Noname manuscript No.
(will be inserted by the editor)

Approximations and Generalized Newton Methods

Diethard Klatte · Bernd Kummer

Submitted 29 January 2016 / Revised: 9 September 2016, 1 February 2017, 24 August 2017

Abstract We present approaches to (generalized) Newton methods in the framework of generalized equations $0 \in f(x)+M(x)$, where f is a function and M is a multifunction. The Newton steps are defined by approximations $\hat{f}$ of f and the solutions of $0 \in \hat{f}(x)+M(x)$. We give a unified view of the local convergence analysis of such methods by connecting a certain type of approximation with the desired kind of convergence and different regularity conditions for $f+M$. Our paper is, on the one hand, thought as a survey of crucial parts of the topic, where we mainly use concepts and results of the monograph [31]. On the other hand, we present original results and new features. They concern the extension of convergence results via Newton maps [31,38] from equations to generalized equations both for linear and nonlinear approximations $\hat{f}$, and relations between semi-smoothness, Newton maps and directional differentiability of f. We give a Kantorovich-type statement, valid for all sequences of Newton iterates under metric regularity, and recall and extend results on multivalued approximations for general inclusions $0 \in F(x)$. Equations with continuous, non-Lipschitzian f are considered, too.

Keywords Generalized Newton method · local convergence · inclusion · generalized equation · regularity · Newton map · nonlinear approximation · successive approximation

Mathematics Subject Classification (2010) 49J53 · 49K40 · 90C31 · 65J05

1 Introduction

Starting more than 30 years ago, the extension of Newton's method to non-smooth and multi-valued settings and their application to nonlinear complementarity problems, KKT systems, variational inequalities, generalized equations and other model classes has become a broad field of research. Important early contributions were e.g. [19,27,35,36,38,45,47,49,52]. The further development of the theory around this subject has been strongly influenced by the progress of stability theory in variational analysis. Some recent monographs [3,15,17,23,24,31,59] reflect the manifold aspects of the theory of generalized Newton methods which is mostly developed in the framework of generalized equations

$$0 \in f(x)+M(x), \text{ where } f: X \to Y \text{ is a function and } M: X \rightrightarrows Y \text{ is a multifunction,} \quad (1.1)$$

and $M \equiv \{0\}$ corresponds to equations. The Newton steps are defined by approximations $f^{(k)}$ of f near the current iterate x_k and the solutions x_{k+1} of

Diethard Klatte
IBW, Universität Zürich, Moussonstrasse 15, CH-8044 Zürich, Switzerland. E-mail: diethard.klatte@uzh.ch

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

$$0 \in f^{(k)}(x) + M(x), \tag{1.2}$$

the generated sequence $\{x_k\}$ is called a *Newton sequence.* If generalized derivatives are used, $f^{(k)}$ may be even multivalued. The standard assumption for smooth equations that $Df(x)^{-1}$ exists on some region of interest (or simply at a zero $\bar{x}$ of f) turns into a regularity condition for $f + M$, which requires particular properties of the inverse mapping $(f + M)^{-1}$.

Our paper follows this line. In the model (1.1), we assume that X and Y are Banach spaces, while f, $f^{(k)}$ and M are specified at the corresponding place. We focus on local convergence analysis for Newton's method and pay particular attention to superlinear convergence. So, except for Prop. 8 which concerns so-called semi-local convergence, we throughout suppose that the initial point is sufficiently close to some solution. Global convergence and numerical implementations are not considered. Our approximations of $f+M$ concern linear approximations of f in section 3, multivalued approximations of f by several generalized derivatives in section 4, particular approximations of f including point-based approximations [52] in section 5 and multivalued approximations of M in section 6.

We aim to characterize the different types of convergence which hold under the supposed regularities and particular types of approximations. For that purpose, we recall several known approaches and results in this context, most of them are taken from the authors' monograph [31]. On the other hand, we extend this by presenting original results in each section. To distinguish between the survey or original character of main results, we use throughout "Theorem" for known statements and "Proposition" or "Corollary" for new ones. As a main feature, we discuss the necessity of our requirements (at least for equations given by locally Lipschitzian $\mathbb{R}^n$ homeomorphisms) and for possible concrete realizations. We point out the limitation of various general approaches, and try also to clarify whether the convergence properties need additional properties of M (besides regularity of $f + M$) or not. Commonly known and new examples play a crucial role in these discussions.

In section 2, we present several tools that are needed throughout the paper, discuss the regularity notions, introduce our iteration scheme and derive basic estimates for the later convergence analysis.

In section 3, we survey known results on Newton maps for functions [31,38], including the basic convergence criterion Thm. 4. Our new topics concern the relation between Newton maps and semismoothness, consequences in view of directional differentiability (by the help of a simple mean-value statement) and an application of Newton maps to inclusions.

Section 4 extends known results for equations to the model (1.1). It shows how various types of generalized derivatives can be used for approximating locally Lipschitz functions f in (1.1) and that the Newton map condition now becomes the condition (CA)*, introduced in [31,38]. The Propositions 6 and 7 permit approximate solutions of the auxiliary problems, the other statements of section 4 characterize the imposed assumptions.

In section 5, we present and apply the successive approximation scheme of [31] which has various applications under pseudo-regularity. Originally used for implicit functions, we exploit it directly for Newton's method. The resulting Prop. 8 is a Kantorovich-type statement which holds for all constructed Newton sequences, in contrast to standard formulations in the literature, where only the existence of a Newton sequence is asserted under pseudo-regularity.

In section 6, we investigate multivalued approximations for inclusions $0 \in F(x)$, which were introduced in [39] and allow applications in convex analysis, too. Here, pseudo-regularity is replaced by an upper Lipschitz property, and Prop. 9 concerns the model (1.1) with non-differentiable f as in section 5.

Section 7 is devoted to approaches which do not use known generalized derivatives or Lipschitz properties of f. First we recall an observation for PC^1 equations, which justifies the application of the usual Newton method to a fixed C^1-function, and add some consequences. Then, following [20], we motivate the need of piecewise linear approximations and automatic differentiation for handling functions which are composed by differentiable and (mostly simple) non-smooth functions as well. The approximations are then not necessarily defined via generalized derivatives. Further, we discuss there the role of the key condition (CA)* of section 4 for convergence of Newton's method when non-Lipschitz functions f are permitted. In particular, we present examples of such f which allow to find a zero of them via a superlinearly converging Newton method.

Finally, let us note that hypotheses of strong or metric regularity for the model (1.1), which often prevent applications to nonsmooth examples (like the trivial equation $|x| = 0$) will not appear in most of our new results, excepted for Prop. 8 and Prop. 11.

2 Pre-Requisites

2.1 Basic definitions and consequences of Lipschitz behavior

In the whole paper, if nothing else is specified, X and Y will denote Banach spaces with elements x, y, respectively, and the symbol B is used for the closed unit ball in X or Y. Given $x \in X$, $Z \subset X$ and $\varrho \in \mathbb{R}$, we put as usual $x + \varrho Z := \{x + \varrho z \mid z \in Z\}$.

$\mathrm{Lin}(X,Y)$ means the normed space of linear and bounded operators $A : X \to Y$. In order to say that $f : X \to Y$ is locally Lipschitz we write $f \in C^{0,1}(X,Y)$, while $f \in C^{1,1}(X,Y)$ means that f has a locally Lipschitzian first derivative. A multifunction $F : X \rightrightarrows Y$ is closed if $\mathrm{gph}\, F := \{(x,y) \mid y \in F(x)\}$ is closed in $X \times Y$.

As usual we shall use o-type functions which have the property $o(x) / \|x\| \to 0$ if $0 \neq \|x\| \to 0$, where $o(0)$ may be arbitrary. A function of type O will be understood as $O(x) \to 0$ if $\|x\| \to 0$. Note that $\limsup_{x\to 0} \frac{\|O(x)\|}{\|x\|} < \infty$ (as in a more common definition of the Big-O symbol) is not required. We say that some property holds *for x near $\bar{x}$*, if it holds for all x with sufficiently small $\|x - \bar{x}\|$.

Now we recall several concepts of local Lipschitz properties for mappings $F : X \rightrightarrows Y$ and/or its inverse $S = F^{-1} : Y \rightrightarrows X$ at a point $(x_0, y_0) \in \mathrm{gph}\, F$ or $(y_0, x_0) \in \mathrm{gph}\, S$, respectively. Though we restrict us to Banach spaces, generalizations to metric spaces are easily possible.

The terminology for the subsequent properties (D1) - (D4) is rather different (and permanently changing and extending) in the literature. Therefore, we shall often recall our definitions which follow the authors' book [31].

Let $(y_0, x_0) \in \mathrm{gph}\, S$ be given.

(D1) S is said to be *pseudo-Lipschitz* at (y_0, x_0) if there are neighborhoods $U \ni x_0, V \ni y_0$ and some $L > 0$ such that

$$\forall (y,x) \in (V \times U) \cap \mathrm{gph}\, S \text{ and } y' \in V : \ \exists x' \in S(y') : \ \|x' - x\| \leq L\, \|y' - y\|. \tag{2.1}$$

This notion was introduced and investigated in [4]; it is also called *Aubin property* [54].

(D2) S is called *locally upper Lipschitz* (briefly *locally u.L.*) at (y_0, x_0) if there are neighborhoods $U \ni x_0, V \ni y_0$ and some $L > 0$ such that

$$\forall (y,x) \in (V \times U) \cap \mathrm{gph}\, S : \ \|x - x_0\| \leq L\, \|y - y_0\|. \tag{2.2}$$

If the same holds with $U = X$, we call S *globally upper Lipschitz.*

(D3) S is called *Lipschitz lower semi-continuous (Lipschitz l.s.c.)* at (y_0, x_0) if there is a neighborhood $V \ni y_0$ and some $L > 0$ such that

$$\forall y' \in V \ \ \exists x' \in S(y') : \|x' - x_0\| \leq L\, \|y' - y_0\|. \tag{2.3}$$

(D4) S is called *strongly Lipschitz at* (y_0, x_0) if the neighborhoods in (D1) can be taken in such a way that, in addition, $S(y') \cap U$ is single-valued for all $y' \in V$. Then $S(y') \cap U$ is locally (near y_0) a Lipschitz function.

Under (D2), $S(y_0) \cap U$ is equal to $\{x_0\}$, and empty sets $S(y)$ are permitted for y near y_0. The constant L is called a *rank* (or *modulus*) of the related stability. If F stands for a function $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, all these properties coincide with $\det Df(x_0) \neq 0$.

Definition 1 *(Regularity)* If S is strongly Lipschitz at (y_0, x_0), then $F = S^{-1}$ is called *strongly regular* at (x_0, y_0). Similarly, at the related points: If S is pseudo-Lipschitz then F is called *pseudo-regular.* If S is locally u.L. and $S(y) \cap U \neq \emptyset$ for all $y \in V$, then F is said to be *upper regular.* As before, a constant L for the related Lipschitz property of S is said to be a *rank of regularity.*

Each of these types of *regularity* of F implies that F^{-1} is Lipschitz l.s.c. at (y_0, x_0). Obviously, strong and pseudo-regularity are *persistent under small variations* of (x_0, y_0) in $\operatorname{gph} F$. Often, pseudo-regular is called *metrically regular*; the slightly different definitions are equivalent. The fundamental pseudo-regularity example is any linear function $F : X \to Y$ with $F(X) = Y$. The constant mapping $F(.) = Y_0$ is pseudo-regular at (x_0, y_0) if $y_0 \in \operatorname{int} Y_0$, but not upper regular. The function $f(x) = x + x^2 \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$ (discussed, e.g., in [29]) is upper, but not pseudo-regular at $(0, 0)$. On the other hand, it has been shown, the deep

Theorem 1 [18] *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be pseudo-regular at $(x_0, 0)$. If f is directionally differentiable* at x_0, *then the zero x_0 is isolated and f is upper regular at $(x_0, 0)$. Further, if f is directionally differentiable* for x near x_0, *then $\inf_{x\in\Omega,\, \|u\|=1} \|f'(x; u)\| > 0$ holds for some neighborhood $\Omega \ni x_0$.*

Here $f : \mathbb{R}^n \to \mathbb{R}^m$ is called *directionally differentiable* at x_0 if the standard (one-sided) directional derivative $f'(x_0; u)$ exists for every $u \in \mathbb{R}^n$. Nonsmooth Newton methods are usually developed for functions satisfying at least the assumptions of the last theorem. Therefore, requiring upper regularity (as in section 6) is, at least formally, weaker than pseudo-regularity in that context.

Remark 1 (equivalent definitions) For deriving estimates, the neighborhoods U and V may be replaced, in the definitions, by open or closed balls around x_0, y_0, respectively. We shall use that they have the same radius after multiplying one norm with some factor.

Remark 2 (regularity and methods) In the context of solution methods, pseudo-regularity of F or its approximation at some $(\bar{x}, \bar{y})$ is often the key assumption, and x' and x in (D1) are the iterations x_{k+1} and x_k, respectively. Accordingly, then *some* x_{k+1} fulfills the desired estimate. So convergence requires to choose the "right one", which is insufficient for an algorithm without an appropriate selection rule. Similarly, strong and upper regularity imply that x_{k+1} exists in some ball $\bar{x} + \rho B$ (uniquely or not) where it satisfies the estimate in question. For small ρ, this is again insufficient if also $x_{k+1} \notin \bar{x} + \rho B$ may happen. However, the latter is a usual situation in the world of nonlinear methods.

In the spirit of Remark 2, it is useful to know whether a pseudo-regular mapping F is even strongly regular. For real functions, this is always true. For $X = Y = \mathbb{R}^n$, this holds, e.g., in these cases:

1. $F(x) = \Phi(x) + N_M(x)$, where $\Phi \in C^1$ and N_M is the usual normal-map of a convex polyhedron M, cf. [14]. In consequence, the mapping $S = S(a, b)$ of KKT-points (x, λ) for nonlinear parametric optimization problems
$$P(a, b) : \quad \min\{f(x) - \langle a, x\rangle \mid x \in \mathbb{R}^n,\, g_i(x) \le b_i, i = 1, \ldots, m\} \quad (f, g \in C^2) \tag{2.4}$$
is strongly Lipschitz at some point if S is pseudo-Lipschitz there. This fails under less differentiability, see [31, Expl. BE4] for an example without constraints and $f \in C^{1,1}(\mathbb{R}^2, \mathbb{R})$.
2. F is a so-called generalized Kojima-function, with C^1 data, cf. [31, Cor. 7.22].
3. F is assigned to critical points of certain cone-constraint variational problems [34].
4. $F = \partial f$ is the subdifferential of a convex function $f : \mathbb{R}^n \to \mathbb{R}$, [31, Thm. 5.4].
5. If X is a (real) Hilbert space, (D1) and (D4) coincide for the mapping $X_{glob}(a, b)$ of global minimizers to (2.4) with arbitrary functions f and g [31, Cor. 4.7].

Though pseudo- and strong regularity are broadly used hypotheses for stability of solutions or convergence of Newton-type methods, one has to consider the limits of these assumptions. It might be hard to verify them for concrete applications. Let us point out some facts.

In spite of many known equivalent conditions, one finds nowhere verifiable criteria which allow us to check them for all continuous $\mathbb{R}^n$-functions (in contrast to $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, where $\det Df(x_0) \neq 0$ is necessary and sufficient for both pseudo- and strong regularity of f at $(x_0, 0)$). Moreover, in [31, §5.2] the following was stated: If $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and pseudo-regular at $(x_0, 0)$ without being strongly regular, then every function ϕ which assigns, to $y \in \delta B$, some $\phi(y) \in f^{-1}(y)$ with $\phi(0) = x_0$, is discontinuous somewhere on δB. For $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ and $m > n$, pseudo-regularity never holds.

Also for problems (2.4) with C^3 functions f, g, which permit a deeply developed theory of critical points [25,26], intrinsic conditions for obtaining strongly Lipschitz (or pseudo-Lipschitz) *stationary primal solutions* $x = x(a, b)$ do not exist (differently from locally u.L.), as far as only the Mangasarian-Fromovitz constraint qualification is supposed in place of the linear independence condition, cf. [32].

Finally, we remind of [31, Expl. BE.2]: For $X = l^2$ and $Y = \mathbb{R}$, the function $x = (x_1, x_2, \ldots) \mapsto f(x) = \inf_k x_k$ is globally Lipschitz, concave and directionally differentiable. The mapping $F(x) = \{y \in \mathbb{R} \mid f(x) \le y\}$ is everywhere pseudo-regular with rank 2. However, known *sufficient* conditions in terms of contingent and co-derivatives (which are *necessary* in finite dimension), cf. e.g. [4, sect. 7.5], [31, §3.3] or [43], fail to detect the pseudo-regularity of F.

2.2 Some known generalized derivatives

Generalized derivatives may be a help for characterizing regularity or solution methods like for smooth functions, provided they are available. Given a multifunction $F : X \rightrightarrows Y$, the *contingent derivative* $CF(x_0, y_0)(u)$ of F at $(x_0, y_0) \in \operatorname{gph} F$ in direction $u \in X$ (also called *graphical derivative* or *Bouligand derivative*) consists of all limits $v = \lim t_k^{-1}[y_k - y_0]$ where $y_k \in F(x_0 + t_k u_k)$ for certain sequences $t_k \downarrow 0$ and $u_k \to u$. If F is a function then $y_k = F(x_k)$ is unique and one writes simpler $CF(x_0)(u)$. For any function $f : X \to Y$, the set of limits $Tf(x)(u) = \{v \mid \exists t_k \downarrow 0,\, x_k \to x \text{ such that } v = \lim t_k^{-1}\,[f(x_k + t_k u) - f(x_k)]\,\}$ is the *Thibault derivative of f at x in direction u* (notation from [31]). It is also called *strict graphical derivative* or *paratingent derivative* or *limit set.*

For $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$, the set $\mathcal{D}$ of all $x \in \mathbb{R}^n$ such that the Fréchet-derivative $Df(x)$ exists, has full Lebesgue measure [48]. In consequence, the *B-subdifferential of f at x,* defined by $\partial_B f(x) = \{A \mid \exists x_k \to x,\ x_k \in \mathcal{D} \text{ with } A = \lim Df(x_k)\}$ is not empty. Its convex hull $\partial^{CL} f(x) = \operatorname{conv} \partial_B f(x)$ is *Clarke's generalized Jacobian* [10,11] - a non-empty, compact set. Writing $\partial^{CL} f(x)(u) = \{Au \mid A \in \partial^{CL} f(x)\}$ and similarly $\partial_B f(x)(u)$, the (possibly proper) inclusions $\partial_B f(x)(u) \subset Tf(x)(u) \subset \partial^{CL} f(x)(u)$ and $\operatorname{conv} Tf(x)(u) = \partial^{CL} f(x)(u)$ hold true. Notice however that - as in all double limit constructions (hence also for so-called limiting normals or limiting coderivatives) - computing the sets $Tf(x)(u)$ and $\partial^{CL} f(x)$ may be a hard job, even if f is piecewise linear.

In finite dimension, there are close relations between stability and generalized derivatives. We present and discuss only those facts which are needed later on and cite them from [31]. For further characterizations of stability concepts we also refer, e.g., to the standard monographs [15,54].

