)$ then implies

$$\emptyset \neq Gf(x)(u) \text{ and } Gf(x)(0) = \{0\}. \tag{2.8}$$

Other possible settings for $x,u \in \mathbb{R}^n$ are, e.g.,

$$Gf(x)(u) = \begin{cases} \{f'(x;u)\} & \text{if } f \text{ is directionally differentiable near } \bar{x} \\ \tilde{C}f(x)(u) & \text{Set of all directional limits in direct. } u \\ Cf(x)(u) & \text{Contingent derivative in direct. } u \\ Tf(x)(u) & \text{Thibault derivative in direct. } u \\ \partial f(x)(u) & \text{Clarke's generalized Jacobian applied to } u \text{ if } f \in C^{0,1}. \end{cases} \tag{2.9}$$

The sets $\tilde{C}f(x)(u)$, $Cf(x)(u)$, $Tf(x)(u)$ contain, by definition, exactly all limits of sequences $\{v_k\} \in \mathbb{R}^n$; $k=1,2,...$ where $t_k \downarrow 0$ and

$$\begin{array}{lll} \text{for } \tilde{C}f(x)(u): & v_k = t_k^{-1}\,[f(x+t_ku)-f(x)], & \\ \text{for } Cf(x)(u): & v_k = t_k^{-1}\,[f(x+t_ku_k)-f(x)] & \text{with } u_k \to u, \\ \text{for } Tf(x)(u): & v_k = t_k^{-1}\,[f(x_k+t_ku_k)-f(x_k)] & \text{with } u_k \to u,\ x_k \to x. \end{array} \tag{2.10}$$

These limit sets are written as Limsup in [14]. Cf and $\tilde{C}f$ correspond to definition (2.1) for multivalued F.

To introduce $\partial_B f(x)$, we recall Clarke's [5, 6] definition of $\partial f(x)$ for $f \in C^{0,1}$. Since $\mathcal{N} := \{y \in \mathbb{R}^n \mid Df(y) \text{ does not exist}\}$ has Lebesgue measure zero (Rademacher), the set $\mathcal{M} := \{A \mid A = \lim Df(x_k) \text{ where } x_k \in \mathbb{R}^n \setminus \mathcal{N} \text{ and } x_k \to x\}$ is compact and not empty. The set $\partial f(x) = \operatorname{conv} \mathcal{M}$ (convex hull) is Clarke's *generalized Jacobian*, and $\mathcal{M}$ itself is often called the B-differential $\partial_B f(x)$ of f at x. One easily shows $\partial_B f(x)(u) \subset Tf(x)(u)$.

Using ∂f or $\partial_B f$ in NM (2.3) means to put $M(x) = \partial f(x)$ or $M(x) = \partial_B f$ in (2.7).

Injectivity and $\ker Gf(x)$**:** As in [20], we call $Gf(x)$ *injective* if

$$v \in Gf(x)(u) \text{ implies } \|v\| \geq c\,\|u\| \text{ with some constant } c > 0. \tag{2.11}$$

Since all mappings $Gf(x)$ (2.7), (2.9) - and $D^*f(x)$, too - are positively homogeneous with empty or non-empty images, this is just $0 \notin Gf(x)(u)\ \forall u \in \mathbb{R}^n \setminus \{0\}$ or, in other words,

$$\ker Gf(x) = \{0\}. \tag{2.12}$$

Hence $\ker \partial f(\bar{x}) = \{0\} \Leftrightarrow$ all $A \in \partial f(\bar{x})$ are non-singular; while e.g. [20, formula (3.5)] says

$$Tf(\bar{x}) \text{ is injective } \Leftrightarrow \exists c > 0 \text{ such that } \|f(y)-f(x)\| \geq c\,\|y-x\|\ \forall x,y \text{ near } \bar{x}. \tag{2.13}$$

Inclusions: Let $f \in C^{0,1}$. Then, setting $u_k = u$ in (2.10), one obtains, for x near $\bar{x}$, the same sets $Cf(x)(u)$ and $Tf(x)(u)$. These sets are non-empty, satisfy (2.8) as well as

$$\tilde{C}f(x)(u) = Cf(x)(u) \subset Tf(x)(u) \subset \partial f(x)(u) \subset L\|u\|B, \tag{2.14}$$

if L is bigger than some Lipschitz rank for f near $\bar{x}$. The inclusion $Tf(x)(u) \subset \partial f(x)(u)$ is non-trivial and needs the mean-value theorem for ∂f in [6]. The others follow immediately from (2.10) like

$$\tilde{C}f(x)(u) \subset Cf(x)(u) \subset Tf(x)(u) \quad \text{for arbitrary } f.$$

If f is C^1 near x (not only differentiable), all Gf (2.9) fulfill $Gf(x)(u) = \{Df(x)u\}$ by the usual mean-value theorem. Then the Newton steps (2.3) coincide with the usual ones at all "C^1-points" (which can form the empty set).

In [14], mainly $\tilde{C}f$ and Cf are used. In [20], all settings (2.7) and (2.9) were studied, but mostly by supposing $f \in C^{0,1}$. A strange situation for $f \in \mathcal{D}$ indicates

Example 2.2. *For the real (strongly regular) function* $f(x) = \begin{cases} +\sqrt{x} & \text{if } x > 0 \\ x & \text{if } x \le 0 \end{cases}$ *the usual NM finds the zero after at most two steps. But (2.4) and (2.26) are violated for all mappings* Gf *in (2.9) since* $x' = -x$ *for* $x > 0$.

2.2 The needed tools of variational analysis

2.2.1 Known properties of the inverse

Locally Lipschitz properties of $f^{-1}(y) = \{x \mid f(x) = y\}$ are helpful to ensure solvability of the auxiliary problems (2.3) and to understand the imposed conditions below. For $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, they simply require $\det Df(\bar{x}) \neq 0$.

A (continuous) function $f : \mathbb{R}^n \to \mathbb{R}^n$ is called *metrically regular (MR)* near $\bar{x} \in \mathbb{R}^n$ if, for some $\mu > 0$ and neighborhoods U, V of $\bar{x}$ and $f(\bar{x})$, respectively, it holds

$$(\ x \in U,\ y' \in V\ ) \ \Rightarrow\ \exists x' :\ f(x') = y' \text{ and } \|x' - x\| \le \mu\ \|y' - f(x)\|. \tag{2.15}$$

With contingent derivatives and the unit-ball-notation in $\mathbb{R}^n$, it holds

$$f \text{ is (MR) near } \bar{x} \ \Leftrightarrow\ \exists \mu > 0 :\ B \subset Cf(x)(\mu B)\ \forall x \text{ near } \bar{x}, \tag{2.16}$$

see, e.g. the openness conditions in [1], [7], [9], [15], [34] or, in view of "more regularities", [20, Thm. 5.1]. By (2.2), this is in terms of the multifunction f^{-1},

$$f \text{ is (MR) near } \bar{x} \ \Leftrightarrow\ \exists \mu > 0 :\ \mu\|v\|B \cap Cf^{-1}(f(x), x)(v) \neq \emptyset\ \ \forall v\ \forall x \text{ near } \bar{x}. \tag{2.17}$$

Proof. Indeed, (2.16) means $\forall v \in B\ \exists u \in \mu B : v \in Cf(x)(u)$. This can be written as $\forall v\ \exists u \in \mu\|v\|B : v \in Cf(x)(u)$, *i.e.*, $\forall v\ \exists u \in \mu\|v\|B : u \in Cf^{-1}(f(x), x)(v)$. □

