GENERALIZED NEWTON'S METHOD BASED ON GRAPHICAL DERIVATIVES[1]

Tim Hoheisel[1], Christian Kanzow[1], Boris S. Mordukhovich[2], and Hung Phan[2]

Preprint 298 September 2010

[1] University of Würzburg
Institute of Mathematics
Am Hubland, 97074 Würzburg, Germany
e-mail: hoheisel@mathematik.uni-wuerzburg.de
kanzow@mathematik.uni-wuerzburg.de

[2] Wayne State University
Department of Mathematics
Detroit, MI 48202, USA
e-mail: boris@math.wayne.edu
pmhung@wayne.edu

September 2, 2010

[1]The research of the third and fourth authors was partially supported by the US National Science Foundation under grants DMS-0603846 and DMS-1007132 and by the Australian Research Council under grant DP-12092508.

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

Key words. nonsmooth equations, optimization and variational analysis, Newton's method, graphical derivatives and coderivatives, local and global convergence.

AMS subject classification. 49J53, 65K15, 90C30

1 Introduction

Newton's method is one of the most powerful and useful methods in optimization and in the related area of solving systems of nonlinear equations

$$H(x) = 0 \tag{1.1}$$

defined by continuous vector-valued mappings $H\colon I\!R^n \to I\!R^n$. In the classical setting when H is a continuously differentiable (smooth, C^1) mapping, Newton's method builds the following iteration procedure

$$x^{k+1} := x^k + d^k \text{ for all } k = 0, 1, 2, \ldots, \tag{1.2}$$

where $x^0 \in I\!R^n$ is a given starting point, and where $d^k \in I\!R^n$ is a solution to the linear system of equations (often called "Newton equation")

$$H'(x^k)d = -H(x^k). \tag{1.3}$$

A detailed analysis and numerous applications of the classical Newton's method (1.2), (1.3) and its modifications can be found, e.g., in the books [7, 14, 26] and the references therein.

However, in the vast majority of applications—including those to optimization, variational inequalities, complementarity and equilibrium problems, etc.—the underlying mapping H in (1.1) is nonsmooth. Indeed, the aforementioned optimization-related models and their extensions can be written via Robinson's formalism of "generalized equations," which in turn can be reduced to standard equations of the form above (using, e.g., the projection operator) while with *intrinsically nonsmooth* mappings H; see [8, 19, 33, 29] for more details, discussions, and references.

Robinson originally proposed (see [32] and also [34] based on his earlier preprint) a point-based approximation approach to solve nonsmooth equations (1.1), which then was developed by his student Josephy [11] to extend Newton's method for solving variational inequalities and complementarity problems. Other approaches replace the classical derivative $H'(x_k)$ in the Newton equation (1.3) by some generalized derivatives. In particular, the B-differentiable Newton method developed by Pang [27, 28] uses the iteration scheme (1.2) with d^k being a solution to the subproblem

$$H'(x^k; d) = -H(x^k), \tag{1.4}$$

where $H'(x^k; d)$ denotes the classical directional derivative of H at x^k in the direction d. Besides the existence of the classical directional derivative in (1.4), a number of strong assumptions are imposed in [27, 28] to establish appropriate convergence results; see Section 5 below for more discussions and comparisons.

In another approach developed by Kummer [16] and Qi and Sun [31], the direction d^k in (1.2) is taken as a solution to the linear system of equations

$$A_k d = -H(x^k), \tag{1.5}$$

where A_k is an element of Clarke's generalized Jacobian $\partial_C H(x_k)$ of a Lipschitz continuous mapping H. In [30], Qi suggested to replace $A_k \in \partial_C H(x^k)$ in (1.5) by the choice of A_k from the so-called B-subdifferential $\partial_B H(x^k)$ of H at x^k, which is a proper subset of $\partial_C H(x^k)$; see Section 4 for more details. We also refer the reader to [8, 15, 34] and bibliographies therein for wide overviews, historical remarks, and other developments on Newton's method for nonsmooth Lipschitz equations as in (1.1) and to [13] for some recent applications.