Theorem 2 [31, Thm. 5.1] (regularity and derivatives). *Let $F : X \rightrightarrows Y$ be closed and $z_0 = (x_0, y_0) \in \operatorname{gph} F$. If $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$,*

$$\begin{array}{l} F \text{ is upper regular at } z_0 \\ \Leftrightarrow\ F^{-1} \text{ is Lipschitz l.s.c. at } (y_0, x_0) \text{ and } 0 \in CF(z_0)(u) \text{ implies } u = 0. \end{array} \tag{2.5}$$

$$\begin{array}{l} \text{If } F^{-1} \text{ is Lipschitz l.s.c. at } (y_0, x_0) \\ \text{then there exists } L > 0 \text{ such that } B \subset CF(z_0)(LB). \end{array} \tag{2.6}$$

$$\begin{array}{l} F \text{ is pseudo-regular at } z_0 \text{ with rank } L \\ \Leftrightarrow\ \exists \varepsilon > 0 : B \subset CF(z)(LB) \text{ for all } z \in \operatorname{gph} F \cap (z_0 + \varepsilon B). \end{array} \tag{2.7}$$

$$\begin{array}{l} F = f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n) \text{ is strongly regular at } (x_0, f(x_0)) \\ \Leftrightarrow\ 0 \in Tf(x_0)(u) \text{ implies } u = 0. \end{array} \tag{2.8}$$

If X is a normed space and $Y = \mathbb{R}^m$, the conditions (2.5) remain necessary for upper regularity. The characterization of upper regularity goes back to [30,53]. If X and Y are Banach spaces, the condition (2.7) is sufficient for pseudo-regularity, cf. [4, Thm. 4, §7.5], while the opposite direction fails by [31, Expl. BE.2]. Statement (2.8) is the inverse function theorem of [37] where also Thm. 1F.2 of [15], chain rules for Tf and applications can be found, while Clarke's inverse function theorem [10] says that $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x_0, f(x_0))$ if all matrices

$A \in \partial^{CL} f(x_0)$ are non-singular. Since $Tf(.)(.)$ is both closed and homogeneous in u, (2.8) means equivalently

$$\exists c > 0, \delta > 0 \text{ such that } v \in Tf(x)(u) \text{ implies } \|v\| \geq c\|u\| \ \forall x \in x_0 + \delta B. \tag{2.9}$$

In this case, the derivative of the local inverse function f^{-1} is the inverse of $Tf(x)$

$$u \in T(f^{-1})(f(x))(v) \Leftrightarrow u \in Tf(x)^{-1}(v) \Leftrightarrow v \in Tf(x)(u). \tag{2.10}$$

Thm. 2 allows a simple analysis of perturbed generalized equations $0 \in \hat{F} := g + F$ where the perturbing function $g : X \to Y$ is Lipschitz on some set $\Omega \subset X$. Here, the quantities

$$\begin{aligned} &\sup (g, \Omega) := \sup \{ \|g(x)\|_Y \mid x \in \Omega\} \quad \text{and} \\ &\mathit{Lip}\ (g, \Omega) := \inf \{L > 0 \mid \|g(x) - g(x')\|_Y \leq L\|x - x'\| \ \ \forall x, x' \in \Omega\} \end{aligned}$$

are important since $\mathit{Lip}\ (g, \Omega)$ plays the role of $\sup (\|Dg\|, \Omega)$ for $g \in C^1$. If $x \in \operatorname{int} \Omega$, it follows directly from the definitions that, with $\beta = \mathit{Lip}\ (g, \Omega)$,

$$\begin{array}{c} CF(x,y)(u) \ \subset C\hat{F}(x, y + g(x))(u) + \beta\|u\|B \ \subset CF(x,y)(u) + 2\beta\|u\|B \\ \text{and, if } F = f \text{ is a function,} \\ TF(x)(u) \ \subset T\hat{F}(x)(u) + \beta\|u\|B \ \subset TF(x)(u) + 2\beta\|u\|B. \end{array} \tag{2.11}$$

So one can use (2.7) and (2.9) along with the estimate (2.11) for checking pseudo-regularity of $\hat{F} = g + F$ and strong regularity of $\hat{f} = f + g$, respectively, with $g \in C^{0,1}$. The next theorem is well-known. Since it is usually derived from one of its infinite dimensional versions, we include the simple proof.

Theorem 3 *(i) If $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is pseudo-regular at (x_0, y_0) with rank L then $\hat{F}$ is pseudo-regular at $(x_0, y_0 + g(x_0))$ with rank λ^{-1} whenever $\lambda = L^{-1} - Lip(g, x_0 + rB) > 0$ for some $r > 0$. (ii) If $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ is strongly regular at $(x_0, f(x_0))$ then so is $\hat{f}$ at $(x_0, \hat{f}(x_0))$ if $c - Lip(g, x_0 + rB) > 0$ for some $r > 0$ with c from (2.9).*

Proof (i) To apply condition (2.7), let $v \in Y$ and $\beta = Lip(g, x_0 + rB)$. By assumption, some $u \in L\|v\|B$ fulfills $v \in CF(x,y)(u)$. By (2.11), we may write

$$v = v' - w' \text{ with some } v' \in C\hat{F}(x, y + g(x))(u) \text{ and } w' \in \beta\|u\|B.$$

So we can estimate: $\|v'\| = \|v + w'\| \geq \|v\| - \|w'\| \geq \|v\| - \beta\|u\| \geq L^{-1}\|u\| - \beta\|u\| = \lambda\|u\|$. Hence the assertion follows from (2.7).
(ii) In the situation (2.9), the assertion follows analogously from (2.11).

For Banach spaces, these statements remain true (cf. e.g. [15, Thms. 5E.1, 5F.1], [31, §4.3]). In this case, however, the above proof fails because condition (2.7) in terms of CF is only sufficient, see [31, Expl. BE.2]. Hence one cannot use CF (and similarly TF) for proving Thm. 3 in infinite dimension; one needs more involved arguments, e.g., by using Thm. 10 below.

2.3 Particular nonsmooth functions

Particularly composed functions and point based approximations

We start with composed functions of the type smooth ∘ Lipschitz.

$$\begin{array}{c} \text{Let} \qquad f(x) = h(\gamma(x)) \ \text{ where } \ h \in C^1(Z, Y), \ \gamma \in C^{0,1}(X, Z), \\ \text{and } \Sigma(x', x) = f(x) + Dh(\gamma(x))(\,\gamma(x') - \gamma(x)\,) \text{ be its partial linearization at } x. \end{array} \tag{2.12}$$

Such functions appear in [52] and have useful properties. If we restrict all arguments to some region where γ has Lipschitz rank L_γ, the function

$$g_x(x') = f(x') - \Sigma(x', x) \tag{2.13}$$

satisfies, as known from the usual case of $f = h \in C^1$ and $\gamma = \mathrm{id}$,

$$\begin{aligned} \| g_x(x') \| &= \|f(x') - f(x) - Dh(\gamma(x))(\gamma(x') - \gamma(x))\| \\ &= \| \int_0^1 [Dh(\gamma(x + t(x' - x))) - Dh(\gamma(x))]\, (\gamma(x') - \gamma(x))\, dt \| \\ &\le \sup_{t\in(0,1)} \| Dh(\gamma(x + t(x' - x))) - Dh(\gamma(x)) \|\, L_\gamma \|x' - x\| \\ &= O(\|x' - \bar{x}\| + \|x - \bar{x}\|)\, \|x' - x\|; \end{aligned} \tag{2.14}$$

$$\begin{aligned} \|g_x(x') - g_x(x'')\| &= \| f(x') - f(x'') - Dh(\gamma(x))(\gamma(x') - \gamma(x'')) \| \\ &\le \sup_{t\in(0,1)} \| Dh(\gamma(x' + t(x' - x''))) - Dh(\gamma(x)) \|\, L_\gamma \|x' - x''\| \\ &= O(\|x' - x''\| + \|x - \bar{x}\|)\, \|x' - x''\|. \end{aligned} \tag{2.15}$$

In the context of Robinson's [52] *point-based approximation* (PBA), f and Σ are continuous functions on an open set $\Omega \subset X$ and $\Omega \times \Omega$, respectively. Σ is called a PBA for f on Ω, if there is a constant K such that (for all $x, \bar{x}, x', x'' \in \Omega$),

$$\begin{array}{lll} (a) & \|f(x') - \Sigma(x', x)\| & \le \frac{1}{2} K\, \|x' - x\|^2 \\ (b) & \| [\Sigma(x'', x) - \Sigma(x'', \bar{x})] - [\Sigma(x', x) - \Sigma(x', \bar{x})] \| & \le K\, \|x - \bar{x}\|\, \|x'' - x'\|. \end{array} \tag{2.16}$$

It was a basic observation in [52] that the conditions (2.16) can be (locally) satisfied for f in (2.12) with $h \in C^{1,1}$. Replacing $f(x')$ by $\Sigma(x', \bar{x})$ in (2.13) then, for each $\bar{x} \in \Omega$, the difference

$$g_x(x') = \Sigma(x', x) - \Sigma(x', \bar{x}) \tag{2.17}$$

describes $\Sigma(., x)$ as a perturbation of $\Sigma(., \bar{x})$ by a continuous function g_x which satisfies

$$\|g_x(x')\| \le \|\Sigma(x', x) - f(x')\| + \|f(x') - \Sigma(x', \bar{x})\| \le \tfrac{1}{2} K(\|x' - x\|^2 + \|x' - \bar{x}\|^2); \tag{2.18}$$

$$\|g_x(x') - g_x(x'')\| = \|(\Sigma(x', x) - \Sigma(x', \bar{x})) - (\Sigma(x'', x) - \Sigma(x'', \bar{x}))\| \le K\|x - \bar{x}\|\|x'' - x'\|. \tag{2.19}$$

Next restrict all arguments to a ball $\Omega = \Omega_r = \bar{x} + rB$. Then the estimates (2.14,2.15,2.18,2.19) ensure estimates for $\sup(g_x, \Omega)$ and $Lip\,(g_x, \Omega)$, namely

$$\begin{array}{ll} \|g_x(\bar{x})\| \le o(r), & \|g_x(x') - g_x(x'')\| \le O(r)\|x'' - x'\| \quad \text{and} \\ \|g_x(x')\| \le \|g_x(\bar{x})\| + \|g_x(x') - g_x(\bar{x})\| & \le o(r) + O(r)r \le o(r). \end{array} \tag{2.20}$$

Hence, though generally nonsmooth, the "linearizations" Σ obey properties which are known to be important for Newton's method in the smooth case. In particular, Thm. 3 can be applied (for finite dimensions) to the functions $\Sigma(., x)$ and perturbations g_x near $\bar{x}$, if r is small enough, in order to verify persistence of pseudo- [strong] regularity under small perturbations by g_x and for estimating the related Lipschitz ranks via (2.11) as well.

We finish the current technical part by deriving an estimate for the difference (2.17) under the hypotheses (2.12) only, namely

$$\|g_x(x')\| \le o(x - \bar{x}) + O(x - \bar{x})\, \|x' - \bar{x}\|. \tag{2.21}$$

To prove (2.21) we abbreviate $\bar{H} = Dh(\gamma(\bar{x}))$, $H_x = Dh(\gamma(x))$, $\gamma' = \gamma(x')$, $\gamma = \gamma(x)$, $\bar{\gamma} = \gamma(\bar{x})$, use the mean-value theorem for h and the Lipschitz property of γ for the points in question.

$$\begin{aligned} \|g_x(x')\| &= \|f(x) + H_x(\gamma' - \gamma) - [f(\bar{x}) + \bar{H}(\gamma' - \bar{\gamma})]\| \\ &= \| f(x) - f(\bar{x}) + H_x(\gamma' - \gamma) - \bar{H}(\gamma' - \bar{\gamma}) \| \\ &= \| [h(\gamma) - h(\bar{\gamma}) - \bar{H}(\gamma - \bar{\gamma})] + \bar{H}(\gamma - \bar{\gamma}) + H_x(\gamma' - \gamma) - \bar{H}(\gamma' - \bar{\gamma})\| \\ &= \| o(\gamma - \bar{\gamma}) + \bar{H}(\gamma - \bar{\gamma} + \bar{\gamma} - \gamma') + H_x(\gamma' - \gamma) \| \\ &= \| o(\gamma - \bar{\gamma}) + \bar{H}(\gamma - \gamma') + H_x(\gamma' - \gamma) \| \\ &\le o(x - \bar{x}) + \| (H_x - \bar{H})\, (\gamma' - \gamma) \| \\ &= o(x - \bar{x}) + \| (H_x - \bar{H})\, (\gamma' - \bar{\gamma} - \gamma + \bar{\gamma}) \| \\ &\le o(x - \bar{x}) + O(x - \bar{x})\, \|\gamma' - \bar{\gamma}\| + O(x - \bar{x})\, \|\gamma - \bar{\gamma}\|. \end{aligned}$$

Due to $O(x - \bar{x})\, \|\gamma - \bar{\gamma}\| = o(x - \bar{x})$, we now obtain (2.21) as asserted.

PC^1 functions and pseudo-smooth functions

A function $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ is called *piecewise smooth,* if there are functions $f^s \in C^1(\mathbb{R}^n, \mathbb{R}^m)$ $(s = 1, \dots N)$ such that the sets $I(x) = \{s \mid f(x) = f^s(x)\}$ are non-empty $(\forall x \in \mathbb{R}^n)$; briefly $f =$

$PC^1(f^1,\dots,f^N)$, $f\in PC^1$. For basic properties of these functions we refer to [55]. We only mention the possible description of $\partial^{CL} f(x)$ as $\mathrm{conv}\{Df^s(x)\,|\,x\in \mathrm{cl}\ \mathrm{int}\, I^{-1}(s)\}$ and $\mathrm{conv}\{Df^s(x)\,|\,x\in I^{-1}(s)$ and $\exists x'\to x : Df(x')=Df^s(x')\}$, cf. [55, Prop.A.4.1] and [36, Prop.4], respectively. The needed functions f^s with $x\in \mathrm{cl}\ \mathrm{int}\, I^{-1}(s)$ are called essentially active at x.

A function $f\in C^{0,1}(\mathbb{R}^n,\mathbb{R}^m)$, which is C^1 on an open and dense subset $\Omega\subset\mathbb{R}^n$, is called *pseudo-smooth.* We denote this class of functions by C^1_Ω. Such f appear in many applications, cover the class PC^1 and many so-called NCP-functions [57]. They have locally bounded derivatives on Ω and nonempty sets $D^\circ f(x)=\{A\mid \exists x_k\to x,\ x_k\in\Omega \text{ such that } A=\lim Df(x_k)\}$, which could be called the *small B-subdifferential.* For PC^1 functions, $D^\circ f(x)=\partial_B f(x)$ is valid. Several further relations between these sets become evident by an example which was made for showing that Newton's method may fail to converge when $f\in C^{0,1}$ has a Lipschitzian inverse and is directionally differentiable.

Example 1 In [36, §2.3], a real Lipschitz function $f\in C^1_\Omega\setminus PC^1$ was constructed and analyzed in detail, which is directionally differentiable, strongly regular and satisfies

$$f(0)=0,\ Df(0)=1,\ 0\notin\Omega,\ D^\circ f(0)=\{\tfrac12,2\} \text{ and } \partial_B f(0)=\{\tfrac12,1,2\}\neq\partial^{CL}f(0)=[\tfrac12,2].$$

It was shown that if one starts at any $x_0\neq 0$ where $Df(x_0)$ exists, then the usual Newton method generates an alternating sequence $x_0,x_1,x_2=-x_1,x_3=x_1,\dots$ with $x_k\in\Omega\,\forall k$. f has been also discussed (with pictures) in [31, Expl. BE.1] and [17, Expl. 7.4.1].

2.4 The iteration schemes and their solvability

In order to solve an inclusion $0\in\Gamma(x)$ for given $\Gamma: X\rightrightarrows Y$, our most general iteration schemes are described by a multifunction $\Sigma: X\times X\rightrightarrows Y$ in such a way that some initial point x_0 must be given, $x=x_k$ is the current iteration point and any solution $x'=x_{k+1}$ of $0\in\Sigma(.\,,x)$ is the next one. In other words, the concrete choice of Σ characterizes the considered method, and

$$x'\in\mathcal{S}(x):=\Sigma(.\,,x)^{-1}(0) \text{ defines the next iterates.} \tag{2.22}$$

Usually, $\Sigma(.\,,x)$ stands for some (multi-)function which approximates Γ near x. Let

$$r(.)=o(.) \text{ with } r(0)=0 \quad\text{or}\quad r(.)=q\|.\|,\ 0<q<1. \tag{2.23}$$

Considering a solution $\bar x$ of the inclusion, we will study local convergence

$$\begin{aligned}&\|x_{k+1}-\bar x\|\le r(x_k-\bar x)\ \text{ (as far as } x_{k+1}\in\mathcal{S}(x_k) \text{ exist)},\\ &\text{provided that } \|x_0-\bar x\| \text{ is sufficiently small.}\end{aligned} \tag{2.24}$$

Equivalently, one may require $\mathcal{S}(x)\subset\bar x+r(x-\bar x)B$ if $\|x-\bar x\|$ is sufficiently small, while the additional condition $\mathcal{S}(x)\neq\emptyset$ for x near $\bar x$ ensures, by $\|x'-\bar x\|<\|x-\bar x\|$, the existence of x_{k+1} in each step. Therefore, Σ describes a well defined method that generates a convergent sequence $x_k\to\bar x$ with the local convergence property (2.24) if and only if

$$\emptyset\neq\mathcal{S}(x)\subset\bar x+r(x-\bar x)B \ \text{ for } x \text{ near } \bar x. \tag{2.25}$$

Due to $r(0)=0$, this implies $\mathcal{S}(\bar x)=\{\bar x\}$. In terms of the definitions in §2.1, so (2.25) requires that the mapping $\mathcal{S}$ of next iterates is Lipschitz l.s.c. and globally upper Lipschitz at $(\bar x,\bar x)$ with fixed modulus $q<1$ or, in the stronger $(r=o)$ case, with each modulus $q<1$.

Example 2 For a generalized equation $0\in\Gamma(x):=f(x)+M(x)$ with $f\in C^1$, the setting $\Sigma(x',x)=f(x)+Df(x)(x'-x)+M(x')$ requires to solve

$$0\in f(x_k)+Df(x_k)(x_{k+1}-x_k)+M(x_{k+1}), \tag{2.26}$$

a standard scheme in the literature.

Example 3 If $\Gamma=f$ is a function, one can interpret Σ by some object like a generalized derivative. The difference $\Sigma(x',x)-f(x)$ is a multifunction depending on x' and x (or on $x'-x$ and x). Calling it $Gf(x)(x'-x)$ we obtain $\Sigma(x',x)=f(x)+Gf(x)(x'-x)$ with some "generalized derivative $Gf(x): X\rightrightarrows Y$ of f at x", which describes the method by

$$0\in\Sigma(x_{k+1},x_k)=f(x_k)+Gf(x_k)(x_{k+1}-x_k). \tag{2.27}$$

The inverse of $Gf(x)$ now defines our iterates and solution sets via $\mathcal{S}(x)=x+Gf(x)^{-1}(-f(x))$.

Based on iteration schemes like (2.26) or (2.27), we intend to connect in the following sections the type of approximations with the desired (or possible) kinds of convergence and more or less concrete iteration rules Σ. In some situations, slightly different convergence results can be derived, e.g., that (2.24) holds for $x_k \neq \bar{x}$ or that, for some positive ρ, $\emptyset \neq \mathcal{S}(x) \cap [\bar{x}+\rho B] \subset \bar{x}+r(x-\bar{x})B$ for x in $\bar{x}+\rho B$ is valid.

The concrete meaning of several Newton-auxiliary problems in $\mathbb{R}^n$ of the form (2.27) is discussed in [31, chapter 11].

The existence of the iterates in $\bar{x}+\rho B$ needs solvability of $0 \in \Sigma(.,x)$ in $\bar{x}+\rho B$ for x near $\bar{x}$ while $0 \in \Sigma(\bar{x},\bar{x})$ is supposed. Typical approaches for guaranteeing this are

(a) the use of an implicit (multi-) function theorem, or
(b) the use of additional properties which make the solvability simpler, or
(c) checking convergence of the sequences $x_{k+1} \in \mathcal{S}(x_k)$ directly.