With the coderivative D^*f in [30, 31], based on the behavior of the functions $f_{y^*}(x) = \langle\ y^*, f(x)\ \rangle$ *near* $\bar{x}$, there is a second condition,

$$f \text{ is (MR) near } \bar{x} \ \Leftrightarrow\ \ker D^*f(\bar{x}) = \{0\}. \tag{2.18}$$

In (2.16) and (2.18), implication (⇒) holds due to finite dimension, implication (⇐) due to Ekeland's variational principle or an equivalent statement. Applying, e.g., [20, Thm. 5.3], the *pointwise characterization* (2.18) means explicitly

$$\ker D^*f(\bar{x}) = \{0\} \Leftrightarrow \forall y^* \in \mathbb{R}^n \setminus \{0\}\ \forall x_k \to \bar{x} :\ \limsup_{k\to\infty}\ \sup_{\zeta\, \in\, Cf(x_k)(B)} \langle y^*, \zeta\rangle > 0 \tag{2.19}$$

and is as *"pointwise"* as condition (2.16). The stronger requirement that, for certain U, V, x' is even unique in (2.15), claims equivalently that f has a locally single-valued Lipschitzian inverse f^{-1} sending V into U, and is often called *strong regularity* as in [37]. Thus metric and strong regularity coincide if f is 1-to-1 near $\bar{x}$. For $f \in C^{0,1}$, it holds

$$f \text{ is strongly regular near } \bar{x} \text{ if } \ker \partial f(\bar{x}) = \{0\}, \qquad \text{cf. [5]} \tag{2.20}$$

$$f \text{ is strongly regular near } \bar{x} \Leftrightarrow \ker Tf(\bar{x}) = \{0\}, \text{ cf. [27] or [20, Thm. 5.14].} \tag{2.21}$$

In (2.21), ($\Rightarrow$) holds again due to finite dimension, ($\Leftarrow$) needs Brouwer's principle on invariance of domains. For $f \in \mathcal{D}$, (2.21) remains true; use (2.13) and the Δ-set in [27], but - in contrast to $f \in C^{0,1}$ - there are no tools to handle the condition effectively. Finally, also the condition

$$\ker Cf(\bar{x}) = \{0\} \tag{2.22}$$

characterizes some stability, the *local upper Lipschitz* property of f^{-1} at $\bar{x}$ which requires: There are $\mu > 0$ and neighborhoods U, V of $\bar{x}$ and $f(\bar{x})$, respectively, such that

$$f^{-1}(y) \cap U \subset \bar{x} + \mu \, \|y - f(\bar{x})\| B \quad \forall y \in V, \tag{2.23}$$

cf. [20, Lemma 3.2] or earlier [19]. In this situation, $\bar{x}$ is isolated in $f^{-1}(f(\bar{x}))$.

Needless to say, these definitions and statements are extended to $f : X \to Y$ and to multifunctions in the literature, e.g. [20], [31], [39]. They are well-known, correctly verified and do not need remakes as given at several places in terms in [14] or trivial "weakenings" like [14, Thm. 4.1].

However, even if $f \in C^{0,1}$, *computing* $Tf, \partial f$, D^*f or Cf is a hard problem which calls for exact chain rules (not only trivial ones of inclusion-type). To study primal-dual solutions of variational conditions, the product rule [20, Thm. 7.5] is helpful for Tf and Cf. The difficulties to find Tf or D^*f for the stationary point map f of parametric C^2- optimization or variational inequalities, can be seen in [21], the difficulties for Cf concerning similar models in [28].

2.2.2 Supplements concerning Cf and $\tilde{C}f$

The following properties of Cf and $\tilde{C}f$ are useful for analyzing $f \in \mathcal{D}$.

Lemma 2.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $x, u \in \mathbb{R}^n$.*
(i) If $t_k^{-1}\|f(x + t_k u) - f(x)\| \to \infty$ for certain $t_k \downarrow 0$ then it holds $Cf(x)(0) \neq \{0\}$ or $\tilde{C}f(x)(u) \neq \emptyset$.
(ii) If $Cf(x)(0) = \{0\}$ then $\tilde{C}f(x)(u) \neq \emptyset$ for all u.

Proof. (i) By continuity, certain $s_k \in (0, 1)$ satisfy $t_k^{-1} \, \|f(x + t_k s_k u) - f(x)\| = 1$. If $s_k \to 0$ (for some subsequence), then some $y \in \operatorname{bd} B$ fulfills $y \in Cf(x)(0)$ by definition. If $s_k \to \sigma > 0$ (for some subsequence), we obtain by $s_k^{-1} = (t_k s_k)^{-1}\|f(x + (t_k s_k) \, u) - f(x)\|$ that some $y \in \operatorname{bd} B$ fulfills $y\sigma^{-1} \in \tilde{C}f(x)(u)$. (ii) If, in contrary, $\tilde{C}f(x)(u) = \emptyset$ for some u, then $t_k^{-1}\|f(x + t_k u) - f(x)\| \to \infty$ holds for all $t_k \downarrow 0$. So (i) yields the assertion. □

Lemma 2.4. *If $f : \mathbb{R}^n \to \mathbb{R}$ and $f(x+u) < f(x)+c$ then $\exists t \in (0, 1)$: $\sup \tilde{C}f(x+tu)(u) < c$.*

Proof. We verify the equivalent statement: If $\sup \tilde{C}f(x+tu)(u) \geq c \; \forall t \in (0, 1)$ then $f(x+u) \geq f(x) + c$. For this reason, pick $q < c$ and put $T_q = \{t \in [0, 1] \mid f(x + tu) \geq f(x) + q\, t\}$. T_q is closed and $0 \in T_q$. Hence $s = \max T_q$ exists. We show $s = 1$ by contradiction. Otherwise, there is some $\zeta \in \tilde{C}f(x + su)(u)$ with $\zeta > q$. Accordingly there are $\varepsilon_\nu \downarrow 0$ such that

$$\zeta_\nu := \varepsilon_\nu^{-1} \, [\, f(\, (x + su) + \varepsilon_\nu u\,) - f(x + su)\,] \text{ fulfill } \zeta = \lim \zeta_\nu.$$

The latter implies $\zeta_\nu > q$ and the contradiction $\varepsilon_\nu + s \in T_q$ for large ν since

$$f(x + su + \varepsilon_\nu u) = \varepsilon_\nu \zeta_\nu + f(x + su) > \varepsilon_\nu q + f(x + su) \geq \varepsilon_\nu q + f(x) + q\, s.$$

Now $s = 1$ implies $f(x + u) \geq f(x) + q$, and the assertion follows via $q \to c$. □

Lemma 2.5. *For $f : \mathbb{R} \to \mathbb{R}$, it holds:* $\tilde{C}f(x)(u) = Cf(x)(u)\ \forall u \neq 0$,
$\ker Cf(\bar{x}) = \{0\} \Leftrightarrow \exists\ \mu > 0$ *such that* $|f(x) - f(\bar{x})| \geq \mu^{-1}\ |x - \bar{x}|$ *for x near $\bar{x}$, and*

$$\begin{array}{llll} & \text{(MR)} \quad \Leftrightarrow \quad \text{strong regularity} \quad \Leftrightarrow \quad \exists \mu > 0 \textit{ such that} \\ \textit{either} & f(y) - f(x) \geq \ \ \mu^{-1}\ (y - x) & \forall y > x : & x, y \textit{ near } \bar{x} \\ \textit{or} & f(y) - f(x) \leq -\mu^{-1}\ (y - x) & \forall y > x : & x, y \textit{ near } \bar{x}. \end{array}$$

Proof. For $u \neq 0$, the set $Cf(x)(u)$ consists of the limits of quotients

$$v_k = \frac{f(x + t_k u_k) - f(x)}{t_k} = |u_k| \frac{f(x + t_k u_k) - f(x)}{t_k |u_k|} \quad \text{ as } u_k \to u \text{ and } t_k \downarrow 0. \tag{2.24}$$