It is proved in [31] and [30] that the Newton type method based on implementing the generalized Jacobian and B-subdifferential in (1.5), respectively, superlinearly converges to a solution of (1.1) for a class of *semismooth* mappings H; see Section 4 for the definition and discussions. This subclass of Lipschitz continuous and directionally differentiable mappings is rather broad and useful in applications to optimization-related problems. However, not every mapping arising in applications (from both theoretical and practical viewpoints) is either directionally differentiable or Lipschitz continuous. The reader can find valuable classes of functions and mappings of this type in [24, 35] and overwhelmingly in spectral function analysis, eigenvalue optimization, studying of roots of polynomials, stability of control systems, etc.; see, e.g., [4] and the references therein.

The main goal and achievements of this paper are as follows. We propose a new Newton-type algorithm to solve nonsmooth equations (1.1) described by general continuous mappings H that is based on *graphical derivatives*. It reduces to the classical Newton method (1.3) when H is smooth, being different from previously known versions of Newton's method in the case of Lipschitz continuous mappings H. Based on advanced tools of variational analysis involving *metric regularity* and *coderivatives*, we justify well-posedness of the new algorithm and its *superlinear local* and *global* (of the Kantorovich type) convergence under verifiable assumptions that hold for semismooth mappings but are *not* restricted to them. Detailed comparisons of our algorithm and results with the semismooth and B-differentiable Newton methods are given and certain improvements of these methods are justified.

Note metric regularity and related concepts of variational analysis has been employed in the analysis and justification of numerical algorithms starting with Robinson's seminal contribution; see, e.g., [1, 18, 25] and their references for the recent account. However, we are not familiar with any usage of graphical derivatives and coderivatives for these purposes.

The rest of the paper is organized as follows. In Section 2 we present basic definitions and preliminaries from variational analysis and generalized differentiation widely used for formulations and proofs of the main results.

Section 3 is devoted to the description of the new generalized Newton algorithm with justifying its well-posedness/solvability and establishing its superlinear local and global convergence under appropriate assumptions on the underlying mapping H.

In Section 4 we compare our algorithm with the scheme of (1.5). We also discuss in detail the major assumptions made in Section 3 deriving sufficient conditions for their fulfillment and comparing them with those in the semismooth Newton methods.

Section 5 contains applications of our algorithm to the B-differentiable Newton method (1.4) with largely relaxed assumptions in comparison with known ones. In Section 6 we give some concluding remarks and discussions on further research.

Our notation is basically standard in variational analysis and numerical optimization;

cf. [8, 24, 35]. Recall that, given a set-valued mapping $F\colon I\!R^n \rightrightarrows I\!R^m$, the expression

$$\mathop{\mathrm{Lim\,sup}}_{x\to\bar{x}} F(x) := \Big\{ y \in I\!R^m \Big|\ \exists\, x_k \to \bar{x} \text{ and } y_k \to y \text{ as } k \to \infty \text{ with } y_k \in F(x_k) \text{ for all } k \in I\!N := \{1,2,\ldots\} \Big\} \tag{1.6}$$

defines the *Painlevé-Kuratowski upper/outer limit* of F as $x \to \bar{x}$. Let us also mention that the symbols $\mathrm{cone}\,\Omega$ and $\mathrm{co}\,\Omega$ stand, respectively, for the conic hull and convex hull of the set in question, that $\mathrm{dist}(x;\Omega)$ denotes the Euclidean distance between a point $x \in I\!R^n$ and a set Ω, and that the notation A^T signifies the matrix transposition. As usual, $B_\varepsilon(\bar{x})$ stands for the closed ball centered at $\bar{x}$ with radius $\varepsilon > 0$.

2 Tools of Variational Analysis

In this section we briefly review some constructions and results from variational analysis and generalized differentiation widely used in what follows. The reader may consult the texts [3, 24, 35, 36] for more details and additional material.