Concerning (a), one can apply (extensions of) Banach's principle as, e.g., in several papers of A.L. Dontchev and R. T. Rockafellar [12,14] with implicit function theorems of a form like our Corollary 1 below. To this end, the difference g between $\Sigma(.,x)$ and $\Sigma(.,\bar{x})$ has to be a *small Lipschitz function,* i.e., $\sup(g,\Omega)$ *and* $Lip(g,\Omega)$ are small for some neighborhood Ω of $\bar{x}$. Alternatively, as shown by S. Robinson 1979 [50], Kakutani's principle can be used as long as g is a *small continuous function.* Of course, the latter requires convexity assumptions.

Both approaches cannot ensure solvability (with respect to x') of

$$0 \in \Sigma(x',x) := \partial_{x'} h(x',x) \quad \text{for } h(x',x) := \|x-\bar{x}\| f_1(x') + (1-\|x-\bar{x}\|) f_2(x'), \tag{2.28}$$

where h stands for a homotopy between two convex functions f_1 and f_2 on $\mathbb{R}^n$: Here, $\Sigma(.,x) = g(.) + \Sigma(.,\bar{x})$ fails to hold for a function g. Related statements for such perturbations exist, but their hypotheses are quite restrictive, cf. [1] or [31, Thm. 4.5]. On the other hand, solvability for the *particular* example (2.28) can be handled by applying basic tools of convex analysis. Thus (2.28) is typical for situation (b). Other examples are the approximations by Newton maps, defined in §3.1 below, where the auxiliary problems are invertible linear systems.

The direct way (c) stands behind all implicit function theorems which use the contraction principle in original or extended form and is also the common technique for deriving statements of Kantorovich type, cf. section 5.

3 Linear Auxiliary Problems for Equations and Inclusions

In this section, we recall the concept of Newton maps [31,36] for equations, which allows necessary and sufficient conditions for local superlinear convergence. Motivated by a referee's question, we point out its relations to semismoothness and discuss consequences in view of directional differentiability by the help of a simple mean-value statement. Finally, an application of Newton maps to inclusions is presented.

3.1 Newton maps for nonsmooth equations

Newton's method for computing a zero $\bar{x}$ of a C^1 function $f: X \to Y$ (Banach spaces) is determined by the iterations $f(x_k)+A(x_{k+1}-x_k)=0$, where x_0 is given and $A = Df(x_k) \in Lin(X,Y)$. The *local superlinear convergence* of $\{x_k\}$ to $\bar{x}$ for this method means that

$$\text{with some o-type function, it holds } \|x_{k+1}-\bar{x}\| \leq o(x_k-\bar{x}) \text{ for } x_0 \text{ near } \bar{x}. \tag{3.1}$$

To create a similar method for arbitrary $f: X \to Y$, one has to think about the choice of A. We will now discuss this by recalling the approach and some basic results from [31].

$$\text{Let } \mathcal{N} \text{ be any multifunction which assigns, to } x \in X, \text{ a set } \emptyset \neq \mathcal{N}(x) \subset Lin(X,Y), \tag{3.2}$$

and let $\bar{x} \in X$. We interpret $\mathcal{N}(x)$ as the permitted Newton operators for the iterations

$$f(x_k) + A(x_{k+1}-x_k) = 0 \quad \text{with some } A \in \mathcal{N}(x_k), \quad x_0 \text{ given.} \tag{3.3}$$

Definition 2 *(Newton map)* We call $\mathcal{N}$ a Newton map (briefly N-map) for f at $\bar{x}$ if

$$A(x-\bar{x}) \in f(x)-f(\bar{x})+o(x-\bar{x})B \text{ holds for all } A\in\mathcal{N}(x) \text{ and } x \text{ near } \bar{x}. \tag{3.4}$$

Notice that we require nothing for $x=\bar{x}$ since $o(0)$ and $A\in Lin(X,Y)$ may be arbitrary.

Definition 3 *(Newton-regularity)* We say that $\mathcal{N}$ is Newton-regular (briefly N-regular) at $\bar{x}$ if there are constants K^+, K^- such that A^{-1} exist and

$$\|A\|\le K^+ \text{ and } \|A^{-1}\|\le K^- \text{ hold for all } A\in\mathcal{N}(x) \text{ and sufficiently small } \|x-\bar{x}\|. \tag{3.5}$$

If only the existence of a related K^- is required, we speak about *weak N-regularity.*

The given notions are motivated by their relations to Newton's method. If both conditions (3.4) and (3.5) are satisfied, we say that $\mathcal{N}$ is a *regular N-map* at $\bar{x}$. Similarly, if (3.4) and weak N-regularity hold true, we call $\mathcal{N}$ a *weakly regular N-map.* The elements x_{k+1} in (3.3) depend on the selected elements A. So we specify that the convergence (3.1) should hold *independently* of the choice of $A\in\mathcal{N}(x_k)$.

Theorem 4 [31, Lemma 10.1]. *Suppose that $f(\bar{x})=0$ and the mapping $\mathcal{N}$ in (3.2) is N-regular at $\bar{x}$, with K^-, K^+ according to (3.5). Then, the method (3.3) fulfills condition (3.1) if and only if $\mathcal{N}$ is a N-map of f at $\bar{x}$. In this case, it holds with o from (3.4), $\|x_{k+1}-\bar{x}\|\le K^-o(x_k-\bar{x})$ and*

$$\tfrac{1}{2}(K^-)^{-1}\|x-\bar{x}\| \;\le\; \|f(x)-f(\bar{x})\| \;\le\; 2K^+\|x-\bar{x}\| \quad \textit{for } x \textit{ near } \bar{x}. \tag{3.6}$$

The existence of K^+ in (3.5) is only needed for verifying the "only if" direction. Hence, the convergence (3.1) is already ensured if $\mathcal{N}$ is a weakly regular N-map.

By (3.6), the zero is isolated and f is "pointwise" Lipschitz with rank $2K^+$ at $\bar{x}$.

In [23], f is called *Newton differentiable at $\bar{x}$* if (3.4) is satisfied, and in [21], after requiring this for all $\bar{x}$ near the zero, f is said to be *slantly differentiable.* To investigate Newton's method, mappings $\mathcal{N}$ satisfying (3.4) and particular realizations have been considered already in [36, Prop. 3]. Particular N-maps and the related Newton-regularity are discussed in [31, chap. 10], where also the following general properties and interrelations can be found.

The union of two N-maps or the convex hull of a N-map are again N-maps (for f at $\bar{x}$). The same holds for all nonempty-valued submappings $\hat{\mathcal{N}}\subset\mathcal{N}$ of a N-map, e.g., for $\hat{\mathcal{N}}=\partial_B f$ if $\mathcal{N}=\partial^{CL}f$ is a N-map for $f\in C^{0,1}(\mathbb{R}^n,\mathbb{R}^n)$. Moreover, the definition of $\mathcal{N}(x)$ at $x=\bar{x}$ plays no role for $\mathcal{N}$ being a N-map at $\bar{x}$. Hence we may assume that $\mathcal{N}(\bar{x})=\{E\}$ (E identity map), after which only $x\ne\bar{x}$ must be considered in (3.5), too. Single-valued selections Rf of a N-map for f at $\bar{x}$ are called *Newton functions for f at $\bar{x}$.*

Theorem 5 [31, Thm. 6.12] (existence and chain rule for Newton functions)

(i) Every locally Lipschitz function $f: X\to Y$ possesses, at each $\bar{x}$, a Newton function Rf being locally bounded by a Lipschitz constant L for f near $\bar{x}$.

(ii) Let $h: X\to Y$ and $g: Y\to Z$ be locally Lipschitz with Newton functions Rh at $\bar{x}$ and Rg at $h(\bar{x})$, respectively. Then the canonically composed function $Rf(x)=Rg(h(x))Rh(x)$ defines a Newton function of $f(.)=g(h(.))$ at $\bar{x}$.

The proof of (i) in [31] is not constructive, but it tells us, that N-maps also exist for $C^{0,1}$ functions which are not directionally differentiable. By (ii), Newton functions (hence also N-maps) satisfy a common chain rule.

Remark 3 (Perturbations and inexact Newton methods) If $\mathcal{N}$ is a N-map of $f\in C^{0,1}(X,Y)$ at a zero $\bar{x}$, so is $\tilde{\mathcal{N}}(x)=\mathcal{N}(x)+\alpha\|f(x)\|\,B_{Lin}$ for fixed $\alpha>0$. Hence, approximating $\mathcal{N}(x)$ with accuracy $\alpha\|f(x)\|$ means passing from one N-map to another. So the concept of Newton maps automatically includes also so-called inexact Newton methods. More results on perturbations of $\mathcal{N}$ and persistence of condition (3.5) can be found in [31, chap. 10].

For readers, interested in approximations by Broyden-updates [6] when f is a nonsmooth Hilbert space function, we refer to [19] and (in finite dimension) to [35], two of the first papers concerning nonsmooth Newton methods at all. For generalized equations, recent Broyden-type results can be found in [2] and [8].

Some Newton maps in finite dimensions

For pseudo-smooth functions $f \in C^1_\Omega$, the single-valued selections of the mapping $D^\circ f$ are natural candidates for being Newton functions (see [31, Thm. 6.18]): If $f \in C^1_\Omega$ and some selection Rf of $D^\circ f$ is a Newton function for f at $\bar{x}$, then $\mathcal{N} = D^\circ f$ is a N-map at the same $\bar{x}$ and $Cf(\bar{x})(u) \subset D^\circ f(\bar{x})u$.

In [31, §6.4], we have presented a subclass of C^1_Ω, called *locally PC^1 functions*. For these functions, $\mathcal{N} = D^\circ f$ is automatically a N-map. For piecewise smooth (and hence directionally differentiable) functions f with $f(\bar{x}) = 0$, N-maps are, e.g.,

$$\begin{array}{ll} \mathcal{N}_1(x) = \{Df^s(x) \mid s \in I(x)\} & \text{where } I(x) = \{s \mid f^s(x) = f(x)\} \\ \mathcal{N}_2(x) = \{Df^s(x) \mid s \in J(x)\} & \text{where } J(x) = \{s \mid \|f^s(x) - f(x)\| \le \|f(x)\|^2\} \\ \text{or} \quad \mathcal{N}_3(x) = D^\circ f(x) = \partial_B f(x), & \mathcal{N}_4(x) = \partial^{CL} f(x). \end{array} \tag{3.7}$$

Clearly, smaller sets $\mathcal{N}(x)$ induce weaker N-regularity conditions (3.5). Under standard assumptions for $f \in PC^1$, even the usual Newton method can be applied to any fixed function f^s which is active at some x_0, sufficiently close to a zero $\bar{x}$, cf. section 7.

3.2 Semismoothness, Newton maps and directional derivatives

Semismoothness has been used for Newton's method by [45,47] and in many subsequent papers. Detailed presentations of semismooth Newton methods can be found e.g. in [17,23,24], with interesting extensions in [59] for function spaces and related complementarity problems.

Throughout this subsection, let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$. Recall that $\mathcal{D} := \{x \mid Df(x) \text{ exists}\}$. The mostly used definition of semismoothness for f at $\bar{x}$ was given by Qi and Sun [47, p.355]:

Definition 4 f is said to be semismooth at $\bar{x}$ if $\lim A^k u$ exists whenever $u \in \mathbb{R}^n, t_k \downarrow 0, u_k \to u$ and $A^k \in \partial^{CL} f(\bar{x} + t_k u_k)$.

For $m = 1$, this is Mifflin's original definition in [42, p.4]. By [47, Lemma 2.1], semismoothness implies that the unique limit $\lim A^k u$ is just the usual directional derivative $f'(\bar{x}; u)$ while [47, Thm. 2.3] says that semismoothness is equivalent to

$$Au - f'(\bar{x}; u) = o(u) \quad \forall u \; \forall A \in \partial^{CL} f(\bar{x} + u) \tag{3.8}$$

$$\text{and to} \quad f'(\bar{x} + u; u) - f'(\bar{x}; u) = o(u) \quad \text{if } \bar{x} + u \in \mathcal{D}. \tag{3.9}$$

Since also $f'(\bar{x}; u) = f(\bar{x} + u) - f(\bar{x}) + o(u)$ holds true (cf. e.g. [31, Lemma A2], [56]), one obtains from (3.8) and Def. 2,

Theorem 6 *Semismoothness of f at $\bar{x}$ means equivalently that $\partial^{CL} f$ is a Newton map for f at $\bar{x}$, and it implies that $f'(\bar{x}; u)$ exists for each u.*

Facchinei and Pang [17, Def. 7.4.2, p.677] used a semismoothness definition which is a bit stronger than Def. 4: f has to be directionally differentiable *near* $\bar{x}$ and

$$f'(x; x - \bar{x}) - f'(\bar{x}, x - \bar{x}) = o(x - \bar{x}). \tag{3.10}$$

Setting $x = \bar{x} + u$, this requires (3.9) for all u near the origin, not only for $\bar{x} + u \in \mathcal{D}$. With this definition, the N-map property of $\partial^{CL} f$ follows from [17, Thm. 7.4.3 (c)]. Hence, Thm. 6 holds for both definitions. For showing the mentioned theorems [17, Thm. 7.4.3] and [47, Thm. 2.3], or only $Cf(x)(u) \subset \partial^{CL} f(x)(u)$, Clarke's mean-value theorem [11] for $\partial^{CL} f$ is the basic tool. Note that in [31], semismoothness of f was defined via the equivalence in Thm. 6.

As well-known, important semismooth functions (according to both definitions) are the Euclidean norm and all PC^1 functions. But knowing examples of non-semismooth functions is also helpful. The first one, Example 1, showed that some assumption is really needed for extending Newton's method to Lipschitz functions. It also indicates, in contrast to an assertion in [22, p. 1339]: The non-singularity of $\partial^{CL} f$ *does not imply* semismoothness of a strongly regular Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^n$. There are even strongly regular, real Lipschitz functions which are *nowhere* semismooth, satisfy $\partial^{CL} f(.) = [a, b]$ with $0 < a < b$ and are *not* directionally differentiable on a dense subset of $\mathbb{R}$, cf. [31, Expl. BE.0] for a construction.

So the relation between N-maps different from $\partial^{CL} f$ and directional derivatives becomes of some interest and will be investigated now. Having any N-map, directional differentiability cannot be guaranteed, due to Thm. 5(i). Thus additional hypotheses are needed. For our purpose, we can use the simple contingent derivative on a line

$$C_{Line} f(x)(u) := \{v \mid v = \lim s_k^{-1}[f(x + s_k u) - f(x)] \text{ for some sequence } s_k \downarrow 0\} \tag{3.11}$$

which satisfies

$$C_{Line} f(x)(u) \subset Cf(x)(u) \subset \partial^{CL} f(x)(u) \tag{3.12}$$

and a mean-value statement, too:

Proposition 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous and $C_{Line} f(x + \theta u)(u) \neq \emptyset \ \forall \theta \in (0, 1)$.*
If $f(x + u) < f(x) + c$ then there is some $\theta \in (0, 1)$ such that $\sup C_{Line} f(x + \theta u)(u) < c$.
If $f(x + u) > f(x) + c$ then there is some $\theta \in (0, 1)$ such that $\inf C_{Line} f(x + \theta u)(u) > c$.

Clearly, the second statement follows from the first one by passing to $-f$.

Proof It is equivalent to prove: If $\sup C_{Line} f(x + \theta u)(u) \geq c \ \forall \theta \in (0, 1)$ then $f(x + u) \geq f(x) + c$. Pick any $q < c$ and put $T_q = \{t \in [0, 1] \mid f(x + tu) \geq f(x) + q\ t\}$. T_q contains the origin and is closed by continuity. Hence $s = \max T_q$ exists. We show $s = 1$ by contradiction. Otherwise, there is some $\eta \in C_{Line} f(x + su)(u)$ with $\eta > q$. Hence there are $s_\nu \downarrow 0$ such that the differences $\eta_\nu := s_\nu^{-1}\ [\ f(\ (x + su) + s_\nu u\ ) - f(x + su)\ ]$ fulfill $\eta = \lim \eta_\nu$. The latter implies $\eta_\nu > q$ for large ν and

$$f(x + su + s_\nu u) = s_\nu \eta_\nu + f(x + su) > s_\nu q + f(x + su) \geq s_\nu q + f(x) + q\ s = f(x) + (s + s_\nu)q.$$

So $s_\nu + s \in T_q$ contradicts to the maximality of s. Now $s = 1$ implies $f(x + u) \geq f(x) + q$, and the assertion follows via $q \to c$.

Proposition 2 *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^m)$ have a N-map at $\bar{x}$ such that $C_{Line} f(\bar{x} + tu)(u) \subset \mathcal{N}(\bar{x} + tu)(u)$ if $0 < t < t_0$ and $\|u\| = 1$. Then $f'(\bar{x}, u)$ exists for all u.*

Proof We write the N-map condition as

$$\mathcal{N}(x)(\bar{x} + tu)(u) \subset \frac{f(\bar{x} + tu) - f(\bar{x})}{t} + O(t)B \quad \text{for all } \|u\| = 1. \tag{3.13}$$

So we know that

$$C_{Line} f(\bar{x} + tu)(u) \subset \frac{f(\bar{x} + tu) - f(\bar{x})}{t} + O(t)B \quad \text{for all } \|u\| = 1. \tag{3.14}$$

Let u be fixed. If $f'(\bar{x}; u)$ does not exist, then this holds for some component of f, say the first one. Further, at least after some evident linear transformation of f, we may assume that $\bar{x} = 0, f(\bar{x}) = 0$ and there are sequences $t_k, \tau_k \downarrow 0$ with $t_k > \tau_k$ such that

$$\begin{array}{ll} \frac{f_1(\bar{x}+t_k u) - f_1(\bar{x})}{t_k} & \to \ \ 1 = \limsup_{t \downarrow 0} \frac{f_1(\bar{x}+tu) - f_1(\bar{x})}{t}, \\ \frac{f_1(\bar{x}+\tau_k u) - f_1(\bar{x})}{\tau_k} & \to -1 = \liminf_{t \downarrow 0} \frac{f_1(\bar{x}+tu) - f_1(\bar{x})}{t}. \end{array} \tag{3.15}$$

Put $\lambda_k = \tau_k + \frac{1}{2}(t_k - \tau_k)$ and $h_k = f_1(\bar{x} + \lambda_k u)$. Then, for certain $k \to \infty$, it holds

$$\lim \frac{f_1(\bar{x} + t_k u) - h_k}{t_k - \lambda_k} \geq 2 \text{ if } h_k \leq 0 \quad \text{and} \quad \lim \frac{h_k - f_1(\bar{x} + \tau_k u)}{\lambda_k - \tau_k} \geq 2 \text{ if } h_k \geq 0.$$

We regard the second case. It yields, for certain $k \to \infty$, $f_1(\bar{x} + \lambda_k u) - f_1(\bar{x} + \tau_k u) > \frac{3}{2}(\lambda_k - \tau_k)$. So Prop. 1 ensures the existence of some $\theta_k \in (\tau_k, \lambda_k)$ such that

$$inf\ C_{Line} f_1(\bar{x} + \theta_k u)(\lambda_k - \tau_k)u > \frac{3}{2}(\lambda_k - \tau_k).$$

Thus, it follows $\inf C_{Line} f_1(\bar{x}+\theta_k u)u > \frac{3}{2}$, a contradiction to the consequence $C_{Line} f_1(\bar{x}+\theta_k u)u \subset [-1,\ 1]$ of (3.14) and (3.15).

If $\partial^{CL} f$ is a N-map, then (3.12) ensures that Prop. 2 can be applied. So the existence of $f'(\bar{x}; u)$ follows from this proposition, too.