They coincide with the limits of $w_k := |u|\ \frac{f(x+\varepsilon_k\ \text{sgn}(u)) - f(x)}{\varepsilon_k}$ as $\varepsilon_k \downarrow 0$. Discussing $u > 0$ and $u < 0$ separately yields

$$Cf(x)(u) = |u|\ \tilde{C}f(x)(\text{sgn}(u)) = \tilde{C}f(x)(u) \qquad \text{if } u \neq 0. \tag{2.25}$$

Equivalences: For $x_1 < x_2$ near $\bar{x}$ and $f(x_1) = f(x_2)$, there is a local maximizer or minimizer $x \in (x_1, x_2)$. There, (2.15) cannot hold with $y' > f(x)$ and $y' < f(x)$, respectively. Hence (MR) implies (local) monotonicity. The rest is left to the reader. □

Example 2.6. *Let $f : \mathbb{R} \to \mathbb{R}$ satisfy $\bar{x} = f(\bar{x}) = 0$ and $\lim_{x \to 0,\ x \neq 0} \frac{|f(x)|}{|x|} = \infty$. Then $Cf(0)(0) \neq \{0\}$ and $Cf(0)(u) = \tilde{C}f(0)(u) = \emptyset\ \forall u \neq 0$. Hence both conditions in (2.8) [and (H1) below since $\tilde{C}f(0)(u) = \emptyset$] are violated for $Gf = Cf$.*

Proof. Assume, with no loss of generality, that $f(x)/x \to \infty$ if $x \downarrow 0$. Then there are $w_k \downarrow 0$ such that $p_k := f(w_k)/w_k \to \infty$ for $k \to \infty$. To show that already the latter implies $1 \in Cf(0)(0)$, we put $t_k = f(w_k)$ and $u_k = p_k^{-1} \to 0$. Continuity ensures $t_k = f(w_k) = p_k w_k \downarrow 0$. By $t_k u_k = t_k/p_k = w_k$ we thus obtain $\frac{f(t_k u_k)}{t_k} = \frac{f(w_k)}{t_k} = 1$. In consequence, $1 \in Cf(0)(0)$.

For $u \neq 0$, $Cf(0)(u)$ consists of the limits of $v_k = \frac{f(t_k u_k)}{t_k} = u_k \frac{f(t_k u_k)}{t_k u_k}$ as $u_k \to u$ and $t_k \downarrow 0$. Since $|v_k| \to \infty$, it follows $\emptyset = Cf(0)(u) \supset \tilde{C}f(0)(u)$. □

2.2.3 Existence of x_{k+1} and convergence in terms of f^{-1} for $Gf = Cf$

Proving (2.4): Usually, one shows first, without requiring solvability of (2.3), that

$$\forall x \text{ near } \bar{x} \text{ and any } x' \text{ satisfying (2.3), it holds } \tfrac{x' - \bar{x}}{\|x - \bar{x}\|} \to 0 \text{ as } x \to \bar{x},\ x \neq \bar{x}. \tag{2.26}$$

The solvability of the auxiliary problem

$$-f(x) \in Gf(x)(u) \qquad \text{with } x' = x + u \tag{2.27}$$

is obviously ensured (and (2.26) implies (2.4)) if $Gf(x)$ is surjective for x near $\bar{x}$, i.e., $\mathbb{R}^n = Gf(x)(\mathbb{R}^n)$. At the zero $\bar{x}$, the algorithm stops with $u = 0$, hence condition

$$\mathbb{R}^n = Gf(x)(\mathbb{R}^n) \qquad \forall x \text{ near } \bar{x},\ x \neq \bar{x} \tag{2.28}$$

is also sufficient to guarantee superlinear local convergence along with (2.26). Property (2.16) tells us that (MR) yields surjectivity for $Gf = Cf$. But (MR), throughout supposed in [14], is stronger than (2.28) - consider $f(x) = |x|$ - and could be replaced by (2.28) in [14, Thm. 3.3] without any problem. Nevertheless, solvability is based on an extra statement in [14].

Proposition 2.7. *[14, Prop. 3.2] Let f be (MR) near $\bar{x}$. Then, for all x near $\bar{x}$, the inclusion*

$$-f(x) \in Cf(x)(u) \tag{2.29}$$

admits a solution u. Furthermore, the set $S(x)$ of solutions is computed by

$$S(x) = \text{Limsup}_{\,t\downarrow 0,\ h\to\ -f(x)}\ t^{-1}\ [f^{-1}(f(x)+th) - x].$$

Proof. Again, (2.16) yields solvability. The formula is (2.2) for f, f^{-1} and $v = -f(x)$: $v \in Cf(x)(u) \Leftrightarrow\ u \in Cf^{-1}(f(x),x)(v) = \text{Limsup}_{\,t\downarrow 0,\ h\to v}\ t^{-1}\ [f^{-1}(f(x)+th)-x]$, and holds without supposing (MR). □

Remark 2.8. *We already know that, for $Gf = Cf$, the iterations (2.5) (or (2.6)) are*

$$x_{k+1}\ \in\ x_k + Cf^{-1}(f(x_k), x_k)\ (-f(x_k));\quad k = 0, 1, 2, ... \tag{2.30}$$

Using (2.17), (MR) ensures that some x_{k+1} (but not necessarily all) *satisfies*

$$\|x_{k+1} - x_k\| \le \mu\ \|f(x_k)\|. \tag{2.31}$$

When proving [14, Thm.3.4] (= Prop. 3.17 below), the authors used (2.31) for all x_{k+1} *and overlooked that $f^{-1}(f(x)+th) - x$ may contain elements with much bigger norm than $\mu t\|h\|$. In other words, their proof works only under strong regularity.*

3 Conditions for local and global convergence

Recall that $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous. Regularity of f means always regularity near the reference point $\bar{x}$. All our conditions have to hold for x near $\bar{x}$, $(x \ne \bar{x})$ only.

3.1 Local conditions in [20]

In this section, we suppose $f \in C^{0,1}$ (locally Lipschitz) if nothing else is said.

3.1.1 The conditions (CI), (CA), (CA)*

To ensure (2.4) or (2.26), the following conditions are used in [20].

$$\begin{array}{lll} (CI) & \exists c > 0: \quad \|v\| \ge c\|u\| & \forall v \in Gf(x)(u),\ u \in \mathbb{R}^n,\ x \text{ near } \bar{x}, \\ (CA) & f(x) - f(\bar{x}) + Gf(x)(u) \subset Gf(x)(u + x - \bar{x})\ + o(x-\bar{x})B & \forall u \in \mathbb{R}^n, \\ (CA)^* & f(x) - f(\bar{x}) + Gf(x)(\bar{x} - x) \subset o(x - \bar{x})B\ . & \end{array} \tag{3.1}$$

Condition (CI) requires *uniform* injectivity (2.11) of $Gf(x)$.
(CA)* requires (CA) for $u = \bar{x} - x$ only (if $Gf(x)(0) = \{0\}$) and stands for the usual type of approximation if $f \in C^1$. Condition (CA) is useful due to

Lemma 3.1. *The conditions (CI) and (CA) together imply (2.26) for any Gf in (2.3).*

Proof. . Having $0 \in f(x) + Gf(x)(u)$ where $u = x' - x$, (CA) yields $0 \in Gf(x)(u + x - \bar{x}) + v$ for some $v \in o(x - \bar{x})B$. So (CI) implies (2.26): $c\|x' - \bar{x}\| = c\|u + x - \bar{x}\| \le \|-v\| \le o(x - \bar{x})$. □

Because (CA)* looks simpler than (CA), the next statements are useful.