Given a nonempty set $\Omega \subset I\!R^n$ and a point $\bar{x} \in \Omega$, the (Bouligand-Severi) *tangent/contingent cone* to Ω at $\bar{x}$ is defined by

$$T(\bar{x};\Omega) := \mathop{\mathrm{Lim\,sup}}_{t\downarrow 0} \frac{\Omega - \bar{x}}{t} \tag{2.1}$$

via the outer limit (1.6). This cone is often nonconvex while its polar/dual cone

$$\widehat{N}(\bar{x};\Omega) := \big\{ p \in I\!R^n \big|\ \langle p, u\rangle \le 0 \text{ for all } u \in T(\bar{x};\Omega) \big\} \tag{2.2}$$

is always convex and can be intrinsically described by

$$\widehat{N}(\bar{x};\Omega) = \Big\{ p \in I\!R^n \Big|\ \limsup_{x \xrightarrow{\Omega} \bar{x}} \frac{\langle p, x-\bar{x}\rangle}{\|x-\bar{x}\|} \le 0 \Big\}, \quad \bar{x} \in \Omega,$$

where the symbol $x \xrightarrow{\Omega} \bar{x}$ signifies that $x \to \bar{x}$ with $x \in \Omega$. The construction (2.2) is known as the *prenormal cone* or the *Fréchet/regular normal cone* to Ω at $\bar{x} \in \Omega$. For convenience we put $\widehat{N}(\bar{x};\Omega) = \emptyset$ if $\bar{x} \notin \Omega$. Observe that the prenormal cone (2.2) may not have natural properties of generalized normals in the case of nonconvex sets Ω; e.g., it often happens that $\widehat{N}(\bar{x};\Omega) = \{0\}$ when $\bar{x}$ is a boundary point of Ω and the cone (2.2) does not possesses required calculus rules. The situation is dramatically improved when we consider a robust regularization of (2.2) via the outer limit (1.6) and arrive at the construction

$$N(\bar{x};\Omega) := \mathop{\mathrm{Lim\,sup}}_{x\to\bar{x}} \widehat{N}(x;\Omega) \tag{2.3}$$

known as the (limiting, basic, Mordukhovich) *normal cone* to Ω at $\bar{x} \in \Omega$. If Ω is locally closed around $\bar{x}$, the basic normal cone (2.3) can be equivalently described as

$$N(\bar{x};\Omega) = \mathop{\mathrm{Lim\,sup}}_{x\to\bar{x}} \big[\mathrm{cone}\big(x - \Pi(x;\Omega)\big)\big], \quad \bar{x} \in \Omega,$$

via the Euclidean projector $\Pi(\cdot;\Omega)$ on Ω; this was in fact the original definition of the normal cone in [21]. Despite its nonconvexity, the normal cone (2.3) and the corresponding subdifferential and coderivative constructions for extended-real-valued functions and set-valued mappings enjoy comprehensive calculus rules, which are particularly based on variational/extremal principles of variational analysis.

Consider next a set-valued mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ with the graph

$$\operatorname{gph} F := \big\{(x,y) \in I\!R^n \times I\!R^m \big|\ y \in F(x)\big\}$$

and define the graphical derivative and coderivative constructions generated by the tangent and normal cones, respectively. Given $(\bar{x},\bar{y}) \in \operatorname{gph} F$, the *graphical/contingent derivative* of F at $(\bar{x},\bar{y})$ is introduced in [2] as a mapping $DF(\bar{x},\bar{y})\colon I\!R^n \rightrightarrows I\!R^m$ with the values

$$DF(\bar{x},\bar{y})(z) := \big\{w \in I\!R^m \big|\ (z,w) \in T\big((\bar{x},\bar{y});\operatorname{gph} F\big)\big\}, \quad z \in I\!R^n, \tag{2.4}$$

defined via the contingent cone (2.1) to the graph of F at the point $(\bar{x},\bar{y})$; see [3, 35] for various properties, equivalent representation, and applications. The *coderivative* of F at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ is introduced in [22] as a mapping $D^*F(\bar{x},\bar{y})\colon I\!R^m \rightrightarrows I\!R^n$ with the values

$$D^*F(\bar{x},\bar{y})(v) := \big\{u \in I\!R^n \big|\ (u,-v) \in N\big((\bar{x},\bar{y});\operatorname{gph} F\big)\big\}, \quad v \in I\!R^m, \tag{2.5}$$

defined via the normal cone (2.3) to the graph of F at $(\bar{x},\bar{y})$; see [24, 35] for extended calculus and a variety of applications. We drop $\bar{y}$ in the graphical derivative and coderivative notation when the mapping in question is single-valued at $\bar{x}$. Note that the graphical derivative and coderivative constructions in (2.4) and (2.5) are *not dual* to each other, since the basic normal cone (2.3) is *nonconvex* and hence cannot be tangentially generated.