3.3 Newton maps for selections and projections

Now we intend to use Thm. 4 for dealing with inclusions $0 \in F(x)$ where $F : X \rightrightarrows Y$. We want to solve them again via certain linear auxiliary problems which modify (3.3), namely

$$\begin{array}{c} f_k + A_k(x_{k+1} - x_k) = 0, \quad A_k \in \mathcal{N}(x_k), \quad (k \geq 0); \\ \text{having } x_{k+1}, \text{ select some } \ f_{k+1} \in F(x_{k+1}), \end{array} \tag{3.16}$$

where $x_0 \in X$ and $f_0 \in F(x_0)$ are given and $\emptyset \neq \mathcal{N}(x) \subset Lin(X, Y)$. Once more, we ask for superlinear convergence (3.1) to a zero $\bar{x}$ of F. Accordingly, all N-maps below are regarded as N-maps *at* $\bar{x}$. In addition, we suppose

(S) All f_{k+1} are uniquely defined by x_{k+1} $(k \geq 0)$.

Then there is a selection *function* $f(.) \in F(.)$, continuous or not, such that $f(\bar{x}) = 0$ and $f_k = f(x_k)$ holds for all steps, and method (3.16) coincides with (3.3), i.e., $f(x_k) + A_k(x_{k+1} - x_k) = 0$, $A_k \in \mathcal{N}(x_k)$. Hence we obtain a trivial but useful

Proposition 3 *Supposing* (S)*, the method (3.16) satisfies the convergence condition (3.1) if and only if so does (3.3) for some selection function $f \in F$ which is defined near $\bar{x}$ and fulfills $f(\bar{x}) = 0$.*

So we may apply Thm. 4 which asks for a regular N-map of f, requires small $\|x_0 - \bar{x}\|$ and the conditions (3.4) and (3.5).

In order to *find* a suitable selection f along with a Newton map, F should be sufficiently simple. In particular, (3.6) tells us that $f \in F$ has to satisfy

$$L_1\|x - \bar{x}\| \ \leq \ \|f(x)\| \ \leq \ L_2\|x - \bar{x}\| \text{ with certain constants } L_1, L_2 \text{ and for } x \text{ near } \bar{x},$$

which implies

$$\operatorname{dist}(0, F(x)) \leq L_2\|x - \bar{x}\|. \tag{3.17}$$

Thus F is necessarily Lipschitz l.s.c. at $(\bar{x}, 0)$. Then the Euclidean projections of the origin onto $F(x)$, i.e. $f(x) \in \operatorname{argmin}_{y \in F(x)} \|y\|$, are interesting candidates whenever they exist. It is even hard to find a better f if (3.17) is valid and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ is polyhedral, i.e., if $\operatorname{gph} F$ is the union of a finite number of convex polyhedrons since f is piecewise linear and continuous near $\bar{x}$.

Applied to a generalized equation $0 \in F(x) = h(x) + M(x)$ the selection becomes $f = h + m$ with $m \in M$. Let $\mathcal{N}h$ be any N-map for h. Then the existence of a N-map $\mathcal{N}f$ implies the existence of a N-map for m (and vice versa) namely $\mathcal{N}m = \mathcal{N}f - \mathcal{N}h$ defined by the sets $\{A_1 - A_2 \mid A_1 \in \mathcal{N}f(x), A_2 \in \mathcal{N}h(x)\}$. Hence we should again look for a selection $m \in M$ with a simple N-map. Recalling the projections onto $F(x)$, our candidates are the elements $m(x) \in \operatorname{argmin}_{y \in M(x)} \|y - h(x)\|$. If $h \in PC^1$ and $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^p$ is polyhedral then, again under (3.17) and with Euclidean norm, m and $h + m$ are PC^1-functions near $\bar{x}$. Thus, aiming at superlinear local convergence, various generalized Newton methods can be applied, even Prop. 10 if $p = n$. The remaining question of weak N-regularity for $\mathcal{N}h + \mathcal{N}m$ turns into a regularity condition for $h + m$ at $\bar{x}$.

4 Nonlinear Auxiliary Problems for (Generalized) Equations

In this section, the concept of Newton maps and the related convergence analysis will be extended to nonlinear, multivalued approximations of the function in generalized equations

$$0 \in \Gamma(x) = f(x) + M(x) \quad \text{where } f: X \to Y \text{ and } M: X \rightrightarrows Y,\ M(x) \neq \emptyset\, \forall x. \tag{4.1}$$

We follow the description of Example 3 in §2.4 and solve, for given x, the inclusion

$$0 \in \Sigma(x', x) = f(x) + Gf(x)(x'-x) + M(x'), \qquad Gf(x): X \rightrightarrows Y\ \forall x \in X. \tag{4.2}$$

By $\mathcal{S}(x)$, we denote again the solution set of (4.2). The following results cover those obtained (by similar ideas) in [31, §10.3] for equations $f(x) = 0$, i.e. $M \equiv \{0\}$, and for the same notion of feasibility as defined below. Like in [31], our Propositions 6 and 7 will deal with approximate solutions of (4.2) without supposing $\mathcal{S}(x) \neq \emptyset$. However, in the general model, particular properties of M come into play. To complete the picture, we also add the characterization of the imposed conditions for the case of $M \equiv \{0\}$, known from [31]. Next we define generalized Newton maps for generalized equations.

Definition 5 We call Σ a *generalized Newton map* for $\Gamma = f + M$ at $\bar{x}$ if $Gf(x)(0) = \{0\}$, $Gf(x)(u) \neq \emptyset\ \forall x, u$ and if there exist positive c, δ and an o-type function $o(.)$ such that

$$\begin{array}{lll} \text{(CI)} & \|v\| \geq c\|u\| \quad \forall v \in f(\bar{x}) + Gf(x)(u) + M(u+\bar{x}) & \\ \text{(CA)} & f(x) + Gf(x)(u) \subset f(\bar{x}) + Gf(x)(x - \bar{x} + u) + o(x-\bar{x})B & \end{array} \quad \forall x \in \bar{x} + \delta B, \forall u \in X.$$

Condition (CI) claims some injectivity of $u \mapsto f(\bar{x}) + Gf(x)(u) + M(u+\bar{x})$ or, in other words,

the mappings $\Phi(x) := (f(\bar{x}) + Gf(x)(.) + M(. + \bar{x}))^{-1}$ of type $Y \rightrightarrows X$
are globally u.L. at $(0,0)$ (with uniform rank for x near $\bar{x}$).

(CI) is stronger, but comparable with condition (G1) in section 6 below. (CA) means some particular approximation. These conditions are closely related to superlinear convergence.

Proposition 4 *Let Σ be a generalized Newton map for $\Gamma = f + M$ at a zero $\bar{x}$ of Γ. Then $0 \in \Sigma(x', x)$ implies $\|x' - \bar{x}\| \leq c^{-1}\, o(x - \bar{x})$ for x near $\bar{x}$.*
So *all* iterates $x_{k+1} \in \mathcal{S}(x_k)$ (if they exist) fulfill $\|x_{k+1} - \bar{x}\| \leq c^{-1}\, o(x_k - \bar{x})$ for x_0 near $\bar{x}$.

Proof Let $0 \in \Sigma(x', x)$. Setting $u = x' - x$ in (CA) we have

$$\begin{aligned} 0 \in f(x) + Gf(x)(x'-x) + M(x') &\subset f(\bar{x}) + Gf(x)(x - \bar{x} + x' - x) + M(x') + o(x-\bar{x})B \\ &= f(\bar{x}) + Gf(x)(x' - \bar{x}) \qquad\quad + M(x') + o(x-\bar{x})B. \end{aligned}$$

So some $v \in f(\bar{x}) + Gf(x)(x'-\bar{x}) + M(x')$ belongs to $o(x-\bar{x})B$. Condition (CI) yields with $u = x' - \bar{x}$ via $v \in f(\bar{x}) + Gf(x)(u) + M(u + \bar{x})$ the inequality $\|v\| \geq c\|u\|$. This ensures $c\,\|u\| \leq \|v\| \leq o(x - \bar{x})$. Hence $c\,\|x' - \bar{x}\| \leq \|v\| \leq o(x-\bar{x})$ presents us already the claimed estimate for the iterates x_{k+1} if $\|x_0 - \bar{x}\|$ is small.

The simpler condition (CA) and convergence of approximate solutions*
Next we suppose $Gf(x)(0) = \{0\}\ \forall x$ and put $u = \bar{x} - x$ in (CA). This leads us to

$$\text{(CA)*} \qquad f(x) + Gf(x)(\bar{x} - x) \subset f(\bar{x}) + o(x-\bar{x})B \quad \forall x \in \bar{x} + \delta B.$$

Since $\Sigma(\bar{x}, x) = f(x) + Gf(x)(\bar{x} - x) + M(\bar{x})$, (CA)* yields, after adding $M(\bar{x})$ on both sides,

$$\Sigma(\bar{x}, x) \subset f(\bar{x}) + M(\bar{x}) \ + o(x-\bar{x})B = \Sigma(\bar{x}, \bar{x}) \ + o(x - \bar{x})B. \tag{4.3}$$

If $\Sigma(\bar{x}, \bar{x})$ contains only the origin (as for $M \equiv \{0\}$), we obtain $\Sigma(\bar{x}, x) \subset o(x - \bar{x})B$. Thus, for arbitrary $\alpha > 0$ and small $\|x - \bar{x}\|$ such that $o(x - \bar{x}) \leq \alpha\|x - \bar{x}\|$, the point $x' = \bar{x}$ solves

$$\emptyset \neq \Sigma(x', x) \cap \alpha\|x - \bar{x}\|B. \tag{4.4}$$

The latter becomes of numerical interest, whenever the norm can be replaced by an observable quantity $\phi(x)$ which fulfills

$$\phi(x) \geq L_1\|x-\bar{x}\| \quad \text{for fixed } L_1 > 0 \text{ and } x \text{ near } \bar{x}. \tag{4.5}$$

Related to the projection discussed after Prop. 3, we identify ϕ with the error measure

$$\phi(x) = \inf\{\, \|y\| \mid y \in f(x) + M(x)\} \tag{4.6}$$

and study also *algorithm $ALG(\alpha)$*, $\alpha > 0$, defined by the following iterations:

$$\text{for given } x_k, \text{ find } x_{k+1} \text{ such that } \emptyset \neq \Sigma(x_{k+1}, x_k) \cap \alpha\, \phi(x_k) B. \tag{4.7}$$

First we discuss the special case $\Sigma(\bar{x}, \bar{x}) = \{0\}$.

Proposition 5 *Let Σ be a generalized Newton map at a zero $\bar{x}$ of $\Gamma = f + M$ with $\Sigma(\bar{x}, \bar{x}) = \{0\}$. Suppose that either $M \equiv \{0\}$ or G fulfills*

$$Gf(x)(-u) = -Gf(x)(u) \tag{4.8}$$

Then, the estimate (4.5) holds for $L_1 = \frac{1}{2}c$ and ϕ in (4.6).

Thus, given $\alpha > 0$, (4.7) is solvable for x_k near $\bar{x}$. In the proof, we may replace (CA) by the (formally) weaker condition (CA)*. Note that (4.8) holds, e.g., for $Gf = Tf$ and $Gf = \mathcal{N}f$.

Proof Put $u = x - \bar{x}$, choose any $\varepsilon > 0$ and select some $m(x) \in M(x)$ with $\|f(x) + m(x)\| \leq \phi(x) + \varepsilon$. Note that

$$\text{(CA)* says} \quad f(\bar{x}+u) - f(\bar{x}) + Gf(x)(-u) \subset o(u)B, \tag{4.9}$$

$$\text{(CI) implies } \|v\| \geq c\|u\| \quad \forall v \in f(\bar{x}) + m(u+\bar{x}) + Gf(x)(u). \tag{4.10}$$

Case 1: Using (4.8), we thus obtain

$$f(\bar{x}+u) - (f(\bar{x}) + Gf(x)(u)) \subset o(u)B \text{ and } \|v\| \geq c\|u\| \text{ if } v \in f(\bar{x}) + Gf(x)(u) + m(u+\bar{x}).$$

Selecting any $g \in f(\bar{x}) + Gf(x)(u)$, so $\|f(\bar{x}+u) - g\| \leq o(u)$ and $\|g + m(\bar{x}+u)\| \geq c\|u\|$ are true. Therefore,

$$\|f(\bar{x}+u) + m(\bar{x}+u)\| \geq \|g + m(\bar{x}+u)\| - \|f(\bar{x}+u) - g\| \geq c\|u\| - o(u).$$

For small $\|u\| = \|x - \bar{x}\|$, this implies: $\phi(x) + \varepsilon \geq \|f(x) + m(x)\| \geq \frac{1}{2}c\|x - \bar{x}\|$.
Since $\varepsilon > 0$ was arbitrary, the claimed inequality follows.
Case 2: Using $M \equiv \{0\}$, now (CI) with $f(\bar{x}) = 0$ also ensures $\|v\| \geq c\|u\|$ $\forall v \in Gf(x)(-u)$. Along with (4.9) this yields $\phi(x) = \|f(x)\| \geq \frac{1}{2}c\|x - \bar{x}\|$ like above.

Definition 6 (feasibility) We call the triple $(f, \Sigma, \bar{x})$ *feasible* if, for each $q \in (0,1)$, there are positive r and α such that, whenever $\|x_0 - \bar{x}\| \leq r$, the auxiliary problems (4.7) of ALG(α) remain solvable and *all* solutions x_{k+1} satisfy the estimate $\|x_{k+1} - \bar{x}\| \leq q\,\|x_k - \bar{x}\|$.

If only the *existence* of solutions x_{k+1} of (4.7) satisfying $\|x_{k+1} - \bar{x}\| \leq q\,\|x_k - \bar{x}\|$ is ensured, we call the triple *weakly feasible*.

Unfortunately, under weak feasibility, the drawbacks discussed in Remark 2 remain.

Proposition 6 *(approximate solutions) Suppose the hypotheses of Prop. 5 and let, for some constant L_2 and small $\|x - \bar{x}\|$, also an* upper estimate $\phi(x) \leq L_2\|x - \bar{x}\|$ *be valid. Then $(f, \Sigma, \bar{x})$ is feasible. Given $q \in (0,1)$, it suffices to choose $\alpha \in (0, \frac{1}{2}c\,L_2^{-1}q)$ in order to satisfy the requirements of Def. 6.*

Evidently, if $M \equiv \{0\}$ then the needed conditions for $\phi = \|f\|$ coincide with condition (3.6) in Thm. 4 and the upper estimate for ϕ is trivially true whenever $f \in C^{0,1}$.

Proof Solutions exist due to Prop. 5. To verify the estimate of Def. 6, assume that $y \in \Sigma(x', x)$ holds with some $y \in \alpha\,\phi(x)B$. Setting $u = x' - x$ in (CA) we have

$$\begin{aligned} y \in f(x) + Gf(x)(x'-x) + M(x') &\subset f(\bar{x}) + Gf(x)(x - \bar{x} + x' - x) + M(x') + o(x-\bar{x})B \\ &= f(\bar{x}) + Gf(x)(x' - \bar{x}) \qquad\quad + M(x') + o(x - \bar{x})B. \end{aligned}$$

So $y = v + h$ holds with some $v \in f(\bar{x}) + Gf(x)(x' - \bar{x}) + M(x')$ and $h \in o(x - \bar{x})B$. Condition (CI) yields, with $u = x' - \bar{x}$ via $v \in f(\bar{x}) + Gf(x)(u) + M(u + \bar{x})$, the inequality $\|v\| \geq c\|u\|$.
Thus, $c\,\|u\| \leq \|y - h\| \leq \|y\| + \|h\| \leq \alpha\phi(x) + o(x - \bar{x}) \leq \alpha L_2\|x - \bar{x}\| + o(x - \bar{x})$.
With small α such that $c^{-1}\alpha L_2 < \frac{1}{2}q$ and sufficiently small $\|x - \bar{x}\|$ such that $c^{-1}o(x - \bar{x}) \leq \frac{1}{2}q\|x - \bar{x}\|$, we so obtain the required estimate $\|x' - \bar{x}\| \leq q\|x - \bar{x}\|$.

The assumption $\Sigma(\bar{x}, \bar{x}) = \{0\}$ used in the Propositions 5 and 6 requires $M(\bar{x}) = \{-f(\bar{x})\}$ and is not typical for generalized equations. Deleting it, one can still prove a weaker statement.

Proposition 7 *Let f, M, G fulfill all assumptions of Prop. 6 except for $\Sigma(\bar{x}, \bar{x}) = \{0\}$. Suppose also a pointwise Lipschitz condition for f,*

$$\|f(x) - f(\bar{x})\| \le L_f \|x - \bar{x}\| \quad \textit{holds for some constant } L_f \textit{ and } x \textit{ near } \bar{x}. \tag{4.11}$$

Then the triple $(f, \Sigma, \bar{x})$ is weakly feasible with the same $q - \alpha-$ relation as in Prop. 6.

Proof Knowing that $-f(\bar{x}) = \bar{m} \in M(\bar{x})$ we replace M by any mapping

$$\tilde{M}(x) = M(x) \cap (\bar{m} + \tilde{L}\|x - \bar{x}\|B) \ \text{ where } \tilde{L} > L_2 + L_f \ \ (L_2 \text{ from Prop. 6})$$

and keep G unchanged. In (4.2), we so obtain a submapping of Σ, namely

$$\tilde{\Sigma}(x', x) = f(x) + Gf(x)(x' - x) + \tilde{M}(x').$$

Now $\tilde{\Sigma}(\bar{x}, \bar{x}) = f(\bar{x}) + \tilde{M}(\bar{x}) = \{0\}$ holds true. Next we show that f, G and $\tilde{M}$ satisfy all hypotheses of Prop. 6 and that even

$$\tilde{\phi}(x) := \inf_{y \in f(x) + \tilde{M}(x)} \|y\| \ = \ \phi(x) \ \text{ for } x \text{ near } \bar{x} \tag{4.12}$$

is valid. The suppositions (CA), (CA)* are still satisfied since f and G do not change. Because of $\tilde{M} \subset M$, also (CI) and $\tilde{\phi}(x) \ge \phi(x)$ hold for the submapping.

To verify $\tilde{\phi}(x) \le \phi(x)$ for x near $\bar{x}$, let $\varepsilon > 0$ and choose some $m(x) \in M(x)$ such that $\|f(x) + m(x)\| \le \phi(x) + \varepsilon$. Then we also obtain

$$\|f(x) + m(x)\| \ \le \ \varepsilon + L_2\|x - \bar{x}\|. \tag{4.13}$$

If $m(x) \in \tilde{M}(x)$, it follows $\tilde{\phi}(x) \le \phi(x) + \varepsilon$, the inequality we want to show. Assume $m(x) \notin \tilde{M}(x)$. Because of $m(x) \in M(x)$ this implies

$$m(x) \notin -f(\bar{x}) + \tilde{L}\|x - \bar{x}\|B, \ \ \textit{i.e.,} \ \ \|f(\bar{x}) + m(x)\| > \tilde{L}\|x - \bar{x}\|. \tag{4.14}$$

Recalling $-f(\bar{x}) = \bar{m} \in M(\bar{x})$ and $f(\bar{x}) + \bar{m} = 0$, so (4.13) and (4.14) yield (after adding zero)

$$\|f(x) - f(\bar{x}) + m(x) - \bar{m}\| = \|f(x) + m(x)\| \le \varepsilon + L_2\|x - \bar{x}\| \ \text{ and } \ \|m(x) - \bar{m}\| > \tilde{L}\|x - \bar{x}\|.$$

This induces (via 4.11) for each x near $\bar{x}$ and all $\varepsilon > 0$,

$$\begin{aligned} \tilde{L}\,\|x - \bar{x}\| &< \|m(x) - \bar{m}\| \le \|f(\bar{x}) - f(x)\| + \|f(x) - f(\bar{x}) + m(x) - \bar{m}\| \\ &\le L_f\,\|x - \bar{x}\| + \varepsilon + L_2\,\|x - \bar{x}\| = \varepsilon + (L_f + L_2)\,\|x - \bar{x}\|, \end{aligned}$$

which contradicts $\tilde{L} > L_f + L_2$. Hence, the choice of $\tilde{L}$ ensures that the mappings $f, G, \tilde{M}$, fulfill all hypotheses of Prop. 6, including $\tilde{\Sigma}(\bar{x}, \bar{x}) = \{0\}$ from Prop. 5. Even

$$\tfrac{1}{2}c\|x - \bar{x}\| \le \tilde{\phi}(x) = \phi(x) \le L_2\|x - \bar{x}\| \quad \text{for } x \text{ near } \bar{x}$$

is true. In consequence, the triple $(f, \tilde{\Sigma}, \bar{x})$ is feasible. By Prop. 6, we also know that, given $q \in (0, 1)$, it suffices to choose $\alpha \in (0, \frac{1}{2}c\,L_2^{-1}q)$ in Def. 6 in order to obtain the existence and convergence $\|x_{k+1} - \bar{x}\| \le q\|x_k - \bar{x}\|$ of all iterates satisfying

$$\emptyset \ne \tilde{\Sigma}(x_{k+1}, x_k) \cap \alpha\,\phi(x_k)B \ \text{ if } \|x_0 - \bar{x}\| \text{ is small enough.}$$

Obviously, these x_{k+1} are particular solutions of ALG(α), assigned to Σ.