Remark 3.2. *If Gf has the form (2.7), condition (CI) means regularity of A and uniform boundedness of A^{-1} for all matrices $A \in M(x)$, x near $\bar{x}$. This yields $Gf(x)(0) = \{0\}$, after which (CA)* ensures (CA).*

Proof. Indeed, having $f(x) - f(\bar{x}) + A(\bar{x} - x) = o(x - \bar{x})$ from (CA)*, it follows
$f(x) - f(\bar{x}) + Au \;=\; A(x - \bar{x}) + o(x - \bar{x}) + Au \;=\; A(u + x - \bar{x}) + o(x - \bar{x}).$ □

Theorem 3.3. *[20, Thm. 10.8]. It holds (CA)* $\Leftrightarrow$ (CA) if Gf is given by (2.7) or coincides with Cf, Tf or ∂f.*

The proof is more involved only if $Gf = Cf$. In view of NM (2.3) for $f \in C^{0,1}$, Lemma 3.1 and Thm. 3.3 ensure

Corollary 3.4. *Suppose (CI), (CA)* and surjectivity (2.28). Then (2.4) holds true.*

If Gf has the simpler form (2.7), already (CI) implies (2.28). Thus already Lemma 3.1 and Rem. 3.2 show that (CI) and (CA)* ensure (2.4). This statement can be sharpened by saying that (CA)* is even necessary for (2.4).

Lemma 3.5. *[20, Lemma10.1]. For Gf (2.7), suppose that all $A \in M(x)$ as well as A^{-1} are uniformly bounded for x near $\bar{x}$. Then, method (2.3) is locally superlinear convergent $\Leftrightarrow$ Gf satisfies (CA)*. In this case, the $o-$functions in (CA)* and (2.4) differ by a constant only.*

Hence quadratic approximation in (CA)* , $|o(x - \bar{x})| \le K\|x - \bar{x}\|^2$, yields quadratic order of convergence as well. For the other settings of Gf, the role of (CI) under the viewpoint of regularity and the necessity of (CA)* for superlinear convergence (2.4) was characterized by

Theorem 3.6. *[20, Thm. 10.9].*
(a) Let $Gf = Tf$. Then (CI) holds true $\Leftrightarrow$ f is strogly regular near $\bar{x}$. Having (CI), condition (CA) is necessary and sufficient for (2.4).*
(b) Let $Gf = \partial f$. Then (CI) holds true $\Leftrightarrow$ all $A \in \partial f(\bar{x})$ are non-singular. This might be stronger than strong regularity.
(c) Let $Gf = Cf$. Then (CI) holds at $x = \bar{x}$ $\Leftrightarrow$ f^{-1} is locally upper Lipschitz at $\bar{x}$.
(d) Let $Gf(x)(u) = f'(x; u)$ provided that $f'(x; u)$ exists near $\bar{x}$. Then, under strong regularity, (CA) is necessary and sufficient for (2.4). Under (MR), condition (CI) is satisfied.*

The proofs of (b) and (c) need only the conditions of section 2.2.1 while (a) and (d) require more effort. In particular (d), (MR) $\Rightarrow$ (CI) is deep and applies the powerful result of [10], which is included in [20] as Thm. 5.12. The formulation of Thm. 3.6 in [20] is a bit more involved since local convergence of the following algorithm, where $\alpha > 0$ is some error-constant, has been also taken into account:

$$\text{ALG}(\alpha): \quad \text{Find } x_{k+1} \text{ such that } \; \emptyset \neq \alpha\|f(x_k)\|B \;\cap\; [f(x_k) + Gf(x_k)(x_{k+1} - x_k)]. \quad (3.2)$$

For $\alpha = 0$, this is algorithm (2.5), and (2.4) yields solvability for any $\alpha \ge 0$. Solution estimates and solvability of (3.2) are collected in [20, Thm. 10.7] and [24]. The "Inexact Nonsmooth Newton Method" 7.2.6 in [8] is exactly algorithm ALG(α), specified to Gf (2.7).

Having *any function* $f : \mathbb{R}^n \to \mathbb{R}^n$, Rem. 3.2 and the Lemmas 3.1, 3.5 remain true without changing any proof. The restriction to $f \in C^{0,1}$ was motivated after [20, Lemma 10.1] (= Lemma 3.5) by the following

Remark 3.7. *"In the current context, the function f may be arbitrary (even for normed spaces) as long as $M(x)$ consists of linear bijections. Nevertheless, we will suppose that f is locally Lipschitz near $\bar{x}$: This is justified by two reasons:*

(i) If f is only continuous, we cannot suggest any practically relevant definition for $M(x)$.

(ii) Having uniformly bounded (by K) matrices, the convergence condition (2.4) implies that f satisfies a pointwise Lipschitz condition at $\bar{x}$, namely

$$\|f(x) - f(\bar{x})\| \leq 2K\|x - \bar{x}\| \text{ for small } \|x - \bar{x}\|. \tag{3.3}$$

Since the solution $\bar{x}$ is unknown, our assumptions should hold for all $\bar{x}$ near the solution. Then, $f \in C^{0,1}$ (near the solution) follows necessarily from (3.3)."

Requiring (CA)* for all $\bar{x}$ near the solution leads to the *slantly differentiable functions* in [12].

3.1.2 Two types of semismoothness for $f \in C^{0,1}$

With $Gf = \partial f$, condition (CA)* defines Mifflin's [29] *semismoothness* (original for $f(x) \in \mathbb{R}$ and analogue for $f(x) \in \mathbb{R}^n$) which is supposed in many papers, e.g., in [32], [33], [35], [36] or [8], where the reader finds more references. Because the existence of directional derivatives $f'(\bar{x}, u)$ at $\bar{x}$ follows easily from (CA)*, this existence is sometimes already supposed in order to define semismoothness at $\bar{x}$ equivalently via $Au - f'(\bar{x}; u) = o(u)\ \forall A \in \partial f(\bar{x} + u)(u)$.

In other papers, e.g. in [8] and [14], semismoothness at $\bar{x}$ requires per definition directional differentiability *also for x near $\bar{x}$,* which makes sense from the practical point of view since the zero $\bar{x}$ is unknown. However, directional differentiability *near* $\bar{x}$ is not important for the convergence (2.4) with $Gf = \partial f$ as the (necessary and sufficient) conditions (CI) and (CA)* show.

The example [14, Ex. 4.11] is a strongly regular $C^{0,1}$ function satisfying (CI) and (CA)* without being directionally differentiable *near* $\bar{x}$. So it is semismooth in the first (Mifflin's) sense or in the sense of [36] and not semismooth in the stronger sense of [14] due to an unnecessary (for convergence (2.4)) requirement in the definition. Hence it is far from an example the (more general) semismoothness-theory cannot be applied to.

In view of methods which use $\partial f, \partial_B f$, Cf or $f'(x; u)$ as Gf, a further remark is useful.

3.1.3 Variation of the generalized derivatives

Remark 3.8. *(cf. [20, § 10.1.1]*
(i) If Gf (2.7) satisfies (CI) and (CA) for a mapping $M = M(x)$ then also for each mapping M' satisfying (for the unit ball of (n, n) matrices)*

$$\emptyset \neq M'(x) \subset M(x) + O(x - \bar{x})B_{n,n}\ , \quad \text{where } O(x - \bar{x}) \to 0 \text{ as } x \to \bar{x}. \tag{3.4}$$

(ii) Methods with different mappings $G_1 f$ and $G_2 f$ can be directly compared whenever $\emptyset \neq G_1 f(x)(u) \subset G_2 f(x)(u)$. (CI) and (CA) for $G_2 f$ imply (CI) and (CA)* for $G_1 f$.*
(iii) Again evident for Gf (2.7): (CA) for M implies (CA)* for $M' = \operatorname{conv} M$.*

Consequently, the method based on $G_1 f$ in (ii) inherits the convergence (2.26) from $G_2 f$ whenever $G_2 f$ satisfies (CI) and (CA)*. The same holds, in particular, for Gf (2.7) if $\emptyset \neq M' \subset M$.