In this paper we employ, together with (2.4) and (2.5), the following modified derivative construction for mappings, which seems to be new in generality although constructions of this (radial, Dini-like) type have been widely used for extended-real-valued functions.

Definition 2.1 (restrictive graphical derivative of mappings). *Let $F\colon I\!R^n \rightrightarrows I\!R^m$, and let $(\bar{x},\bar{y}) \in \operatorname{gph} F$. Then a set-valued mapping $\widetilde{D}F(\bar{x},\bar{y})\colon I\!R^n \rightrightarrows I\!R^m$ given by*

$$\widetilde{D}F(\bar{x},\bar{y})(z) := \operatorname*{Lim\,sup}_{t\downarrow 0} \frac{F(\bar{x}+tz)-\bar{y}}{t}, \quad z \in I\!R^n, \tag{2.6}$$

is called the RESTRICTIVE GRAPHICAL DERIVATIVE *of F at $(\bar{x},\bar{y})$.*

The next proposition collects some properties of the graphical derivative (2.4) and its restrictive counterpart (2.6) needed in what follows.

Proposition 2.2 (properties of graphical derivatives). *Let $F\colon I\!R^n \rightrightarrows I\!R^m$, and let $(\bar{x},\bar{y}) \in \operatorname{gph} F$. Then the following assertions hold:*

(a) *We have $\widetilde{D}F(\bar{x},\bar{y})(z) \subset DF(\bar{x},\bar{y})(z)$ for all $z \in I\!R^n$.*

(b) *There are inverse derivative relationships*

$$DF(\bar{x},\bar{y})^{-1} = DF^{-1}(\bar{y},\bar{x}) \ \text{ and } \ \widetilde{D}F(\bar{x},\bar{y})^{-1} = \widetilde{D}F^{-1}(\bar{y},\bar{x}).$$

(c) *If F is single-valued and locally Lipschitzian around $\bar{x}$, then*

$$\widetilde{D}F(\bar{x})(z) = DF(\bar{x})(z) \;\; \textit{for all} \;\; z \in I\!\!R^n.$$

(d) *If F is single-valued and directionally differentiable at $\bar{x}$, then*

$$\widetilde{D}F(\bar{x})(z) = \big\{F'(\bar{x}; z)\big\} \;\; \textit{for all} \;\; z \in I\!\!R^n.$$

(e) *If F is single-valued and Gâteaux differentiable at $\bar{x}$ with the Gâteaux derivative $F'_G(\bar{x})$, then we have*

$$\widetilde{D}F(\bar{x})(z) = \big\{F'_G(\bar{x})z\big\} \;\; \textit{for all} \;\; z \in I\!\!R^n.$$

(f) *If F is single-valued and (Fréchet) differentiable at $\bar{x}$ with the derivative $F'(\bar{x})$, then*

$$DF(\bar{x})(z) = \big\{F'(\bar{x})z\big\} \;\; \textit{for all} \;\; z \in I\!\!R^n.$$

Proof. It is shown in [35, 8(14)] that the graphical derivative (2.4) admits the representation

$$DF(\bar{x}, \bar{y})(z) = \mathop{\mathrm{Lim\,sup}}_{t \downarrow 0,\, h \to z} \frac{F(\bar{x} + th) - \bar{y}}{t}, \quad z \in I\!\!R^n. \tag{2.7}$$

The inclusion in (a) is an immediate consequence of Definition 2.1 and representation (2.7).

The first equality in (b), observed from the very beginning [2], easily follows from definition (2.4). We can similarly check the second one in (b).