It remains a future task to find a practically more relevant characterization of the existing points x_{k+1} (like in §5.2) in terms of f, G and M.

The imposed conditions for locally Lipschitz functions

Let $f \in C^{0,1}(X,Y)$ and $M \equiv \{0\}$. For this case, Prop. 6 reduces to Thm. 10.7 in [31].

We will speak of a *standard setting of* Gf, if

$$Gf(x)(u) = \mathcal{N}(x)u := \{Au \mid A \in \mathcal{N}(x)\} \text{ where } \emptyset \neq \mathcal{N}(x) \subset \text{Lin}(X,Y). \tag{4.15}$$

Then (CA)* is just the N-map condition (3.4) and coincides with (CA) by the following argumentation. For linear functions $A \in \mathcal{N}(x)$, (CA)* yields $f(x) - f(\bar{x}) + A(\bar{x} - x) \subset o(x - \bar{x})B$ which ensures (CA) after adding Au to both sides. Moreover, for $X = Y = \mathbb{R}^n$, weak N-regularity and (CI) coincide under (4.15), and we are in the framework of §3.1.

The next example shows that $(CA)^* \Rightarrow (CA)$ may fail in non-standard settings.

Example 4 If $\bar{x} = f(\bar{x}) = 0$ and $Gf(x)(u) = f(x+u) - f(x)$, then (CA) requires $f(x) + f(x+u) \in f(2x+u) + o(x - \bar{x})B$ which usually fails to hold. (CA)* turns into $0 \in o(x - \bar{x})B$.

Nevertheless, for several generalized derivatives, (CA) and (CA)* coincide, and the imposed conditions can be characterized in detail for $M \equiv \{0\}$.

Theorem 7 [31, Thm. 10.8] (Condition (CA)). *Let $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$. Then for all standard settings of Gf and for $Gf(x)(u) = Tf(x)(u)$, $Gf(x)(u) = Cf(x)(u)$ and $Gf(x)(u) = \{f'(x;u)\}$, (CA) and (CA)* are equivalent and $Gf(x)(0) = \{0\}$ as well as $Gf(x)(u) \neq \emptyset$ hold true. For standard settings and $Gf = Tf$, also (4.8) is valid.*

Theorem 8 [31, Thm. 10.9] (Condition (CI)). *Suppose that $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ and $f(\bar{x}) = 0$.*

(i) Let $Gf(x)(u) = Tf(x)(u)$. Then (CI) holds at $x = \bar{x}$ $\Leftrightarrow$ (CI) holds for x near $\bar{x} \Leftrightarrow f$ is strongly regular at $(\bar{x}, 0)$. Under (CI), condition (CA) holds true if and only if $(f, \Sigma, \bar{x})$ is feasible.

(ii) Let $Gf(x)(u) = \partial^{CL} f(x)u$. Then (CI) holds at $\bar{x}$ $\Leftrightarrow$ (CI) holds for x near $\bar{x}$ $\Leftrightarrow$ $\partial^{CL} f(\bar{x})$ is non-singular. This condition is stronger than strong regularity.

(iii) Let $Gf(x)(u) = Cf(x)(u)$. Then (CI) holds at $\bar{x}$ $\Leftrightarrow$ f^{-1} is locally u.L. at $(0, \bar{x})$.

(iv) Let $Gf(x)(u) = \{f'(x;u)\}$, provided that directional derivatives exist near $\bar{x}$. Then, under strong regularity of f, (CA) holds true if and only if $(f, \Sigma, \bar{x})$ is feasible. Under pseudo-regularity, (CI) is satisfied *for x near $\bar{x}$ [by Thm. 1].*

Due to Thm. 7, one may replace (CA) by (CA)* everywhere in Thm. 8. Summarizing, the conditions (CI) and (CA)* are, at least for ALG(α) in the context of $C^{0,1}$ functions and nonlinear approximations, similarly crucial as N-regularity and N-maps in Thm. 4.

Let us finish this section by referring to related literature in the case of equations. Strong regularity or surjectivity of f were not explicitly required in our analysis of procedure (4.7). These are realistic assumptions for equations arising from control problems. Basic ideas on this topic (where the correct choice of related function spaces is important, too) can be found, e.g., in [21, 58, 59].

In [7], for $f \in C^{0,1}(X,Y)$ and $M \equiv \{0\}$, the iteration scheme (4.7) is replaced by a non-monotone path search, which leads to a convergence result similar to that of Prop. 6; this was successfully implemented for semi-infinite programs and Nash equilibrium problems.

The “Inexact Nonsmooth Newton Method” [17, 7.2.6] is algorithm ALG(α), specified to bounded sequences $\alpha = \alpha_k$, $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ and $M \equiv \{0\}$. Condition (CA)* there defines a so-called *Newton approximation* of f, and $Gf(x)$ consists of possibly nonlinear functions $u \mapsto \mathcal{A}(x,u)$ which replace Au for $A \in \mathcal{N}(x)$. The convergence theorem [17, Thm. 7.2.8] does not use (CA); it requires that all $\mathcal{A}(x, .) : \mathbb{R}^n \to \mathbb{R}^n$ are strongly regular at (0,0) with uniform rank. In particular, this claims the existence of solutions for $0 \in \Sigma(\,.\,, x_k)$.

5 Modified Successive Approximation for Generalized Equations

In this section, we discuss convergence of Newton methods for generalized equations under pseudo-regularity and propose a possible way to overcome the drawback explained in Remark 2 (cf. §2.4). For that purpose, we study zeros of closed mappings $F : X \rightrightarrows Y$ after small nonlinear variations near $(x_0, y_0) \in \operatorname{gph} F$ for the basic model

$$g(x) \in F(x),\ x \in \Omega \quad \text{and their solution sets } \tilde{S}(g) \subset \Omega, \tag{5.1}$$

where Ω is some ball around x_0 and $g : \Omega \to Y$ is Lipschitz on Ω.

Often, F is the right-hand side $f + M$ of another generalized equation and $g = f - f^{(k)}$ stands for the approximation error at x_k as in (1.2). Alternatively, $F(x)$ may coincide with $f(x, t_0) + M(x)$ in parametric settings with $g = f(., t_0) - f(., t)$.

Particular inclusions (5.1) and assigned Newton-type methods have been studied already in [49] and [27], and its parametric form $g(x, t) \in F(x)$ was the subject of Robinson's pioniering paper [51] for C^1-functions g under *strong regularity* of F. It uses the obvious fact that (5.1) coincides with the fixed point condition $x \in F^{-1}(g(x)),\ x \in \Omega$.

To outline the key idea of [51] (which can be easily extended to $g \in C^{0,1}$) and to compare it with the case of pseudo-regular F, now we are going to analyze the convergence of a sequence generated by successive approximation to a solution x^* of (5.1). Let us suppose that for $(x_0, y_0) \in \operatorname{gph} F$,

$$\begin{array}{l} F \text{ is strongly or pseudo-regular at } (x_0, y_0) \text{ with rank } L \text{ and assigned closed} \\ \text{balls } U, V \text{ around } x_0, y_0 \text{ in (D1); let } \alpha = \|g(x_0) - y_0\|,\ \beta = Lip\,(g, \Omega),\ U \subset \Omega \\ \text{and } L\beta < 1. \text{ Suppose also that } \alpha \text{ and } \beta \text{ are small enough such that } g(U) \subset V. \end{array} \tag{5.2}$$

(R1) Under strong regularity in (5.2), the intersection $F_U^{-1} := F^{-1} \cap U$ is single-valued on V, and one has $\|F_U^{-1}(y) - F_U^{-1}(y')\| \le L\,\|y - y'\|\ \forall y, y' \in V$. Hence, the function

$$x \in U \mapsto \phi(x) := F^{-1}(g(x)) \cap U \tag{5.3}$$

has Lipschitz rank $\theta = L\beta < 1$ on U, and a fixed point $x^* = \phi(x^*) \in \tilde{S}(g) \subset U$ can be obtained and estimated via the well-known successive approximations $x_{k+1} = \phi(x_k)$. Since $x_0 \in F^{-1}(y_0)$, the point $x_1 = \phi(x_0)$ fulfills $\|x_1 - x_0\| \le L\,\|g(x_0) - y_0\| = L\,\alpha$. Thus the estimates

$$\|x_{k+1} - x_k\| \le \theta\,\|x_k - x_{k-1}\| \qquad (k \ge 1) \tag{5.4}$$

and, in consequence (adding the distances),

$$\|x_{k+1} - x_0\| \le \|x_1 - x_0\|(1-\theta)^{-1} \le L\,\|g(x_0) - y_0\|(1-\theta)^{-1} = L\,\alpha(1-\theta)^{-1}\ (k > 1) \tag{5.5}$$

are valid. By (5.4), so convergence of the well-defined sequence $\{x_k\}_{k \ge 0}$ to some (unique) $x^* \in \tilde{S}(g) \cap (x_0 + L\,\alpha(1-\theta)^{-1}B)$ in the complete space X follows.

(R2) Under pseudo-regularity of F, these conclusions are similarly applicable, but ϕ in (5.3) is a *multifunction*. One obtains first the *existence* of some $x_1 \in \phi(x_0) = F^{-1}(g(x_0))$ with $\|x_1 - x_0\| \le L\|g(x_0) - y_0\| = L\alpha$. Further, as long as $(x_k, g(x_k)) \in U \times V$ and $x_k \in \phi(x_{k-1})$, some $x_{k+1} \in \phi(x_k)$ satisfying (5.4) *exists* again. This yields (5.4) and (5.5) for the given k. Using both estimates recursively, one easily sees that, once more for sufficiently small $\alpha = \|g(x_0) - y_0\|$, namely if

$$x_0 + L\,\alpha(1-\theta)^{-1}B \subset U \quad \text{and} \quad g(x_0 + L\,\alpha(1-\theta)^{-1}B) \subset V, \tag{5.6}$$

the supposed inclusions $(x_k, g(x_k)) \in U \times V$ remain true. Hence *there exists* a sequence $x_k \to x^* \in x_0 + L\,\alpha(1-\theta)^{-1}B$ satisfying $x_{k+1} \in \phi(x_k)$, (5.4) and (5.5). Finally, $x^* \in \tilde{S}(g)$ follows from $x_{k+1} \in \phi(x_k)$ and closedness of F and ϕ.

The arguments of (R1) and (R2) are known from the literature on implicit mappings and generalized Newton methods for generalized equations. In both situations, only the properties of F^{-1} are needed. It plays no role whether F itself is a continuous function, a fixed set, a "normal map" as in [51] or any other closed mapping.

In the second case, uniqueness of $x_{k+1} \in \phi(x_k)$ is no longer true and $\theta = L\beta$ is usually not explicitly known. So it may be hard to *find* some existing $x_{k+1} \in \phi(x_k)$ satisfying (5.4) in order to determine some $x^* \in \tilde{S}(g)$ numerically. If $\dim Y < \infty$, one can select $x_{k+1} \in \phi(x_k)$ with minimal distance to x_k. Otherwise, one could look for $x_{k+1} \in \phi(x_k)$ such that $\|x_{k+1} - x_k\| \le \operatorname{dist}(x_k, \phi(x_k)) + \varepsilon_k$ holds with appropriate $\varepsilon_k \downarrow 0$. This requires to adapt the above estimates which will be done below for $\varepsilon_k = \|g(x_k) - g(x_{k-1})\|$.

It is well known [28] that, under strong regularity, the successive approximation can be used to solve a C^1- equation $f(x) = 0$ by a *modified Newton method* $0 = f(x_k) + Df(x_0)\,(x_{k+1} - x_k)$ with derivative at x_0. Setting

$$F(x) = f(x_0) + Df(x_0)(x - x_0) \quad \text{and} \quad g(x) = f(x_0) + Df(x_0)(x - x_0) - f(x), \tag{5.7}$$

then x_{k+1} coincides with $\phi(x_k) = F^{-1}(g(x_k))$.

Under pseudo-regularity of f at $(x_0, f(x_0))$, (5.7) yields, via the estimates (5.4) and (5.5), the existence of a modified Newton sequence with the properties derived in (R2). The supposition $(x_0, y_0) \in \operatorname{gph} F$ means $F(x_0) = f(x_0) = y_0$, and we require that $\alpha = \|g(x_0) - y_0\| = \|f(x_0)\|$ is small. Assumptions on small $\|f(x_0)\|$ instead of small $\|x_0 - \bar{x}\|$ for some (unknown) solution $\bar{x}$ characterize Kantorovich-type statements for Newton's method (also called semi-local or Kantorovich-Newton methods). They are particularly useful, provided one knows how small $\|f(x_0)\|$ must be chosen. Basic ideas and estimates in [28, sect. XVIII] were the key for many later papers and results on numerical methods for solving equations. Classical statements of this type can be found, e.g., in [5] and [44]. To illustrate the interplay of the related hypotheses and for convenience of the reader let us add

Theorem 9 [44]. *Let $f \in C^{1,1}(X, Y)$, $\Omega \subset X$ be open and convex, $x_0 \in \Omega$ and $Lip\,(Df, \Omega) \le \beta$. Suppose that $Df(x_0)^{-1}$ exists with norm L and that $h := L\,\beta\,\eta \le \frac{1}{2}$ holds with $\eta = \|Df(x_0)^{-1} f(x_0)\|$. Finally, put $\delta^* = \frac{1}{L\beta}(1 - \sqrt{1 - 2h})$, $\delta^{**} = \frac{1}{L\beta}(1 + \sqrt{1 - 2h})$ and suppose that $S := x_0 + \delta^* B \subset \Omega$. Then, the (usual) Newton iterates are well defined, lie in S and converge to a zero x^* of f which is unique in $\Omega \cap [x_0 + \operatorname{int} \delta^{**} B]$. If even $h < \frac{1}{2}$ the convergence is quadratic.*

For an overview on the state of art until 2010, various deeper results (including variational inequalities, too) and more references, we refer to [3]. In what follows, we shall modify (5.7) for generalized equations and propose a selection rule for $x_{k+1} \in \phi(x_k)$, based on the errors ε_k above. Note our approach differs from that in the recent papers [1,8,9,15,16] handling generalized equations, too.

5.1 The approximation scheme and its properties

Here we follow the study in [31, §4.1]. Though (5.1) is our main inclusion, we construct elements $x_k \in X$ and $v_k \in Y$, independently of any function $g : X \to Y$. This generalizes the iterations $x_{k+1} \in \phi(x_k)$ motivated above, and allows additional applications like in the setting S2 below.

Given any mapping $F : X \rightrightarrows Y$, $(x_0, y_0) \in \operatorname{gph} F$ and $v_0 \in Y$ we consider the

Process $\mathbf{P}(\lambda, \beta, x_0, y_0, v_0)$. Let $\lambda > 0$, $\beta > 0$. For describing the initial step at (x_0, v_0) like the others, put $v_{-1} = y_0$. Hence $x_0 \in F^{-1}(v_{-1}) = F^{-1}(y_0)$. Beginning with $k = 0$,

$$\begin{array}{c} \text{find } x_{k+1} \in F^{-1}(v_k) \text{ with } \|x_k - x_{k+1}\| \le \operatorname{dist}(x_k, F^{-1}(v_k)) + \lambda \|v_k - v_{k-1}\| \\ \text{and choose any } v_{k+1} \text{ such that } \|v_{k+1} - v_k\| \le \beta\,\|x_{k+1} - x_k\|. \end{array} \tag{5.8}$$

Clearly, x_{k+1} is an approximate projection of x_k onto $F^{-1}(v_k)$ with error $\le \lambda \|v_k - v_{k-1}\|$. The process stops if and only if $F^{-1}(v_k) = \emptyset$, and it becomes stationary if $v_{k+1} = v_k$.

Particular settings for process $P(\lambda, \beta, x_0, y_0, v_0)$

- S1 In order to solve (5.1), put $v_k = g(x_k)$ and $\beta = \operatorname{Lip}(g, \Omega)$ as mentioned above.
- S2 To solve $H(x) \cap F(x) \ne \emptyset$, i.e., $0 \in -H(x) + F(x)$, for closed $H, F : X \rightrightarrows Y$, assume also $v_0 \in H(x_0)$ and select $v_{k+1} \in H(x_{k+1})$ with $\|v_{k+1} - v_k\| \le \beta \|x_{k+1} - x_k\|$. The latter is possible if H is pseudo-Lipschitz with rank β on $\Omega \times H(\Omega)$.
- S3 Let $F = \partial f$ be the subdifferential of a convex function $f : X = \mathbb{R}^n \to \mathbb{R}$. Put $g(x) = \beta x$. Then $x \in F^{-1}(g(x_k))$ means $\beta x_k \in \partial f(x)$ and $0 \in -\beta x_k + \partial f(x)$. Hence, given x_k, we require $x_{k+1} \in \operatorname{argmin}_{x \in X}(f(x) - \beta\langle x_k, x\rangle)$. A solution x^* of $g(x) \in F(x)$ now solves the conjugate problem $x^* \in \operatorname{argmin}_{x \in X}(f(x) - \beta\langle x^*, x\rangle)$.

S4 Let f and g be as in S3. Put $F(x) = \beta x + \partial f(x)$. Then $x \in F^{-1}(g(x_k))$ means $g(x_k) \in F(x)$ and $0 \in \beta(x - x_k) + \partial f(x)$. Hence, x_{k+1} minimizes the Moreau-Yosida approximation $f(x) + \frac{1}{2}\beta\|x - x_k\|^2$, and one has $g(x^*) \in F(x^*) \Leftrightarrow x^* \in \operatorname{argmin}_{x\in X} f(x)$. In this case, the algorithm minimizes f by a proximal point method.

Next, according to Remark 1, we take the same radius δ for the balls U and V around x_0 and y_0. Unfortunately, δ disappeared in the formulation (not in the proof) of [31, Thm. 4.2].