These observations explain completely the relation between NM based on $M'(x) = \partial_B f(x)$ and $M(x) = \partial f(x) = \operatorname{conv} \partial_B f(x)$, in particular the trivial statement [14, Thm. 5.1] where strong regularity of $f \in C^{0,1}$ is hidden under (MR) and 1-to-1.

They also show the relation between methods based on $G_1 f(x)(u) = \{f'(x; u)\}$ (if directional derivatives exist) or $G_1 f(x)(u) = Cf(x)(u)$ on the one hand and Clarke's Jacobians $G_2 f(x)(u) = \partial f(x)(u)$ or $G_2 f(x)(u) = Tf(x)(u)$ on the other hand. For these particular settings of G_1 and G_2, the method assigned to $G_1 f$ also inherits the stronger convergence (2.4) since solvability follow via (CI) $\Rightarrow$ $\ker Tf(\bar{x}) = \{0\} \Rightarrow$ strong regularity $\Rightarrow$ (MR) $\Rightarrow$ surjectivity for Cf. For $G_1 f(x)(u) = \{f'(x; u)\}$, solvability is implied by $\{f'(x; u)\} = Cf(x)(u)$ if $f \in C^{0,1}$. More details presents Thm. 3.6 (d).

Remark 3.9. *For Gf (2.7), the conditions (CI) and (CA)* have been used in [8, Sect. 7.2] to define so-called* (regular) Newton approximation schemes.

Mappings $M = M(x)$ such that Gf (2.7) satisfies (CA) are called* Newton maps *in [20, § 6.4.2]. They satisfy usual chain rules for composed functions, exist for "locPC1 functions", do not necessarily coincide with Tf or ∂f, but can replace these mappings for studying methods which use directional derivatives or Cf by "inheritance".*

Condition (3.4) describes possible approximations or, e.g. if $M'(x) = M(x) + \|f(x)\|E$, some "regularization".

3.2 Local and global conditions in [14]

Now we suppose throughout that $Gf = Cf$. Thus (2.29) is the crucial Newton inclusion. Again, regularity of f means always regularity near the reference point $\bar{x}$.

3.2.1 Convergence under (H1) and (H2)

In [14], the following conditions have been imposed (we write u for d and f for H).

(H1) Exist $c > 0$ and neighborhoods Ω, V of $\bar{x}$ and 0_n, respectively, satisfying:

$$\text{If } x \in \Omega \text{ and } -f(x) \in Cf(x)(u), \text{ then}$$
$$\forall z \in V \ \exists w \in \tilde{C}f(x)(z) \text{ with } c\|u - z\| \le \|w + f(x)\| + o(x - \bar{x})$$

and, improving a typing error in [14],

$$\text{(H2)} \qquad \|f(x) - f(\bar{x}) + w\| \le o(x - \bar{x}) \quad \forall w \in \tilde{C}f(x)(\bar{x} - x).$$

If $f \in C^{0,1}$, (H2) is (CA)*. Again, it holds, like Lemma 3.1,

Lemma 3.10. *(H1) and (H2) together imply (2.26).*

Proof. Assume (as in [14]) for x near $\bar{x}$, that (2.29) holds true, $x' = x + u$, and put $z = \bar{x} - x$. By (H1), some $w \in \tilde{C}f(x)(\bar{x} - x)$ fulfills $c\ \|x' - \bar{x}\| = c\ \|u - z\| \le \|w + f(x)\| + o(x - \bar{x})$. (H2) tells us $\|f(x) + w\| \le o(x - \bar{x})$. Hence $c\ \|x' - \bar{x}\| \le 2\ o(x - \bar{x})$ implies (2.26). □

The Lemma ensures directly the analogon of Corollary 3.4.

Theorem 3.11. *[14, Thm. 3.3]. Under (H1), (H2) and (MR), convergence (2.4) holds true.*

Proof. Indeed, (2.26) ⇒ (2.4) follows from(MR) which could be replaced by (2.28). □

The proof in [14] applies Prop. 2.7.

Remark 3.12. *(i) By the proof of Lemma 3.10, $\tilde{C}f(x)$ could be* any multifunction *satisfying (H1) and (H2), and one needs (H1) for $z = \bar{x} - x$ only. Since (H1) is only used for showing Thm. 3.11, there is no reason to involve other directions z. (ii) For $f \in C^{0,1}$, Thm. 3.11 is just Corollary 3.4 with condition (H1) in place of (CI).*

3.2.2 Analysing (H1) and (H2)

The conditions (H1) and (H2) bite each other: (H1) requires that $\tilde{C}f(x)(z)$ is big, (H2) claims that $\tilde{C}f(x)(\bar{x} - x)$ is small. Thus passing to smaller mappings as in Rem. 3.8 is impossible. Since the technical condition (H1) compares the usually different mappings $\tilde{C}f \subset Cf$, it is hard to find any sufficient condition for (H1) if $f \in \mathcal{D}$. In particular, (H1) fails for all f in example 2.6 where $\tilde{C}f(\bar{x})(z) = \emptyset \ \forall z \ne 0$.

The only sufficient condition for (H1) and $f \in \mathcal{D}$ requires *directional boundedness*

$$\limsup_{t \downarrow 0} \ t^{-1} \|f(x + tu) - f(x)\| < \infty \quad \forall x \text{ near } \bar{x} \text{ and } u \in \mathbb{R}^n. \tag{3.5}$$

Proposition 3.13. *[14, Prop. 4.4] (H1) holds true if f satisfies (3.5) and is strongly regular.*

Remark 3.14. *Under these assumptions, condition (CI) holds for Cf, too. Hence, for strongly regular $f \in C^{0,1}$, (H1) and (H2) ensure the well-known sufficient convergence conditions (CI) and (CA)*.*

Proof. (3.5) ensures $\emptyset \neq \tilde{C}f(x)(u) \subset Cf(x)(u)$. Using $v \in Cf(x)(u) \;\Leftrightarrow\; u \in (Cf^{-1})(f(x))(v)$ and applying (2.14) to $f^{-1} \in C^{0,1}$, (CI) follows from $\|u\| \leq L\|v\|$. $\square$

Remark 3.15. *For strongly regular* real *functions $f \in \mathcal{D}$, both Prop. 3.13 and condition (3.5) are completely useless in view of* NM*: If such f satisfies (3.5) at $x = \bar{x}$,* NM cannot superlinearly converge; *cf. Thm. 4.1.*

Condition (H2)

In [14], there is no sufficient condition for (H2) if $f \in \mathcal{D}$. Hence all sufficient conditions for (H1) *and* (H2) concern only strongly regular $f \in C^{0,1}$. Moreover, even semismoothness is required, cf. [14, Prop. 4.8].

Next we turn to the only function $f \in \mathcal{D}$ in [14] which satisfies (H2). If also (H1) and (MR) would hold, the paper had at least one justification by one example. But these conditions cannot hold together (by Thm. 3.11) since NM does not superlinearly converge.