To justify the equality in (c), it remains to verify by (a) the opposite inclusion '$\supset$' when F is single-valued and locally Lipschitzian around $\bar{x}$. In this case fix $z \in I\!\!R^n$, pick any $w \in DF(\bar{x})(z)$, and find by representation (2.7) sequences $h_k \to z$ and $t_k \downarrow 0$ such that

$$\frac{F(\bar{x} + t_k h_k) - F(\bar{x})}{t_k} \to w \;\; \text{as} \;\; k \to \infty.$$

The local Lipschitz continuity of F around $\bar{x}$ with constant $L \geq 0$ implies that

$$\begin{aligned} \left\| \frac{F(\bar{x} + t_k h_k) - F(\bar{x})}{t_k} - \frac{F(\bar{x} + t_k z) - F(\bar{x})}{t_k} \right\| &= \left\| \frac{F(\bar{x} + t_k h_k) - F(\bar{x} + t_k z)}{t_k} \right\| \\ &\leq L \|h_k - z\| \end{aligned}$$

for all $k \in I\!\!N$ sufficiently large, and hence we have the convergence

$$\frac{F(\bar{x} + t_k z) - F(\bar{x})}{t_k} \to w \;\; \text{as} \;\; k \to \infty.$$

Thus $w \in \widetilde{D}F(\bar{x})(z)$, which justifies (c). Assertions (d) and (e) follow directly from the definitions. Finally, assertion (f) is implied by (e) in the local Lipschitzian case (c) while it can be easily derived from the (Fréchet) differentiability of F at $\bar{x}$ with no Lipschitz assumption; see, e.g., [35, Exercise 9.25(b)]. $\triangle$

Proposition 2.2 reveals important differences between the graphical derivative (2.4) and the coderivative (2.5). Indeed, assertions (c) and (d) of this proposition show that the graphical derivative of locally Lipschitzian and *directionally differentiable* mappings

$F\colon I\!R^n \to I\!R^m$ is always *single-valued.* At the same time, the coderivative single-valuedness for locally Lipschitzian mappings is equivalent to the *strict/strong* Fréchet *differentiability* of F at the point in question; see [24, Theorem 3.66]. It follows from the well-known formula

$$\mathrm{co} D^* F(\bar{x})(z) = \big\{ A^T z \big|\ A \in \partial_C F(\bar{x}) \big\} \tag{2.8}$$

that the latter strict differentiability condition characterizes also the single-valuedness of the generalized Jacobian of F at $\bar{x}$.

In fact, in the case of $F = (f_1, \ldots, f_m)\colon I\!R^n \to I\!R^m$ being locally Lipschitzian around $\bar{x}$ the coderivative (2.5) admits the subdifferential description

$$D^* F(\bar{x})(z) = \partial \Big(\sum_{i=1}^{m} z_i f_i \Big)(\bar{x}) \ \text{ for any } \ z = (z_1, \ldots, z_m) \in I\!R^m, \tag{2.9}$$

where the (basic, limiting, Mordukhovich) *subdifferential* $\partial f(\bar{x})$ of a general scalar function f at $\bar{x}$ is defined geometrically by

$$\partial f(\bar{x}) := \big\{ p \in I\!R^n \big|\ (p, -1) \in N\big((\bar{x}, f(\bar{x})); \mathrm{epi}\, f \big) \big\} \tag{2.10}$$

via the normal cone (2.3) to the epigraph $\mathrm{epi}\, f := \{ (x, \mu) \in I\!R^{n+1} |\ \mu \geq f(x) \}$ and admits analytical descriptions in terms of the outer limit (1.6) of the Fréchet/regular and proximal subdifferentials at points nearby; see [24, 35] with the references therein. Note also that the basic subdifferential (2.10) of a continuous function f can be also described via the coderivative of f by $\partial f(\bar{x}) = D^* f(\bar{x})(1)$; see [24, Theorem 1.80].

Finally in this section, we recall the notion of metric regularity and its coderivative characterization that play a significant role in the paper. A mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ is *metrically regular* around $(\bar{x}, \bar{y}) \in \mathrm{gph}\, F$ if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ as well as a number $\mu > 0$ such that

$$\mathrm{dist}\big(x; F^{-1}(y) \big) \leq \mu\, \mathrm{dist}\big(y; F(x) \big) \ \text{ for all } \ x \in U \ \text{ and } \ y \in V. \tag{2.11}$$

Observe that it is sufficient to require the fulfillment of (2.11) just for those $y \in V$ satisfying the estimate $\mathrm{dist}(y; F(x)) \leq \gamma$ for some $\gamma > 0$; see [24, Proposition 1.48].