Theorem 10 [31, Thm. 4.2] (modified successive approximation). *Let $F : X \rightrightarrows Y$ be closed and pseudo-regular at $(x_0, y_0) \in \operatorname{gph} F$ with rank L and balls $U = x_0 + \delta B_X$, $V = y_0 + \delta B_Y$. Suppose that $\alpha := \|v_0 - y_0\|$ and β are small enough such that*

$$\theta := \beta(L + \lambda) < 1 \quad \text{and} \quad \alpha < \delta\,(1 - \theta)\,(\max\{1, L + \lambda\})^{-1}. \tag{5.9}$$

Then, $P(\lambda, \beta, x_0, y_0, v_0)$ generates convergent sequences $\{x_k\} \subset U$, $\{v_k\} \subset V$ such that

(i) The limit (x^*, v^*) belongs to $\operatorname{gph} F$, $x_k, x^* \in x_0 + (1-\theta)^{-1}(L+\lambda)\alpha B \subset U$, $v_k, v^* \in y_0 + (1-\theta)^{-1}\alpha B \subset V$, and

$$\|x_{k+1} - x_k\| \le \theta\,\|x_k - x_{k-1}\|, \quad \|v_{k+1} - v_k\| \le \theta\,\|v_k - v_{k-1}\| \quad \forall k \ge 1. \tag{5.10}$$

(ii) If $v_k = g(x_k)$ for all $k \ge 0$ in $P(\lambda, \beta, x_0, y_0, v_0)$ and $Lip(g, U) \le \beta$, then $g(x^*) \in F(x^*)$.

(iii) If $v_k \in H(x_k)$ for all $k \ge 0$ in $P(\lambda, \beta, x_0, y_0, v_0)$ and $H : X \rightrightarrows Y$ is closed as under S2, *then $v^* \in H(x^*) \cap F(x^*)$, hence $0 \in -H(x^*) + F(x^*)$.*

Remark 4 By an estimate in [1, p.181] (though some index is wrong), (5.10) for all x_k implies linear convergence $\|x_{k+1} - x^*\| \le q\,\|x_k - x^*\|$ with factor $q = \theta\,(1 - 2\theta)^{-1} < 1$ if $\theta < \frac{1}{3}$.

Remark 5 (Strongly regular F) Under strong regularity of F, we obtain that all x_{k+1} of our construction are uniquely defined by v_k and belong to U. In the situation S1, then $Lip(g, U) \le \beta$ implies that $x \mapsto \phi(x) := F^{-1}(g(x)) \cap U$ is a contraction which maps U into itself. Thus, x^* is also the unique fixed point of ϕ on U.

Remark 6 (Family of mappings) The theorem indicates only one set of assumptions which ensures the asserted estimates as well as $F^{-1}(v_k) \neq \emptyset$. So it is quite obvious that Thm. 10(ii) also holds for a family of mappings F_k and functions g_k with $\operatorname{Lip}(g_k,\, x_0 + \delta B) \le \beta$ as long as they fulfill the requirements concerning pseudo-regularity of F and the initial conditions (5.9) for g with $v_0 := g_k(x_0)$. Thus all estimates for x_k and $v_k = g_k(x_k)$ remain true. In particular, our assumptions for the initial point even hold with $\alpha_k := \|v_{k+1} - v_k\| \le \theta\|v_k - v_{k-1}\| < \alpha$. Hence the limits x^* and v^* exist again.

Remark 7 To simplify our applications, we put $\lambda = 1$ in $\mathrm{P}(\lambda, \beta, x_0, y_0, v_0)$ and require both $\beta < \frac{1}{2}(L + 1)^{-1}$ and $\alpha := \|v_0 - y_0\| < \frac{1}{2}\delta(L + 1)^{-1}$. Then (5.9) is satisfied with $\theta := \beta(L + 1) < \frac{1}{2}$, after which (i) presents the simpler estimates $\|x_k - x_0\| \le 2(L+1)\alpha < \delta$ and $\|v_k - y_0\| \le 2\alpha < \delta$. For the setting S1, we then have $\alpha = \|g(x_0) - y_0\|$ and may put $\beta = Lip(g, U)$.

5.2 Generalized equations under strong or pseudo-regularity

We are now going to consider the approach of §2.4 for solving $0 \in \Gamma(x)$ with closed $\Gamma : X \rightrightarrows Y$ via a method given by $\Sigma : X \times X \rightrightarrows Y$, where *$\Sigma(., x)$ is a continuous translation of $\Sigma(., \bar{x})$*.

With the following corollary of Thm. 10, both implicit mappings with parameter x and convergence of Newton sequences can be studied in a unified manner. The close connection between these two topics for generalized equations is a main subject in [16].

Corollary 1 *Suppose that $0 \in \Gamma(\bar{x}) \cap \Sigma(\bar{x}, \bar{x})$, let $\Sigma(.\,, \bar{x})$ be closed and pseudo regular at $(\bar{x}, 0)$ with rank L and let $\Sigma(x', x) = g_x(x') + \Sigma(x', \bar{x})$ hold with some* function *g_x satisfying*

$$\sup(g_x,\, \Omega_r) \le o(r) \quad \text{and} \quad Lip(g_x,\, \Omega_r) \le O(r), \quad \forall x \in \Omega_r := \bar{x} + rB. \tag{5.11}$$

Then one has:

(i) If $r > 0$ is sufficiently small and $x \in \Omega_r$, there is some $x' \in \Omega_r$ with $0 \in \Sigma(x', x)$ and $\|x' - \bar{x}\| \le 2(L+1)\,\|g_x(\bar{x})\|$.

(ii) For the method $0 \in \Sigma(x_{k+1}, x_k)$, there exist such iterates x_{k+1} for all k which satisfy $\|x_{k+1} - \bar{x}\| = o(x_k - \bar{x})$, provided that $\|x_0 - \bar{x}\|$ is sufficiently small.

Proof We apply Thm. 10(ii), based on pseudo regularity of $F = \Sigma(., \bar{x})$ at $(\bar{x}, 0)$ for $g = g_x$ with initial point $(x_0, v_0) := (\bar{x}, g_x(\bar{x}))$. If $x \in \Omega_r$ and r is small, then $(\bar{x}, g_x(\bar{x}))$ is arbitrarily close to $(\bar{x}, 0)$ and $Lip(g_x, \Omega_r)$ is arbitrarily small. So Thm. 10 and Remark 7 ensure assertion (i) with $x' = x^*$. By (5.11), it also holds $\|g_x(\bar{x})\| = o(x - \bar{x})$. Hence, $\|x' - \bar{x}\| = o(x - \bar{x})$. Identifying $x_k = x,\ x_{k+1} = x'$, so also (ii) is true.

This way we do not obtain the inclusion of (2.25), but at least

$$\emptyset \ne \mathcal{S}(x) \cap [\bar{x} + o(x - \bar{x})B] \quad \text{for } x \text{ near } \bar{x}. \tag{5.12}$$

Under *strong regularity* in place of pseudo regularity in Corollary 1, and again for r small enough, the solutions $x' \in \Omega_r$ for $0 \in \Sigma(x', x)$ are unique, and it follows stronger

$$\Omega_r \cap \mathcal{S}(x) \ \text{ is single-valued and contained in } \ \bar{x} + o(x - \bar{x})B \ \text{ for } x \text{ near } \bar{x}. \tag{5.13}$$

This tells us for the same method: If $r > 0$ and $\|x_0 - \bar{x}\|$ are small enough, then there exist unique iterates $x_{k+1} \in \Omega_r$ for all k, and they fulfill $\|x_{k+1} - \bar{x}\| = o(x_k - \bar{x})$.

Applications to generalized equations

For $\Gamma = f + M$, where M is closed, let us apply Corollary 1 to the two situations

Case 1: If $f = h \circ \gamma$, with $h \in C^1$ and $\gamma \in C^{0,1}$ as under (2.12) put

$$\Sigma(x', x) = \begin{cases} f(x) + Dh(\gamma(x))(\,\gamma(x') - \gamma(x)\,) + M(x') & \text{if } x \ne \bar{x} \\ f(x') \qquad\qquad\qquad\qquad\qquad\qquad\quad + M(x') & \text{if } x = \bar{x} \end{cases}$$

with $g_x(x') = f(x') - \ f(x) + Dh(\gamma(x))(\,\gamma(x') - \gamma(x)\,)$.

Case 2: If f is continuous and there is a PBA $\hat{\Sigma}$ for f near $\bar{x}$, put $\Sigma(x', x) = \hat{\Sigma}(x', x) + M(x')$ and $g_x(x') = \hat{\Sigma}(x', x) - \hat{\Sigma}(x', \bar{x})$.

In both situations, (5.11) holds true due to (2.20) since, adding the multivalued term $M(x')$, does not change the needed estimate. Also the proofs remain the same with or without M.

Finally, if we replace $\Sigma(x', x)$ by any mapping $\tilde{\Sigma}(x', x) = \tilde{g}(x') + \Sigma(x', x)$ such that $\tilde{g}$ fulfills (5.11), then we are in the context of inexact Newton methods and obtain, evidently, the same statements. Slight generalizations of the case $f \in C^1$ by passing to uniform strict differentiability as, e.g., in [16], ensure condition (5.11) [by definition] too.

A Kantorovich-type statement for generalized equations

Statement (ii) of Corollary 1 is a typical example for the message of Remark 2 in §2.4 since we used statement (i) [and Thm. 10] only for verifying the existence of the next iterates $x' = x_{k+1}$. In fact, under pseudo-regularity, all similar statements we found - of Kantorovich-type or not - only state the existence of such iterates, as far as strong regularity is not required; cf. e.g. [1,2,8,9,12,13]. Clearly, since the statements are correct, related sequences have been constructed in the proofs by different means. On the other hand, the assertion of Corollary 1 and similar statements of the papers just mentioned could be improved if appropriate solutions of $0 \in \Sigma(x_{k+1}, x_k)$ would be directly assigned to the iterates of process $\mathrm{P}(\lambda, \beta, x_0, y_0, v_0)$. Then *all such sequences* converge automatically in the corresponding manner whenever Thm. 10 can be applied. We obtain a Kantorovich-type statement since small $\|x_0 - \bar{x}\|$ for some solution $\bar{x}$ is not required.

In the following, we shall do this for generalized equations

$$0 \in f(x) + M(x) \ \text{ with } \ f = h \circ \gamma \ \text{ as under case 1 above} \tag{5.14}$$

by the help of a simple selection rule. Let us first consider the modified Newton method

$$0 \in f(x_k) + H_0(\,\gamma(x_{k+1}) - \gamma(x_k)\,) + M(x_{k+1}) \tag{5.15}$$

with fixed operator $H_0 := Dh(\gamma(x_0))$. If the set X_{k+1} of the related solutions has more than one element, select any $x_{k+1} \in X_{k+1}$ with

$$\|x_{k+1} - x_k\| \le \begin{cases} \text{dist}(x_k, X_{k+1}) + \| g(x_k) - g(x_{k-1}) \| & \text{if } k > 0 \\ \text{dist}(x_k, X_{k+1}) + \| g(x_0) - y_0 \| & \text{if } k = 0, \end{cases} \tag{5.16}$$
$$\text{where} \qquad g(x) = f(x_0) + H_0\,(\gamma(x) - \gamma(x_0)) - f(x) \ \text{ and } \ y_0 \in f(x_0) + M(x_0).$$

While the definition of the auxiliary problems (5.15) is well-known standard, the particular selection rule (5.16) is crucial for the following statement. Needless to say that possible numerical realizations depend essentially on the structure of X_{k+1}. Our Kantorovich-type assumptions require that $\|y_0\| = \|f(x_0) + m_0\|$ is small enough for some $m_0 \in M(x_0)$.

Notice that $O_h(r, x_0) := \sup_{x,x' \in x_0 + rB} \|Dh(\gamma(x)) - Dh(\gamma(x'))\|$ vanishes as $r \downarrow 0$ due to the hypotheses (5.14).

Proposition 8 *Let $\delta_0 > 0$, $L_0 > 0$, $x_0 \in X$ and $m_0 \in M(x_0)$ be given in such a way that*

(i) the mapping $\Gamma = f + M$ is pseudo-regular at $(x_0, y_0) = (x_0, f(x_0) + m_0)$ with rank L_0, balls $U = x_0 + \delta_0 B$, $V = y_0 + \delta_0 B$, and $L_\gamma = Lip\,(\gamma, x_0 + \delta_0 B) < \infty$,

(ii) some δ with $0 < \delta < c := \min\{1, \frac{\delta_0}{12(L_0+1)}\}$ is small enough such that, with $L = 2L_0$, both $\beta := O_h(\delta, x_0)\, L_\gamma < \frac{\delta_0}{3(L_0+1)}$ and $\theta := \beta\,(L+1) < \frac{1}{2}$ hold true, and

(iii) $\|f(x_0) + m_0\| < \frac{1}{2}\delta\,(L+1)^{-1}$ is satisfied.

Then, the procedure (5.15) with selection rule (5.16) generates a sequence which converges to a zero x^* of Γ with $\|x^* - x_0\| \le 2(L+1)\|f(x_0) + m_0\| < \delta$ and

$$\|x_{k+1} - x_k\| \le \theta\, \|x_k - x_{k-1}\|, \quad \|x_k - x_0\| \le 2(L+1)\, \|f(x_0) + m_0\| \quad \forall k > 0. \tag{5.17}$$

Proof Define

$$\begin{aligned} F(x) &= f(x_0) + H_0\,(\gamma(x) - \gamma(x_0)) + M(x), \\ g(x) &= f(x_0) + H_0\,(\gamma(x) - \gamma(x_0)) - f(x). \end{aligned} \tag{5.18}$$

The initial mapping $\Gamma = f + M$ fulfills $g + \Gamma = F$, $y_0 = f(x_0) + m_0 \in F(x_0)$ and the equivalences $g(x) \in F(x) \Leftrightarrow -f(x) \in M(x) \Leftrightarrow 0 \in \Gamma(x)$. In addition, the definitions ensure

$$\begin{aligned} & x_{k+1} \in F^{-1}(g(x_k)) \ \Leftrightarrow \ g(x_k) \in F(x_{k+1}) \\ & \Leftrightarrow f(x_0) + H_0\,(\gamma(x_k) - \gamma(x_0)) - f(x_k) \in f(x_0) + H_0\,(\gamma(x_{k+1}) - \gamma(x_0)) + M(x_{k+1}) \\ & \Leftrightarrow 0 \in f(x_k) + H_0\,[\gamma(x_{k+1}) - \gamma(x_k)] + M(x_{k+1}). \end{aligned} \tag{5.19}$$

Thus the iterations of $\mathrm{P}(\lambda, \beta, x_0, y_0, v_0)$ (for $\lambda = 1$) correspond exactly to the method (5.15), (5.16). We continue by checking the assumptions of Thm. 10. First we investigate the Lipschitz rank of g with respect to $x', x \in x_0 + rB$ and $0 < r < \delta_0$:

$$\begin{aligned} & \|g(x') - g(x)\| = \|H_0\,\gamma(x') - f(x') - (H_0\,\gamma(x) - f(x))\| \\ & = \|h(\gamma(x)) - h(\gamma(x')) - H_0\,(\gamma(x) - \gamma(x'))\| \\ & \le \sup_{x,x' \in x_0 + rB} \|Dh(\gamma(x)) - Dh(\gamma(x'))\|\ \|\gamma(x) - \gamma(x')\|. \end{aligned} \tag{5.20}$$

Hence

$$Lip(g, x_0 + rB) \le O_h(r, x_0)\, L_\gamma = O(r). \tag{5.21}$$

This yields due to $g(x_0) = 0$,

$$\sup(g, x_0 + rB) \ \le\ r\, Lip(g, x_0 + rB) = o(r). \tag{5.22}$$

Because of (i) and since $F = g + \Gamma$ where g is a small Lipschitz function, also F is pseudo-regular with some rank L at (x_0, y_0) and neighborhoods $U = x_0 + \delta B, V = y_0 + \delta B$, provided that

$$\mu(\delta) := \max\,\{\sup(g, x_0 + \delta B),\ Lip(g, x_0 + \delta B)\,\}$$

is small enough. Using [40, proof of Thm. 2.4], this is true if L, δ and $\mu(\delta)$ satisfy

$$L = 2L_0, \quad \delta < \frac{\delta_0}{12(L_0+1)} \quad \text{and} \quad \mu(\delta) < \frac{\delta_0}{3(L_0+1)}.$$

By (5.21) and (5.22), the second inequality holds for $\delta < 1$ and $O_h(\delta, x_0)\, L_\gamma < \frac{\delta_0}{3(L_0+1)}$. After fixing δ in this way, which is possible due to $\mu(\delta) \to 0$ and $O_h(\delta, x_0) \to 0$ as $\delta \downarrow 0$, we have to investigate α and β in Remark 7. The inequality $\beta(L+1) < \frac{1}{2}$ is guaranteed by our choice of δ. The same holds for the second condition since $y_0 = f(x_0) + m_0$ and $\frac{1}{2}\delta\,(L+1)^{-1} > \alpha = \|v_0 - y_0\| = \|g(x_0) - y_0\| = \|f(x_0) + m_0\|$. Thus the sequence converges, with the estimates (5.17), to some x^* which solves $0 \in f + M$ by taking (5.15) into account.

Let us add some comments on Prop. 8. Our estimates are not sharp, and the convergence statement is less detailed than Thm. 9. If Γ is strongly regular at (x_0, y_0) then so is F and all x_{k+1} are unique by Remark 5. Next we emphazise another fact which is different for equations and generalized equations. Under assumption (i), some small $\delta \in (0, c)$ with the properties (ii) really exists, while (iii) is an additional requirement at the initial point, satisfied for equations if $\|x_0 - \bar{x}\|$ is small and $\bar{x}$ is a zero. However, even if f is linear, $0 \in \Gamma(\bar{x})$ and $\Gamma = f + M$ is strongly regular at $(\bar{x}, 0)$, the hypotheses of Prop. 8 are not automatically fulfilled for sufficiently small $\|x_0 - \bar{x}\|$.

Example 5 (condition (iii) for x_0 near $\bar{x}$) Consider $f : \mathbb{R}^2 \to \mathbb{R}$ and closed $M : \mathbb{R}^2 \rightrightarrows \mathbb{R}$ as

$$f(x_1, x_2) = x_1, \quad M(x_1, x_2) = \begin{cases} \{1/x_2\} & \text{if } x_2 \neq 0 \\ \{0\} & \text{if } x_2 = 0 \end{cases}. \qquad \text{Put } \Gamma = f + M.$$

For $|y| < 1$ and $\|x\| < 1$, the solution x of $y \in \Gamma(x)$ is unique and Lipschitz: $x = (y, 0)$. Hence Γ is strongly regular at the origin. If $x \to (0,0)$ and $x_2 \neq 0$ then $|y| \to \infty\ \forall y \in \Gamma(x)$ follows. Thus, for such initial points x_0 near $x^* = \bar{x} = (0,0)$, one cannot satisfy the assumptions of Prop. 8 since $|m_0|$ becomes big. The assigned y-values $1/x_{02}$ and 0 are too far in order to apply regularity at $(\bar{x}, 0)$. Nevertheless, the first Newton step at any x_0 generates the solution $\bar{x}$.

Therefore, in statements similar to Prop. 8, e.g., [13, Thm. 4], often additional hypotheses occur about the situation after the first iteration. The undesired effect disappears for equations or, more general, if $M(x)$ is fixed or Γ fulfills, as under condition (3.17),

$$\limsup_{x \to \bar{x}}\ \operatorname{dist}(0, \Gamma(x)) = 0 \qquad \text{(i.e., } \Gamma \text{ is l.s.c. at } (\bar{x}, 0)). \tag{5.23}$$

Then, it is not difficult to see that, due to persistence of pseudo-regularity of Γ near $(\bar{x}, 0)$, we find, for each x_0 close enough to $\bar{x}$, some appropriate m_0 satisfying all hypotheses.