Example 3.16. *(= Example 4.10 in [14]) The interesting two-dimensional function*

$$f(x) = \begin{pmatrix} x_2\sqrt{|x_1| + |x_2|^3} \\ x_1 \end{pmatrix}.$$

belongs to $\mathcal{D}$ and fulfills (H2) at $\bar{x} = 0$ as shown in [14]. To check convergence of NM, let $x_1 > 0, x_2 > 0$. The derivatives Df and $(Df)^{-1}$ there exist, and

$$Df(x)^{-1} = \begin{pmatrix} 0 & 1 \\ 2\frac{\sqrt{|x_1|+|x_2|^3}}{2|x_1|+5|x_2|^3} & -\frac{x_2}{2|x_1|+5|x_2|^3} \end{pmatrix}.$$

Newton steps at x define x' with $x' - x = -Df(x)^{-1}\begin{pmatrix} x_2\sqrt{|x_1| + |x_2|^3} \\ x_1 \end{pmatrix}$, *i.e.,*

$$x' - x = -\begin{pmatrix} x_1 \\ 2x_2\sqrt{|x_1| + |x_2|^3}\;\frac{\sqrt{|x_1|+|x_2|^3}}{2|x_1|+5|x_2|^3} - x_1\;\frac{x_2}{2|x_1|+5|x_2|^3} \end{pmatrix}. \tag{3.6}$$

Next take small positive variables, such that $2x_1 = x_2 - 5x_2^3$. This ensures with max-norm

$$x_2^3 = \frac{x_2 - 2x_1}{5}, \quad x_2 > 2x_1 \;\; i.e., \;\; 3x_2 - x_1 > 5x_1\;, \quad \text{as well as} \quad \|x\| \leq 4x_1\;, \tag{3.7}$$

and (3.6) implies

$$x' = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ \frac{x_1}{5} + \frac{2x_2}{5} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{x_1}{5} + \frac{3x_2}{5} \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 0 \\ 3x_2 - x_1 \end{pmatrix}.$$

Now (3.7) yields $\|x'\| \geq x_1 \geq \frac{\|x\|}{4}$. Thus local superlinear convergence is violated.

3.2.3 The Kantorovich-type statement

for continuous f requires a new set of hypotheses and asserts for method (2.5) with $Gf = Cf$,

Proposition 3.17. *[14, Thm 3.4] Let $x_0 \in \mathbb{R}^n$ and $r > 0$ be given, such that for all $x \in \Omega := x_0 + rB$ (uniformly) the following holds:*
(a) f is (MR) near x with the same modulus μ.
(b) $\forall \varepsilon > 0\ \exists \delta > 0$ such that $\|w\| \le \varepsilon$ if $w \in Cf(x)(z)$ and $\|z\| \le \delta$.
(c) For some $\alpha \in (0,\ \mu^{-1})$ it holds $\mu\,\|f(x_0)\| \le r(1-\alpha\mu)$ and

$$\|f(y) - f(x) - v\| \le \alpha\|x - y\| \quad \text{if } v \in Cf(x)(y - x) \text{ and } x, y \in \Omega. \tag{3.8}$$

Then the sequence x_k is well defined, remains in Ω and converges to some zero $\bar{x}$ of f where $\|x_k - \bar{x}\| \le \frac{\alpha\mu}{1-\alpha\mu}\|x_k - x_{k-1}\|$.

Condition (3.8) in [14] begins with $\|f(x) - f(y) - v\|$ which is a mistake. The proof of [14, Thm 3.4] is wrong: our Remark 2.8 says why. Hence Prop. 3.17 is not proven. Nevertheless we again investigate the suppositions which seem to permit that certain sets $Cf(x)(y - x)$ in (3.8) are empty (confusing the interested reader perfectly).

- Condition (b) requires $Cf(x)(z) \subset K\|z\|B$ for some constant K (put $\varepsilon = 1$ and $K = 1/\delta$). Hence $Cf(x)(0) = \{0\}$. By Lemma 2.3, so all sets $\tilde{C}f(x)(z) \subset Cf(x)(z)$ are non-empty, and elements $v \in Cf(x)(y - x)$ exist in (3.8). For that reason, the triangle inequality $\|f(y) - f(x)\| \le \|f(y) - f(x) - v\| + \|v\| \le (\alpha + K)\|x - y\|$ yields

  Remark 3.18. *The conditions (b) and (3.8) imply that f is globally Lipschitz on Ω. Thus, the proposition (true or not) says nothing for $f \in \mathcal{D}$.*

- Because $Cf(x)$ is positively homogeneous, (3.8) requires for $x \ne y$ in Ω

$$\left\| \frac{f(y) - f(x)}{\|y - x\|} - w \right\| \le \alpha \qquad \forall w \in Cf(x)\left(\frac{y - x}{\|y - x\|}\right). \tag{3.9}$$

  Hence, up to error α, all difference quotients $\frac{f(y)-f(x)}{\|y-x\|}$ - for $\|y - x\|$ big or not - have to coincide with arbitrary $w \in Cf(x)(\frac{y-x}{\|y-x\|})$. This condition is strong even for $f \in C^2$ where it requires small (compared with r) second derivatives on Ω. It also claims $\operatorname{diam} Cf(x)(\frac{y-x}{\|y-x\|}) \le 2\alpha$.

- In consequence, already for real *piecewise linear homeomorphisms* f, the assumptions of Prop. 3.17 are violated whenever $\operatorname{int}\Omega$ contains a "sufficiently big kink" of f. Put, e.g., $f(\xi) = 2\xi + |\xi|$ and suppose $0 \in \operatorname{int}\Omega$ in order to see (setting $y = -x < 0$) that the hypotheses of Prop. 3.17 are not satisfied with $\alpha \in (0, \mu^{-1}) = (0, 1)$.

4 Newton-convergence for real functions in $\mathcal{D}$

In this section, we study NM with $Gf = Cf$ for *real* (MR) functions and derive necessary conditions for convergence (2.4). By (MR) and Lemma 2.5, f is strongly monotone. We may assume that

$$\bar{x} = f(\bar{x}) = 0 \text{ and } f(y) - f(x) \ge \mu(y - x)\ \forall y > x \text{ near } 0,\ \mu > 0. \tag{4.1}$$

Lemma 2.5 then also ensures $Cf(x)(u) = \tilde{C}f(x)(u)$ for $u \ne 0$, and

$$0 < \mu u \le \inf \tilde{C}f(x)(u)\ \forall u > 0, \qquad 0 > \mu u \ge \sup \tilde{C}f(x)(u)\ \forall u < 0. \tag{4.2}$$

Here and below, all arguments of f will be taken close to 0 without saying it explicitly. The convergence (2.4) requires, if $x \neq 0$ and $x \to 0$,

$$-f(x) \in Cf(x)(u) \quad \Rightarrow \quad \frac{x'}{x} \to 0, \ \frac{x'-x}{x} = \frac{u}{x} \to -1 \quad (x' = x+u). \tag{4.3}$$

Setting $\frac{u}{x} = -1+\beta$, so (2.4) means exactly $\beta = \beta(x,u) \to 0$ uniformly as $x \to 0$. For $x > 0$, we have $u < 0$ and $Cf(x)(u) = |u|Cf(x)(-1) = (1-\beta)\ x\ Cf(x)(-1)$, thus

$$-f(x) \in Cf(x)(u) \ \Leftrightarrow \ -\frac{f(x)}{x} \in (1-\beta)\ Cf(x)(-1). \tag{4.4}$$

4.1 Violation of superlinear convergence for $f \in \mathcal{D}$

Next we show for *real*, (MR) functions in $\mathcal{D}$: If superlinear convergence (2.4) holds true, then directional boundedness (3.5) at $\bar{x}$ is not satisfied.

Theorem 4.1. *Let $f \in \mathcal{D}$ be a real (MR) function with* $\limsup_{t\downarrow 0} \frac{|f(\bar{x}+tz)-f(\bar{x})|}{t} < \infty \ \forall z \in \mathbb{R}$. *Then local superlinear convergence (2.4) cannot hold.*

Proof. We use the above preparations. Because of $f \in \mathcal{D}$ there are $x_k < y_k$ which tend to 0 and satisfy $C_k := \frac{f(y_k)-f(x_k)}{y_k-x_k} \to \infty$. Assume first $x_k < 0 < y_k$ (for some subsequence). Writing

$$C_k = \frac{f(y_k)-f(0)}{y_k-x_k} + \frac{f(0)-f(x_k)}{y_k-x_k} = A_k + B_k,$$

A_k or B_k has to diverge. If $A_k \to \infty$, then $y_k - x_k > y_k > 0$ and $A_k \le \frac{f(y_k)-f(0)}{y_k} \to \infty$. If $B_k \to \infty$, then $y_k - x_k > -x_k > 0$ and $B_k \le \frac{f(0)-f(x_k)}{-x_k} \to \infty$. Both situations violate the lim sup-condition. Thus, it holds $0 < x_k < y_k$ or $x_k < y_k < 0$. We consider the first case.
Let $u_k = x_k - y_k < 0$ and $c_k = \frac{1}{2}C_k(x_k - y_k)$.
Then $f(y_k+u_k) - f(y_k) = f(x_k) - f(y_k) = C_k(x_k - y_k) < c_k < 0$. By Lemma 2.4, some $t_k \in (0,1)$ fulfills $\sup \tilde{C}f(y_k + t_k u_k)(u_k) < c_k$. Put $\theta_k = y_k + t_k u_k$. Lemma 2.5 ensures $Cf(\theta_k)(u_k) = \tilde{C}f(\theta_k)(u_k)$ as well as

$$\sup Cf(\theta_k)(-1) = \sup Cf(\theta_k)(u_k/|u_k|) < \frac{c_k}{|y_k - x_k|} = -\tfrac{1}{2}C_k \to -\infty. \tag{4.5}$$

Thus there are $\theta_k \downarrow 0$ with $\sup Cf(\theta_k)(-1) \le -\frac{C_k}{2} \to -\infty$. By our assumption, all $f(\theta_k)/\theta_k$ remain bounded. This yields a contradiction to (4.4) since all $w \in (1-\beta)\ Cf(\theta_k)(-1)$ diverge to $-\infty$. The situation $x_k < y_k < 0$ can be similarly handled. ☐

4.2 Superlinear convergence for real, "almost C^1-functions"

The subsequent functions are continuously differentiable near x for $x \neq \bar{x}$, strongly regular near $\bar{x} = 0$ with $f(\bar{x}) = 0$ and have the limit-property of example 2.6. Thus (2.8) and (H1) are violated. Below, $f'(x)$ denotes the usual derivative of f at x.

For $f'(x) \neq 0$, the next Newton iterate is $x' = x - \frac{f(x)}{f'(x)}$ if $x \neq 0$. To study superlinear convergence we define

$$O_1(x) = \frac{x'}{x} = 1 - \frac{f(x)}{xf'(x)}\ , \tag{4.6}$$

and for condition (H2), which is $f(x) - f(0) - f'(x)x = o(x)$, the function

$$O_2(x) = \frac{f(x)}{x} - f'(x). \tag{4.7}$$

Hence the crucial conditions require equivalently $O_1, O_2 \to 0$ as $x \to 0$. Because of

$$-O_1(x) \;=\; \frac{f(x)}{xf'(x)} - 1 \;=\; \frac{O_2(x)}{f'(x)} \tag{4.8}$$

the $O(.)$ functions are closely connected and, if (H2) is true and $|f'(x)| \to 0$ is excluded, the superlinear local convergence $O_1(x) \to 0$ follows automatically.
First we check all imposed conditions for the simplest functions $f \in \mathcal{D}$, which also motivate our question (vii) of the introduction.

Example 4.2. *Let* $f(x) = \operatorname{sgn}(x)\,|x|^q$, $0 < q < 1$. *Obviously, we have*
$x > 0 \;\Rightarrow\; f = x^q, \quad f' = qx^{q-1}, \quad x' = x - f/f' = x - \frac{1}{q}x.$
Convergence (2.4) claims for $x \downarrow 0$*:* $O_1(x) = x'/x = 1 - \frac{1}{q} \to 0$ *which is impossible. Condition (H2) requires* $O_2(x) := \frac{f(x)}{x} - f'(x) \to 0$ *if* $x \downarrow 0$ *and fails to hold since* $O_2 = (1-q)x^{q-1} \to \infty$.

The following strongly regular examples indicate that, nevertheless, NM may superlinear converge for $f \in \mathcal{D}$. The examples also show that condition (H2) which, for $f \in C^{0,1}$, coincides with (CA)* and is crucial due to Lemma 3.5, may hold or not, in this situation.

Example 4.3. *Superlinear local convergence, though (H2) is violated.*

$$f(x) = \begin{cases} x\,(1 - \ln x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -f(-x) & \text{if } x < 0. \end{cases}$$

Evidently, f *is continuous and, for* $x > 0$*, it holds* $f' = -\ln x$ *and* $x' = x - \frac{x(1-\ln x)}{-\ln x} = x + \frac{x}{\ln x} - x = \frac{x}{\ln x}$. *This implies* $O_1(x) \to 0$ *due to*

$$O_1 = \; 1 - \frac{f(x)}{xf'(x)} = 1 - \frac{x\,(1-\ln x)}{-x\ln x} = 1 - (-\frac{1}{\ln x} + 1) = \frac{1}{\ln x},$$

and (H2) fails due to $O_2(x) \;=\; \frac{x\,(1-\ln x)}{x} + \ln x \;\equiv\; 1.$

Example 4.4. *Superlinear local convergence* and *(H2) hold true.*

$$f(x) = \begin{cases} x\,(\,1 + \ln(-\ln x)\,) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -f(-x) & \text{if } x < 0. \end{cases}$$

Consider small $x > 0$ *which yields* $f > 0$ *and, for* $x \downarrow 0$,

$$f' = (\,1 + \ln(-\ln x)\,) + \; x\,(\frac{1}{-\ln x}\,\frac{1}{-x}) \;=\; 1 + \ln(-\ln x) \;+ \frac{1}{\ln x} \to \infty.$$

$$O_2 = \frac{f}{x} - f' = (\,1 + \ln(-\ln x)\,) - (\,1 + \ln(-\ln x) \;+ \frac{1}{\ln x}) = -\frac{1}{\ln x} \to +0.$$

$$O_1 = 1 - \frac{f}{xf'} = 1 - \frac{1 + \ln(-\ln x)}{1 + \ln(-\ln x) \;+ \frac{1}{\ln x}} = \frac{\frac{1}{\ln x}}{1 + \ln(-\ln x) \;+ \frac{1}{\ln x}} \to -0.$$

Similarly, negative x *can be handled. Thus the assertion is verified.*

5 Summary

The statements and tools of [14] nowhere establish convergence of NM for $f \in \mathcal{D}$. Though certain functions $f \in \mathcal{D}$ allow the application of contingent derivatives in Newton's method, it looks hard to find any function $f \in \mathcal{D}$ which satisfies all hypotheses for statement [14, Thm. 3.3]. This is also true when (MR) is replaced by the weaker condition (2.28). Moreover, there is no reason for optimism when searching such real f because of Thm. 4.1. For the (possibly incorrect) Kantorovich-type statement [14, Thm. 3.4] the Lipschitz-property of f on the crucial set Ω is always necessary in order to satisfy the hypotheses.