Extension to the proper Newton method

We apply Remark 6 to the model (5.15) with $H_k = Dh(\gamma(x_k))$ in place of H_0. Increasing L if necessary and decreasing the δ assigned to pseudo regularity, we may suppose that $O_h(\delta, x_0)$ is already small enough such that the mappings

$$F_k(x) = f(x_0) + H_k\,(\gamma(x) - \gamma(x_0)) + M(x), \quad g_k(x) = f(x_0) + H_k\,(\gamma(x) - \gamma(x_0)) - f(x)$$

satisfy our assumptions on L and β, too. With $x_{k+1} \in F_k^{-1}(g_k(x_k))$, they realize now the steps of the proper Newton method with the related adapted g_k-selection rule (5.16). For the limit x^*, it follows from $g_k(x_k) \in F_k(x_{k+1})$,

$$f(x_0) + H_k(\gamma(x_k) - \gamma(x_0)) - f(x_k) \in f(x_0) + H_k(\gamma(x_{k+1}) - \gamma(x_0)) + M(x_{k+1})$$

and, as desired, $0 \in f(x^*) + M(x^*)$. Hence, the claimed convergence and estimates remain true. For x_k near x^*, the hypotheses are satisfied with certain vanishing $\theta_k \downarrow 0$. By Remark 4, also the factor of linear convergence vanishes. So the convergence is superlinear.

Remark 8 Under (5.23), the current method can be applied for all x_0 near $\bar{x}$: Adding our adapted g_k-selection rule (5.16) to the iterates of Corollary 1 (case 1), we obtain automatically a sequence whose existence was asserted.

General point-based approximations

For Robinson's [52] PBA, f and Σ are continuous functions on open sets Ω and $\Omega \times \Omega$, respectively; cf. §2.3. To ensure the existence and convergence - in a Kantorovich-type manner - of the iterates $0 \in \Sigma(x_{k+1}, x_k)$, the main suppositions in [52, Thm. 3.2] require for the initial point x_0:

$$\begin{array}{ll} (i) & \text{for small } \|y\| \text{ a solution } x \text{ to } \Sigma(x,x_0)=y \text{ exists in } \Omega \\ (ii) & c := \inf \{ \ \|\Sigma(x',x_0)-\Sigma(x,x_0)\| \, \|x'-x\|^{-1} \mid x' \neq x \in \Omega \ \} > 0. \end{array} \tag{5.24}$$

For the skillful interplay of the constants K and c in Robinson's convergence-theorem (under the additional hypothesis that $\|x_1 - x_0\|$ is sufficiently small for the first iteration x_1 and Ω is big enough), we refer to [52] which uses [46, Thm. 1.9]. Condition (5.24) (ii) ensures Lipschitz behavior of the solutions $x = x(y)$ on some ball. So the function $x' \mapsto F(x') := \Sigma(x', x_0)$ is strongly regular at $(x_0, f(x_0))$.

With $g_x(x') = \Sigma(x',x) - \Sigma(x',x_0)$, the equation $0 = g_x(x') + F(x')$ describes the solutions of $\Sigma(x',x) = f(x) + Gf(x)(x'-x) = 0$ and the assigned Newton method (2.27).

With $\hat{g}_x(x') = f(x) + Gf(x_0)(x'-x) - \Sigma(x',x_0)$, equation $0 = \hat{g}_x(x') + F(x')$ describes the solutions of $\hat{\Sigma}(x',x) = f(x) + Gf(x_0)(x'-x) = 0$ and the modified Newton method (5.15) with $M \equiv \{0\}$.

Both models can be handled, by Prop. 8, for the composed functions (2.12). Unfortunately, for the general setting (2.16) of PBAs, we cannot apply Prop. 8, since (5.14) and hence the existence of Dh were assumed there. Nevertheless, there is some (Kantorovich-type) statement if *arbitrary PBAs* of f replace the linearizations in $f + M$ under pseudo-regularity at the initial point (x_0, y_0), cf. [13, Thm. 4]. However, this asserts once more the existence of a related Newton sequence, and examples of such PBAs are not added. The same is true for Thm. 6.3 in [1] where systems $0 \in f(x) + F(x)$ have been considered with multivalued f and F, and the convergence is based on approximations $\hat{f}$ of f in the Hausdorff-metric. Other approximations will be studied in section 6.

Remark 9 Looking on the class (2.12) of composed functions $f = h \circ \gamma$, one could believe that, after setting $h = \text{id}$, (5.15) or similar schemes produce Newton-type methods for arbitrary Lipschitz functions $f = \gamma$. But we have to stop the readers enthusiasm: Though *we obtain all* $C^{0,1}$ *functions* f, the auxiliary equation $\Sigma(x',x) = 0$ for Newton's method in (2.12) remains just the original one: $f(x') = 0$. Also (5.15) then leads us again to the original problem $0 \in f(x_{k+1}) + M(x_{k+1})$.

6 Approximations for General Multifunctions

In this section, we investigate inclusions under less traditional hypotheses which a priori do not utilize the structure of generalized equations or derivatives of composed functions as under (2.12). They use the framework of the general approximation scheme introduced in §2.4. Accordingly, we study $0 \in \Gamma(x)$ where $\Gamma : X \rightrightarrows Y$ is closed, and consider a mapping $\Sigma : X \times X \rightrightarrows Y$ as well as the iterations $0 \in \Sigma(x_{k+1}, x_k)$ from (2.22) with the (solution) sets $\mathcal{S}(x) = \Sigma(.\,,x)^{-1}(0)$. The difference $\Sigma(.\,,x) - \Sigma(.\,,x')$ is, in general, neither defined nor a function.

We follow [36,39], where solvability of $0 \in \Sigma(.,x_k)$ and the existence of a zero $\bar{x}$ for Γ are assumed and discussed for particular cases. The conditions for convergence are based on

(G1) The inverse of $\Sigma(.\,,\bar{x})$, namely $\Phi = \Sigma(.,\bar{x})^{-1}$, has to be *locally u.L.* at $(0,\bar{x})$.

(G2) A relation between the graphs of $\Sigma(.\,,x)$ and $\Sigma(.\,,\bar{x})$ by the requirement:

$$\text{if } \ x' \in \mathcal{S}(x) \cap [\bar{x} + \varepsilon B] \quad \text{then} \quad \text{dist}(\ (x',0),\ \text{gph}\, \Sigma(.\,,\bar{x})\ ) \leq \tau \tag{6.1}$$

where dist uses the max-norm of $X \times Y$ and $\tau = \tau(\varepsilon,\, x - \bar{x})$ is small; see (G3).
Explicitly, (G2) claims: *Given a solution* $x' \in \bar{x} + \varepsilon B$*, assigned to* x*, there is some* (x'',y') *such that both* $y' \in \Sigma(x'',\bar{x})$ *and* $\max\{\, \|x''-x'\|,\, \|y'-0\|\, \} \leq \tau$.

(G3) An estimate for the function $\tau = \tau(\varepsilon, z)$ near $(0, 0_X)$, namely:

$$\begin{array}{c} \exists \text{ real functions } a_i \geq 0 \ (i=0,1) \text{ such that } \lim_{s\downarrow 0} a_i(s)s^{-i} = 0 \\ \text{and } \ \tau(\varepsilon,z) \ = \ a_0(\varepsilon)\|z\| + a_0(\|z\|)\varepsilon \ + \ a_1(\varepsilon) + a_1(\|z\|). \end{array} \tag{6.2}$$

Notice that ε estimates $\|x' - \bar{x}\|$ and z stands for $x - \bar{x}$ in (G2). The interplay of these conditions for superlinear convergence describes

Theorem 11 [39, p. 244] Suppose (G1), (G2) and (G3). Then, for each $q \in (0,1)$, there is some $\rho > 0$ such that all solutions $x_{k+1} \in \mathcal{S}(x_k) \cap (\bar{x} + \rho B_X)$ (as long as they exist) fulfill $\|x_{k+1} - \bar{x}\| \leq q\ \|x_k - \bar{x}\|$ whenever $\|x_0 - \bar{x}\| \leq \rho$.

The hypotheses (G1), (G2), (G3) look very artificial. So, in the rest of this section, let us discuss them under two viewpoints. How they can be satisfied and what about its necessity for convergence of Newton's method? We consider first a simpler approximation for the images of $\Sigma(.,x)$ only.

The graph-estimate (6.1) in (G2) is obviously satisfied under a stronger condition for the Σ-images alone (which requires additionally $x'' = x'$ in the explicit form), namely if

$$(G2)' \quad \Sigma(x', \bar{x}+z) \subset \Sigma(x', \bar{x}) + \tau(\varepsilon, z)B_Y \quad \forall x' \in \bar{x} + \varepsilon B_X. \tag{6.3}$$

This condition requires some upper semicontinuity of $\Sigma(.\,,x)$ at $\bar{x}$, measured by τ.

For comparing Thm. 11 with Corollary 1 (which assumes pseudo-regularity), let us consider the following two situations.

Case 1: f is of type (2.12) and $\Sigma(x', x) = f(x) + Dh(\gamma(x))(\gamma(x') - \gamma(x)) + M(x')$,
Case 2: f is continuous, there is a PBA $\hat{\Sigma}$ for f near $\bar{x}$ and $\Sigma(x', x) = \hat{\Sigma}(x', x) + M(x')$.

Proposition 9 *Let $\Gamma = f + M$ be a multifunction. In case 1, the conditions (G2)' and (G3) are satisfied with a function τ of the form $\tau = a_1(\|z\|) + a_0(\|z\|)\varepsilon$. In case 2, (G2)' and (G3) are satisfied with τ of the form $\tau = a_0(\varepsilon)\|z\| + a_1(\|z\|) + a_1(\varepsilon)$.*

Hence, under (G1), Thm. 11 can be applied even without supposing pseudo-regularity.

Proof Case 1: Assume $\|x' - \bar{x}\| = \varepsilon$ and $a \in \Sigma(x', x)$. We have to find some $b \in \Sigma(x', \bar{x})$ such that $a - b$ has sufficiently small norm τ. Since $a = f(x) + Dh(\gamma(x))(\gamma(x') - \gamma(x)) + m'$ holds with some $m' \in M(x')$, we obtain $b := f(\bar{x}) + Dh(\gamma(\bar{x}))(\gamma(x') - \gamma(\bar{x})) + m' \in \Sigma(x', \bar{x})$. So let us estimate $\|a - b\|$. This is just the norm $\|g_x(x')\|$ in (2.21) which satisfies $\|g_x(x')\| \le o(x - \bar{x}) + O(x - \bar{x})\,\|x' - \bar{x}\|$. Hence condition (6.3) holds with τ of the form $\tau = a_1(\|z\|) + a_0(\|z\|)\varepsilon$.

Case 2: We may use formula (2.18) and continue with

$$\|x'-x\|^2 + \|x'-\bar{x}\|^2 \le (\|x'-\bar{x}\| + \|\bar{x}-x\|)^2 + \|x'-\bar{x}\|^2 = 2\|x'-\bar{x}\|\|\bar{x}-x\| + \|\bar{x}-x\|^2 + 2\|x'-\bar{x}\|^2.$$

Hence τ has the form $\tau = a_0(\varepsilon)\|z\| + a_1(\|z\|) + a_1(\varepsilon)$.

Obviously, M played no role; the estimate with or without m' is the same as for $M \equiv \{0\}$.

Corollary 2 *After replacing $\Sigma(.\,,\bar{x})$ in Prop. 9, case 1, by the original multifunction $\Gamma = f + M$, i.e., $\Sigma(x', \bar{x}) = f(x') + M(x')$ (without using $Dh(\gamma(\bar{x}))$ explicitly), one obtains a similar result with $\tau = a_1(\|z\|) + a_0(\|z\|)\varepsilon + a_1(\varepsilon)$.*

Proof Indeed, we have only to estimate (deleting m' as above and using the notations from the proof of (2.20)), $d = \|\, f(x) + H_x(\gamma' - \gamma) - f(x')\,\| = \|\, f(x) + H_x(\gamma' - \gamma) - f(\bar{x}) - \bar{H}(\gamma' - \bar{\gamma}) + o(x' - \bar{x})\|$ which yields with the above estimate for $\|a - b\|$: $d \le o(x - \bar{x}) + O(x - \bar{x})\,\|x' - \bar{x}\| + o(x' - \bar{x})$.

It is worth noting that the approximations (G2), (G3) above are not only of interest for the settings under Prop. 9 or in view of Newton's method at all. In [39], two examples coming from convex analysis are discussed to show how the hypotheses can be satisfied. The first one concerns proximal points with large exponents where, for minimizing a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the solvable problems $\min_\xi \; f(\xi) + \|\xi - x\|^p, \quad (p > 2) \quad (x \in \mathbb{R}^n)$ and

$$\Sigma(x', x) \;=\; \partial_{x'}\,[f(x') + \|x' - x\|^p\,] \;=\; p\,\|x' - x\|^{p-2}(x' - x) \;+\; \partial f(x') \tag{6.4}$$

have been studied. The second one concerns the use of so-called $\varepsilon-$ subgradients for minimizing a convex $\mathbb{R}^n$- function. For both examples, the iterates x_{k+1} exist under well-known facts of convex analysis, and (G1) requires upper Lipschitz behavior of $(\partial f)^{-1}$. The regularity properties (strong, pseudo, upper) of ∂f for convex f on $\mathbb{R}^n$ have been completely characterized by [31, Thm. 5.4]. In particular, upper regularity simply means quadratic growth of f at $\bar{x}$, i.e., $\exists \varepsilon > 0 :$ $f(\bar{x} + u) \ge f(\bar{x}) + \varepsilon\|u\|^2 \,\forall u \in \varepsilon B$. Strong and pseudo-regularity of $\partial f$are useful for the process $\mathrm{P}(\lambda, \beta, x_0, y_0, v_0)$ in §5.1.

The stronger condition (6.3) = (G2)' can be also combined with a *simpler function* τ, namely

$$(G3)' \qquad \tau(\varepsilon, z) \;\le\; c\,\|z\| \;+\; a_0(\|z\|)\varepsilon + a_1(\|z\|) \qquad (c \ge 0). \tag{6.5}$$

Then, with rank L under (G1) and possibly larger q than in Thm. 11, one obtains,

Theorem 12 [36, Thm. 1] *Let $\Gamma = f$ be a function and suppose (G1), (G2)' and (G3)' with $cL < 1$. Then, for each $q \in (cL, 1)$, there is some $\rho > 0$ such that all $x_{k+1} \in \mathcal{S}(x_k) \cap (\bar{x} + \rho B_X)$ (as long as they exist) fulfill $\|x_{k+1} - \bar{x}\| \le q\,\|x_k - \bar{x}\|$, provided that $\|x_0 - \bar{x}\| \le \rho$.*

Clearly, $c = 0$ and the existence of all x_{k+1} are sufficient for local superlinear convergence.

The imposed conditions for locally Lipschitz functions

Let $\Gamma = f$ and $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ be strongly regular at a zero $\bar{x}$. According to (2.27), we put

$$\Sigma(x', x) = \begin{cases} \{f(x')\} & \text{if } x = \bar{x} \\ f(x) + Gf(x)(x' - x) & \text{if } x \neq \bar{x} \end{cases} \tag{6.6}$$

and suppose that either $Gf(x)(u) = f'(x; u)$ is the usual directional derivative (provided it exists), or $Gf(x)(u) = Tf(x)(u)$ consists of the Thibault derivative, cf. §2.2. Under these assumptions, it has been shown in [39, §2.3]:

Necessity. If, in Thm. 11 the iterates exist for all k and converge superlinearly, then the conditions (G1), (G2), (G3) are satisfied. The proof used results of [38], consequences of Thm. 2 as well as the conditions (CI), (CA) and (CA)* of section 4.

Sufficiency. Conversely, by Thm. 11, superlinear convergence holds true under the conditions (G1), (G2), (G3), if our auxiliary problems are solvable in $\bar{x} + \rho B$. For $Gf = Tf$, this follows from the inverse derivative rule (2.10) and non-emptyness of $Tf^{-1}(f(x))(v)$. For $Gf = f'$, one may first use that $0 \in f(x) + Cf(x)(u)$ has solutions u due to Thm. 2, see (2.7). Since $f \in C^{0,1}$ and f' exists, $Cf(x)(u)$ consists of $f'(x; u)$ only.

Hence, for strongly regular $f \in C^{0,1}(\mathbb{R}^n, \mathbb{R}^n)$ and both settings of Gf, our conditions are necessary and sufficient for superlinear convergence *and* the existence of x_{k+1} in Thm. 11.

7 Some non-Derivative Approaches for Nonsmooth Functions

After studying Newton maps and other derivative-like objects for nonsmooth functions, let us turn to Newton's method, where f is not necessarily replaced by its (generalized) derivative or is only continuous. We begin with an observation from [31, sect. 10.2] for PC^1 functions and extend it from equations to some generalized equations. Then we follow ideas of [20] and consider piecewise linear approximations and automatic differentiation for handling certain composed nonsmooth functions. Finally, we recall a discussion in [41] and ask for the role of the key condition (CA)* of section 4 in the case of non-Lipschitz functions.

PC^1 equations, the simplest approach

The well-known theory of generalized Newton methods for PC^1 functions $f : \mathbb{R}^n \to \mathbb{R}^n$, generated by $f^1, ..., f^N$, mostly uses the hypothesis of non-singularity for all $Df^s(\bar{x})$ with $s \in I(\bar{x})$. This is a direct and canonical generalization of the usual C^1 case and is even a necessary condition in order to obtain a "regular B-derivative" $\partial_B f(\bar{x})$ at $\bar{x}$ if all $s \in I(\bar{x})$ are essential. Almost all papers, however, do nowhere take into consideration that this hypothesis is strong enough *for avoiding non-smooth Newton methods at all.* This can be seen as follows.

The hypothesis ensures that $\bar{x}$ is a (strongly) regular and isolated zero for each function f^s, $s \in I(\bar{x})$. Hence Newtons method converges as usual to $\bar{x}$ for the C^1 function f^s and initial points $x_0 \in \bar{x} + \varepsilon_s B$ with some $\varepsilon_s > 0$. For small $\delta > 0$ and $\|x_0 - \bar{x}\| \leq \delta$, also $\emptyset \neq I(x_0) \subset I(\bar{x})$ is obviously true by continuity of f. Therefore, it holds,

Proposition 10 *Let $\{x_k\}$ be a sequence generated by choosing any $s_0 \in I(x_0)$ and applying the usual Newton method to the C^1 function f^{s_0}, where s_0 remains fixed even if $f^{s_0}(x_k) \neq f(x_k)$ holds at some iteration point x_k. If all derivatives $Df^s(\bar{x}), s \in I(\bar{x})$, are regular and $\|x_0 - \bar{x}\| \leq \rho :=$ $\min\{\delta, \min_s \varepsilon_s\}$, then $\{x_k\}$ converges superlinearly to $\bar{x}$.*

This is not only simpler than the classical "active index set strategy" (cf., e.g., [17, §7.2.1], [35], [36]) which uses $\mathcal{N}_1(x)$ of (3.7) and, in consequence, only active functions $f^{s(k)}$ at the iteration point x_k. Mainly, it also permits to apply *all* modifications of Newton's method to f^{s_0} and to extend these modifications to active index set strategies in an evident manner. The latter is possible since the proposition allows to replace the function f^{s_0}, at any step k, by another function f^s which is active at x_k since $\|x_k - \bar{x}\| < \|x_0 - \bar{x}\| \leq \rho$ remains true.

These facts do not imply that some $Rf(x) = Df^s(x)$ is a Newton function for f in the sense of §3.1 or is any other generalized derivative of f at $\bar{x}$. This shows already $f(x) = |x|$ which is PC^1 with $f^1 = x, f^2 = -x$ and $f^3 = 7x$. Clearly, f^3 is not essential for describing f and cannot appear as f^{s_0} unless we start at the zero. Nevertheless, even f^3 could be used to compute the zero.