Additional references: Newton's method for continuous functions is the subject in [16, 17](1998). The conditions (CI), (CA), (CA)* in (3.1), many of the mentioned statements including ALG(α), Thm. 3.3 and parts of Thm. 3.6 appeared first in [24, 25](1992) where also relations to point-based approximations [38] have been discussed. [25, 27] also contain criteria for strong stability of KKT-points to optimization problems with original functions having $C^{0,1}$ derivatives. Extensions to NM for multifunctions and graph-approximations can be found in [26](1995). For $f \in C^{0,1}$, the convergence theory based on contingent derivatives was recently studied (locally and globally) in [2, 3, 4]. Quasi-NM for PC1-equations are the subject of the pioneering paper [22](1986). For $C^{0,1}$- equations in Hilbert spaces (hence without semismoothness), even update formulas are examined in [11](1987). Nonsmooth Newton methods in function spaces have been studied in [18], [40, 41] and [12, 13] (after 2000).

Acknowledgment
I like to thank my friend and colleague Diethard Klatte for many fruitful discussions and help concerning this paper and for checking the examples, too (nevertheless, if some of them is wrong, so this is my error).

References

[1] J.-P. Aubin and I. Ekeland. Applied Nonlinear Analysis. Wiley, New York, 1984.

[2] S. Bütikofer. *Globalizing a nonsmooth Newton method via nonmonotone path search.* Mathematical Methods of OR, (2008) DOI: 10.1007/s00186-008-0219-8.

[3] S. Bütikofer. *Generalized Newton type methods for nonsmooth equations in optimization and complemenatarity problems.* Dissertation Doctor of Sciences, ETH Zurich (2008).

[4] S. Bütikofer and D. Klatte. *A nonsmooth Newton method with path search and its use in solving $C^{1,1}$ programs and semi-infinite problems.* SIAM Journal on Optimization, 20: 2381–2412 (2010).

[5] F.H. Clarke. *On the inverse function theorem.* Pacific J. of Mathematics, 64: 97–102 (1976).

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

[7] R. Cominetti. *Metric regularity, tangent sets and second-order optimality conditions.* Applied Mathematics and Optimization, 21: 265–287 (1990).

[8] F. Facchinei and J.-S. Pang. Finite-Dimensional Variational Inequalities and Complementary Problems, Vol II. Springer, New York, 2003.

[9] H. Frankowska. *An open mapping principle for set-valued maps* J. of Math. Analysis and Appl. 127: 172–180 (1987).

[10] P. Fusek. *Isolated zeros of Lipschitzian metrically regular R^n functions. Optimization*, 49: 425–446 (2001).

[11] A. Griewank. *The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces.* SIAM Journal on Numerical Analysis 24: 684-705 (1987)

[12] M. Hintermueller, K. Ito, K. Kunisch. *The primal-dual active set strategy as a semismooth Newton method.* SIAM J. Optimiz. 13, 3: 865–888, (2003).

[13] M. Hintermueller, J.C. de los Reyes. *A Duality-Based Semismooth Newton Framework for Solving Variational Inequalities of the Second Kind.* Interfaces and Free Boundaries 13: 437–462 (2011).

[14] T. Hoheisel, C. Kanzow, B. S. Mordukhovich, H. Phan. *Generalized Newton's method based on graphical derivatives.* Nonlinear Analysis 75 (3): 1324-1340 (2012).

[15] A.D. Ioffe. *Metric regularity and subdifferential calculus.* Russ. Mathematical Surveys, 55: 501–558, (2000).

[16] V. Jeyakumar and H. D. Tuan, *Approximate Jacobian based nonsmooth Newton methods: convergence analysis.* Computational Optimization and Applications (1998).

[17] V. Jeyakumar and D. T. Luc, *Approximate Jacobian matrices for nonsmooth continuous maps and C1-Optimization* SIAM J. Control Optim., 36(5): 1815-1832, (1998).

[18] I. Kazufumi and K. Kunisch *Newton's Method for a Class of Weakly Singular Optimal Control Problems.* SIAM J. Optim. 10(3): 896–916 (2000).

[19] A. King and R.T. Rockafellar. *Sensitivity analysis for nonsmooth generalized equations.* Mathematical Programming, 55: 341–364, (1992).

[20] D. Klatte and B. Kummer. Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications. Ser. Nonconvex Optimization and Its Applications. Kluwer Academic Publ., Dordrecht-Boston-London, 2002.

[21] D. Klatte and B. Kummer. *Strong Lipschitz Stability of Stationary Solutions for Nonlinear Programs and Variational Inequalities.* SIAM Optimization, 16: 96–119, (2005).

[22] M. Kojima and S. Shindo. *Extension of Newton and quasi-Newton methods to systems of PC1-equations.* J. Oper. Res. Soc. Japan 29: 352–375 (1986).

[23] B. Kummer. *Newton's method for non-differentiable functions.* In J. Guddat et al. eds. Advances in Math. Optimization. Akademie Verlag Berlin, Ser. Math. Res. 45: 114–125 (1988).

[24] B. Kummer. *Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis.* In W. Oettli and D. Pallaschke, eds. Advances in Optimization. Springer, Berlin 1992, 171–194.

[25] B. Kummer. *On stability and Newton-type methods for Lipschitzian equations with applications to optimization problems (plenary lecture).* In P. Kall ed. System Modelling and Optimization, Proceedings of the 15th IFIP-Conference, Zurich Sept. 2-6, 1991. Springer, 1992, 3–19.

[26] B. Kummer. *Approximation of multifunctions and superlinear convergence.* in Recent developments in optimization. R. Durier and C. Michelot eds. Lecture Notes in Economics and Mathematical Systems 429. 243–251, Springer, 1995.

[27] B. Kummer. *Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ optimization.* J. of Optimization Theory and Applications, 70: 559–580, (1991).

[28] A. B. Levy, R. A. Poliquin, R. T. Rockafellar. *Stability of Locally Optimal Solutions.* SIAM J. Optim. 10 (2): 580–604 (2000).

[29] R. Mifflin. *Semismooth and semiconvex functions in constrained optimization.* SIAM Journal on Control and Optimization, 15: 957–972, (1977).

[30] B.S. Mordukhovich. *Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions.* Transactions of the American Mathematical Society 340: 1–35 (1993).

[31] B.S. Mordukhovich. Variational Analysis and generalized differentiation. Vol. I: Basic Theory, Vol II: Applications Springer, Berlin, 2005.

[32] J.S. Pang. *Newton's method for B-differentiable equations.* Mathematics of OR, 15 (2): 311–341 (1990).

[33] J.S. Pang, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems.* Mathematical Programming 51: 101-131 (1991).

[34] J.-P. Penot. *Metric regularity, openness and Lipschitz behavior of multifunctions. Nonlinear Analysis*, 13: 629–643, (1989).

[35] L. Qi. *Convergence analysis of some algorithms for solving nonsmooth equations.* Working paper. School of Mathematics, Univ. of New South Wales (1991); Math. Oper. Res. 18: 227–244 (1993).

[36] L. Qi and J. Sun. *A nonsmooth version of Newton's method.* Mathematical Programming, 58: 353–367 (1993).

[37] S.M. Robinson. *Strongly regular generalized equations.* Mathematics of Operations Research, 5: 43–62, (1980).

[38] S.M. Robinson. *Newton's method for a class of nonsmooth functions.* Set-Valued Analysis , 2: 291–305, (1994); and Working Paper Univ. of Wisconsin-Madison, Dptm. of Industrial Engineering (1988).

[39] R. T. Rockafellar and R. J.-B. Wets. Variational Analysis. Springer, Berlin, 1998.

[40] M. Ulbrich. *Nonsmooth Newton-like Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces.* Technische Universitaet Muenchen, Fakultaet fuer Mathematik, June 2001, revised February 2002, Habilitationsschrift.

[41] M. Ulbrich. *Semismooth Newton methods for operator equations in function spaces.* SIAM J. Optim. 13: 805-841, (2003).