Obviously, Prop. 10 does not help without knowing the generating C^1 functions f^s of f explicitly. But this applies also to all active index methods, and they need the condition $\|x_0 - \bar{x}\| \le \rho$, too. The drawback of this condition becomes obvious with a perturbation f_α of the abs-function

$$f_\alpha(x) = \begin{cases} f^1 := x & \text{if } x \ge -\alpha \\ f^2 := -x - 2\alpha & \text{if } x \le -\alpha \end{cases} \tag{7.1}$$

for small $\alpha > 0$. At $\bar{x} = 0$, only f^1 is active. Thus $I(x_0) \subset I(\bar{x})\ \forall x_0 \in \bar{x} + \delta B$ requires $\rho \le \delta < \alpha$, and the sufficient convergence condition $\|x_0 - \bar{x}\| \le \rho$ is very strong. Notice, however, that this condition is not a necessary one for superlinear convergence.

It is well-known that many variational problems can be written as equations of PC^1 functions. Let us mention only a few of them.

The KKT system of optimization problems (2.4) for $(a, b) = (0, 0)$ can be handled in this way since the KKT-points are (up to a simple transformation) the zeros of the (Kojima-) function F where $y_i^+ = \max\{0, y_i\},\ y_i^- = \min\{0, y_i\}$ and

$$F(x, y) = (\, Df(x) + \textstyle\sum_i\ y_i^+ \, Dg_i(x), \quad g_1(x) - y_1^-, \ \dots, \ g_m(x) - y_m^- \,) \ \in \mathbb{R}^{n+m}. \tag{7.2}$$

Here, $F : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ is of type PC^1 since $y_i \mapsto (y_i^+, y_i^-)$ is a PC^1- function of the two functions $y_i \mapsto (y_i, 0)$ and $y_i \mapsto (0, y_i)$. Hence, following Prop. 10 one obtains

Proposition 11 *Let F be strongly regular at a zero $(\bar{x}, \bar{y})$ and $\rho > 0$ be sufficiently small. For $\|(x_0, y_0) - (\bar{x}, \bar{y})\| \le \rho$, replace above all (y_i^+, y_i^-) by $(y_i, 0)$ if $y_{0i} > 0$ and by $(0, y_i)$ otherwise. Then, Newton's method applied to the related C^1 function, say F^s, converges as usually to $(\bar{x}, \bar{y})$.*

Proof The hypothesis of strong regularity implies regularity of $DF^s(\bar{x}, \bar{y})$ for (x_0, y_0) near $(\bar{x}, \bar{y})$ and the index condition holds true for small ρ.

Passing to $(y_i, 0)$ if $y_{0i} > 0$ means to require $g_i(x) = 0$ in the optimization problem, passing to $(0, y_i)$ means to delete the ith constraint. The same holds with additional equality constraints of type C^2 and for *generalized Kojima-functions* where (in particular) any C^1 function $\Phi = \Phi(x)$ of the same dimension may replace $Df(x)$. In this way, variational conditions, games or complementarity problems can be written as equations. For proofs and details, cf. [31, chapter 7]. Other "derivatives" of F can be found in [33].

Using differences and the need of automatic differentiation

Having in mind the trivial setting

$$Gf(x)(u) = f(x + u) - f(x) \text{ after which } \ 0 \in f(x) + Gf(x)(u) \Leftrightarrow f(x + u) = 0,$$

it is evident that $f(x + u) - f(x)$ should be approximated in some appropriate way. We consider two straightforward ideas:

(A1) If f is composed by a finite number of PC^1-functions $h^i = h^i(x_i)$, one could construct $Gf(x)(u)$ by the help of directional derivatives $(h^i)'(x_i; u_i)$ (and chain rules for sums, products and quotients). We then trivially obtain the directional derivative of f.

(A2) If certain directional derivatives, say $(h^j)'(x_j; u_j)$, are not available or difficult to compute, one can replace them by the differences $h^j(x_j + u_j) - h^j(x_j)$. In particular, this is possible if f is composed by C^1-functions h^i and piecewise linear functions h^j. As elaborated in [20], then the resulting models reflect the original structure in an often preferable manner (the drawback of Example 7.1 disappears) and the "derivatives" can be determined by automatic differentiation, as long as f is defined in a hierarchic manner by the functions h^k like a tree in a graph.

To be more concrete, assume simpler that $f = h^N(\dots(h^2(h^1(x)))\dots) : \mathbb{R}^n \to \mathbb{R}^m$ is composed by PC^1 functions h^i which define $x_{i+1} = h^i(x_i)$. Then f is again PC^1, $x_1 = x \in \mathbb{R}^n$ and $h^N(x_N) \in \mathbb{R}^m$. Let us look at the above approximations $\hat{f}$ for $f(x+u) - f(x)$.

(A1) uses the directional derivative $\hat{f} = f'(x;u)$, which is computable via the directional derivatives of all h^i as $f'(x;u) = (h^N)'(x_N;u_N)$ where, recursively, $x_1 = x,\ u_1 = u$ and

$$x_{i+1} = h^i(x_i), \quad u_{i+1} = (h^i)'(x_i;u_i) \quad i = 1\dots N-1. \tag{7.3}$$

Clearly, setting $\hat{f} = f(x+u) - f(x)$ makes no sense since it does not simplify the equation.

(A2) however, requires to take such differences for the involved difficult functions h^j. Then $(h^j)'(x_j;u_j)$ is replaced by $u_{j+1} = h^j(x_j+u_j) - h^j(x_j)$ in (7.3) and $x_{j+1} = h^j(x_j)$ is given as above. Now the approximation $\hat{f}$ is no longer positively homogeneous in the direction u.

Example 6 Let $f(x) = abs(\,5\sin(abs(x)\,+1)) = h^3(h^2(h^1(x))),\ x_1 = x,\ u_1 = u$. Then one may write $x_2 = h^1(x_1) = abs(x_1),\ x_3 = h^2(x_2) = 5\sin(x_2)+1,\ f = h^3(x_3) = abs(x_3)$. Here, (A1) yields

$$\begin{array}{ll} x_2 = abs(x), & u_2 = (abs)'(x;u), \\ x_3 = 5\sin(abs(x)+1), & u_3 = 5\cos(abs(x))\,u_2, \quad \text{which implies} \end{array}$$
$$\hat{f} = f'(x;u) = (h^3)'(x_3;u_3) = (abs)'(\ 5\sin(abs(x)+1)\ ;\ 5\cos(abs(x))(abs)'(x;u)\ )$$

while (A2) with $h^j = abs$ for $j \neq 2$ corresponds to

$$\begin{array}{ll} x_2 = abs(x), & u_2 = abs(x+u) - abs(x), \\ x_3 = 5\sin(abs(x)+1), & u_3 = 5\cos(abs(x))\,u_2 \quad \text{and} \\ (h^3)'(x_3;u_3) \text{ is replaced by} & abs(x_3+u_3) - abs(x_3), \quad \text{which yields} \end{array}$$
$$\hat{f} = abs(\,5\sin(abs(x)+1)\,+5\cos(abs(x))\,[abs(x+u) - abs(x)]\,) - abs(\,5\sin(abs(x)+1)).$$

The example could be easily extended to larger size and shows that - for both approaches - automatic differentiation is an important pre-requisite for dealing with nonsmooth Newton methods; even if only standard functions like *abs* or *max* are involved. Fortunately, in many concrete variational problems, the non-smoothness arises from complementarity conditions only and allows simpler procedures for handling this situation, since *abs* or *max* are involved in an elementary manner. But for hierarchic problems (like so-called MPECs), functions as above may occur and it becomes important that $\hat{f}$ is still of type PC^1 and can be efficiently determined.

Difficulties and condition (CA) for non-Lipschitz functions*

The restriction to $f \in C^{0,1}$ in the context of standard settings (4.15) is motivated by formula (3.6) which requires pointwise Lipschiz behavior. For nonlinear approximations and $f \in \tilde{C} := C \setminus C^{0,1}$, the situation is not simpler. In the paper [22], devoted to Newton's method based on contingent derivatives for continuous f, there is no $f \in \tilde{C}$ such that the proposed method converges or all hypotheses of the convergence statements are satisfied. We also found nowhere an example of a PBA for continuous f, different from the functions $f = h \circ \gamma$ in (2.12) with $h \in C^{1,1}$. So it is even not clear whether any Newton-type method may, in fact, superlinearly converge for $f \in \tilde{C}$.

To give an answer and to characterize the difficulties, let us check superlinear convergence and condition (CA)* for real functions $f \in \tilde{C}$ which are C^1 near $x \neq \bar{x}$. So one can use the usual Newton steps at $x \neq \bar{x}$. For the known example $f(x) = \text{sgn}(x)\,|x|^q$ and $0 < q < 1$, both superlinear convergence and (CA)* are violated. The following strongly regular functions indicate that Newton's method may superlinearly converge for $f \in \tilde{C}$, while (CA)* may hold or not. In both examples, put $f(0) = 0$ and $f(x) = -f(-x)$ for $x < 0 = \bar{x}$.

Example 7 Superlinear local convergence, though (CA)* is violated: $f(x) = x\,(1 - \ln x)$ if $x > 0$. Evidently, f is continuous and, for $x > 0$, it holds $Df = -\ln x$ and $x_{\text{new}} = x - \frac{x(1-\ln x)}{-\ln x} = x + \frac{x}{\ln x} - x = \frac{x}{\ln x}$. This implies

$q_1(x) := f/Df \to 0$ for $x \downarrow 0$ due to $q_1 = 1 - \frac{f(x)}{xDf(x)} = 1 - \frac{x\,(1-\ln x)}{-x\ln x} = 1 - (-\frac{1}{\ln x} + 1) = \frac{1}{\ln x}$,

and (CA)* fails due to $q_2(x) := \frac{f(x)}{x} - Df(x) = \frac{x\,(1-\ln x)}{x} + \ln x \equiv 1.$

Example 8 Superlinear local convergence *and* (CA)* hold true: $f(x) = x\,(\,1{+}\ln(-\ln x)\,)$ if $x > 0$. Consider small $x > 0$ which yields $f > 0$ and, for $x \downarrow 0$,

$$\begin{aligned}
Df &= (\,1 + \ln(-\ln x)\,) + \;x\,(\tfrac{1}{-\ln x}\;\tfrac{1}{-x}) \;=\; 1 + \ln(-\ln x)\; + \tfrac{1}{\ln x} \to \infty.\\
q_1 &= 1 - \tfrac{f}{xDf} = 1 - \tfrac{1+\ln(-\ln x)}{1+\ln(-\ln x)\;+\frac{1}{\ln x}} = \tfrac{\frac{1}{\ln x}}{1+\ln(-\ln x)\;+\frac{1}{\ln x}} \to -0,\\
q_2 &= \tfrac{f}{x} - Df = (\,1 + \ln(-\ln x)\,) - (\,1 + \ln(-\ln x)\; + \tfrac{1}{\ln x}) = -\tfrac{1}{\ln x} \to 0.
\end{aligned}$$

Similarly, negative x can be handled. Thus the assertions are verified.

Both examples violate the crucial pointwise Lipschitz condition (4.11), used in [22], too. This can be explained by [41, Thm. 4.1]: *Let $f \in \tilde{C}$ be a real, strongly regular function which is not locally Lipschitz near a zero $\bar{x}$. Assume that the method $0 \in f(x_k) + Cf(x_k)(x_{k+1} - x_k)$ generates infinite sequences with $|x_{k+1} - \bar{x}| = o(x_k - \bar{x})$ whenever $|x_0 - \bar{x}|$ is small enough. Then (4.11) cannot hold.* The proof applies Prop. 1.

Acknowledgement. The authors are indebted to the referees for their constructive comments that significantly improved the presentation.

References

1. Adly, S., Cibulka, R., van Ngai, H.: Newton's method for solving inclusion using set-valued approximations. SIAM J. Optim. 25, 159–184 (2015)
2. Aragon Artacho, F.J., Belyakov, M., Donchev, A.L., Lopez, M.: Local convergence of quasi-Newton methods under metric regularity. Comput. Optim. Appl. 58, 225–247 (2014)
3. Argyros, I.K.: Convergence and Applications of Newton-type Iterations. Springer, New York (2010)
4. Aubin, J.-P., Ekeland, I.: Applied Nonlinear Analysis. Wiley, New York (1984)
5. Bartle, R.G.:. Newton's method in Banach spaces. Proc. Amer. Math. Soc. 5, 827–831 (1955)
6. Broyden, C.G.: A class of methods for solving nonlinear simultaneous equations. Math. Comp. 19, 577–593 (1965)
7. Bütikofer, S., Klatte, D.: 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)
8. Cibulka, R., Dontchev, A.L., Geoffroy, M.H.: Inexact Newton methods and Dennis-More theorems for nonsmooth generalized equations. SIAM J. Control Optim. 53, 1003–1019 (2015)
9. Cibulka, R., Dontchev, A.L., Preininger, J., Roubal, T., Veliov, V.: Kantorovich-type theorems for generalized equations. Research Report 2015-16, TU Wien (2015)
10. Clarke, F.H.: On the inverse function theorem. Pac. J. Math. 64, 97–102 (1976)
11. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
12. Dontchev, A.L.: Local convergence of the Newton method for generalized equation. C. R. Acad. Sci. Paris, Ser. I 322, 327–331 (1996)
13. Dontchev, A.L.: Local Analysis of a Newton-type Method Based on Partial Linearization. Lectures in Applied Mathematics 32, 295–306 (1996)
14. Dontchev, A.L., Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. 6, 1087–1105 (1996)
15. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings. A View from Variational Analysis. Springer, Dortrecht, Heidelberg, London, New York (2009) (2nd edition 2014)
16. Dontchev, A.L., Rockafellar, R.T.: Newton's method for generalized equations: a sequential implicit function theorem. Math. Program. Ser. B 123, 139–159 (2010)
17. F. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementary Problems, Vol II. Springer, New York (2003)
18. Fusek, P.: Isolated zeros of Lipschitzian metrically regular R^n functions. Optimization 49, 425–446 (2001)
19. Griewank, A.: The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces. SIAM J. Numer. Anal. 24, 684-705 (1987)
20. Griewank, A.: On stable piecewise linearization and generalized algorithmic differentiation. Optim. Methods Softw. 28, 1139–1178 (2013)
21. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semi-smooth Newton method. SIAM J. Optim. 13 865–888 (2003)
22. T. Hoheisel, T., Kanzow, C., Mordukhovich, B.S.: Phan, H.: Generalized Newton's method based on graphical derivatives. Nonlinear Anal. 75, 1324-1340 (2012)
23. Ito, K., Kunisch, K.: Lagrange Multiplier Approach to Variational Problems and Applications. SIAM, Philadelphia (2008)
24. Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer, Cham (2014)

25. Jongen. H. Th., Jonker, P., Twilt, F.: Nonlinear Optimization in R^n, I: Morse Theory, Chebychev Approximation. Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork (1983)
26. Jongen. H. Th., Jonker, P., Twilt, F.: Nonlinear Optimization in R^n, II: Transversality, Flows, Parametric Aspects. Peter Lang Verlag, Frankfurt a.M.-Bern-NewYork (1986)
27. Josephy, N.H.: Newton's method for generalized equations and the PIES energy model. Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison (1979)
28. Kantorovich, L.W., Akilov, G.P.: Funktionalanalysis in normierten Räumen. Akademie Verlag, Berlin (1964)
29. Klatte, D., Kruger, A., Kummer, B.: From convergence principles to stability and optimality conditions. J. Convex Anal. 19, 1043–1072 (2012)
30. King, A.J., Rockafellar, R.T.: Sensitivity analysis for nonsmooth generalized equations, Math. Program. 55(2), 193-212 (1992)
31. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization - Regularity, Calculus, Methods and Applications. Kluwer Academic Publ., Dordrecht-Boston-London (2002)
32. Klatte, D., Kummer, B.: Strong Lipschitz Stability of Stationary Solutions for Nonlinear Programs and Variational Inequalities. SIAM J. Optim. 16, 96–119 (2005)
33. Klatte, D., Kummer, B.: Newton methods for stationary points: an elementary view of regularity conditions and solution schemes. Optimization 56, 441–462, (2007)
34. Klatte, D., Kummer, B.: Aubin property and uniqueness of solutions in cone constrained optimization. Math. Methods Oper. Res. 77, 291–304 (2013)
35. Kojima, M., Shindo, S.: Extension of Newton and Quasi-Newton methods to systems of PC^1 equations. J. Oper. Res. Soc. Japan 29, 352–374 (1986)
36. Kummer, B.: Newton's method for non-differentiable functions. In Guddat, J. et al. (eds.) Advances in Mathematical Optimization, Ser. Math. Res. 45, pp. 114–125. Akademie Verlag, Berlin (1988)
37. Kummer, B.: Lipschitzian inverse functions, directional derivatives and application in $C^{1,1}$ optimization, J. Optim. Theory Appl. 70, 559–580 (1991)
38. Kummer, B.: Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In Oettli, W., Pallaschke, D. (eds.) Advances in Optimization, pp. 171–194 Springer, Berlin (1992)
39. Kummer, B.: Approximation of multifunctions and superlinear convergence. In Durier, R., Michelot, C. (eds.) Recent Developments in Optimization, Lecture Notes in Econom. and Math. Systems 429, pp. 243–251. Springer, Berlin (1995)
40. Kummer, B.: Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear programming. In Fiacco, A.V. (ed.) Mathematical Programming with Data Perturbations, pp. 201–222. M. Dekker, New York (1997)
41. Kummer, B.: Newton's method for continuous functions? script for students (2012) http://www2.mathematik.hu-berlin.de/ kummer/scripts/ContNewton.pdf
42. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. 15, 957–972 (1977)
43. Mordukhovich, B.S.: Complete characterization of openness, metric regularity and Lipschitzian properties of multifunctions. Trans. Amer. Math. Soc. 340, 1–35 (1993)
44. Ortega, J.M.: The Newton-Kantorovich Theorem. Amer. Math. Monthly 75, 658-660 (1968)
45. Pang, J.-S., Qi, L.: Nonsmooth equations: motivation and algorithms. SIAM J. Optim. 3, 443–465 (1993).
46. Potra, F.A., Ptak, V.: Nondiscrete Induction and Iterative Processes. Research Notes in Mathematics 103, Pitman, Boston MA (1984)
47. Qi, L., Sun, L.: A nonsmooth version of Newton's method. Math. Program. 58, 353–367 (1993)
48. Rademacher, H.: Ueber partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln und ueber die Transformation der Doppelintegrale. Math. Ann. 79, 340-359 (1919)
49. Robinson, S.M.: Extension of Newton's method to nonlinear functions with values in a cone. Numer. Math. 19, 341–347 (1972)
50. Robinson, S.M.: Generalized equations and their solutions, Part I: Basic theory. Math. Program. Study 10, 128–141 (1979)
51. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. 5, 43–62 (1980)
52. Robinson, S.M.: Newton's method for a class of nonsmooth functions. Set-Valued Anal. 2, 291–305 (1994) (Preprint 1988)
53. Rockafellar, R.T.: Proto-differentiability of set-valued mappings and its applications in optimization, in Analyse non linéaire (Perpignan 1987), Ann. Inst. H. Poincaré, Anal. Non Linéaire, vol. 6 (suppl.), 448-482 (1989)
54. Rockafellar, R.T., Wets, J.-B.: Variational Analysis. Springer, Berlin (1998)
55. S. Scholtes: Introduction to Piecewise Differentiable Equations. Springer, New York (2012) (Preprint 1998)
56. Shapiro, A.: On concepts of directional differentiability. J. Optim. Theory Appl. 66, 477–487 (1990)
57. Sun, D., Qi, L.: On NCP functions. Comput. Optim. Appl. 13, 201–220 (1999)
58. Ulbrich, M.: Semismooth Newton methods for operator equations in function spaces. SIAM J. Optim. 13, 805-841 (2003)
59. Ulbrich, M.: Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. SIAM, Philadelphia (2011)