We will see below that metric regularity is crucial for justifying the well-posedness of our generalized Newton algorithm and establishing its local and global convergence. It is also worth mentioning that, in the opposite direction, a Newton-type method (known as the Lyusternik-Graves iterative process) leads to verifiable conditions for metric regularity of smooth mappings; see, e.g., the proof of [24, Theorem 1.57] and the commentaries therein. The latter procedure is replaced by variational/extremal principles of variational analysis in the case of nonsmooth and set-valued mappings under consideration; cf. [10, 24, 35].

In this paper we broadly use the following *coderivative characterization* of the metric regularity property for an arbitrary set-valued mapping F with closed graph, known also as the Mordukhovich criterion (see [23, Theorem 3.6], [35, Theorem 9.45], and the references therein): F is metrically regular around $(\bar{x}, \bar{y})$ *if and only if* the inclusion

$$0 \in D^* F(\bar{x}, \bar{y})(z) \ \text{ implies that } \ z = 0, \tag{2.12}$$

which amounts the kernel condition $\ker D^* F(\bar{x}, \bar{y}) = \{0\}$.

3 The Generalized Newton Algorithm

This section presents the main contribution of the paper: a new generalized Newton method for nonsmooth equations, which is based on graphical derivatives. The section consists of three parts. In Subsection 3.1 we precisely describe the algorithm and justify its well-posedness/solvability. Subsection 3.2 contains a local superlinear convergence result under appropriate assumptions. Finally, in Subsection 3.3 we establish a global convergence result of the Kantorovich type for our generalized Newton algorithm.

3.1 Description and Justification of the Algorithm

Keeping in mind the classical scheme of the smooth Newton method in (1.2), (1.3) and taking into account the graphical derivative representation of Proposition 2.2(f), we propose an extension of the Newton equation (1.3) to nonsmooth mappings given by:

$$-H(x^k) \in DH(x^k)(d^k), \quad k = 0, 1, 2, \ldots. \tag{3.1}$$

This leads us to the following generalized Newton algorithm to solve (1.1):

Algorithm 3.1 (generalized Newton's method).

Step 0: *Choose a starting point* $x^0 \in I\!\!R^n$.

Step 1: *Check a suitable termination criterion.*

Step 2: *Compute* $d^k \in I\!\!R^n$ *such that* (3.1) *holds.*

Step 3: *Set* $x^{k+1} := x^k + d^k$, $k \leftarrow k + 1$, *and go to Step* 1.

The proposed Algorithm 3.1 does not require a priori any assumptions on the underlying mapping $H\colon I\!\!R^n \to I\!\!R^n$ in (1.1) besides its continuity, which is the standing assumption in this paper. Other assumptions are imposed below to justify the well-posedness and (local and global) convergence of the algorithm. Observe that Proposition 2.2(c,d) ensures that Algorithm 3.1 reduces to scheme (1.4) in the B-differentiable Newton method provided that H is directionally differentiable and locally Lipschitzian around the solution point in question. In Section 5 we consider in detail relationships with known results for the B-differentiable Newton method, while Section 4 compares Algorithm 3.1 and the assumptions made with the corresponding semismooth versions in the framework of (1.5).

To proceed further, we need to make sure that the generalized Newton equation (3.1) is *solvable*, which is a major part of the well-posedness of Algorithm 3.1. The next proposition shows that an appropriate assumption to ensure the solvability of (3.1) is *metric regularity.*

Proposition 3.2 (solvability of the generalized Newton equation). *Assume that* $H\colon I\!\!R^n \to I\!\!R^n$ *is metrically regular around* $\bar{x}$ *with* $\bar{y} = H(\bar{x})$ *in* (2.11), *i.e., we have* $\ker D^*H(\bar{x}) = \{0\}$. *Then there is a constant* $\varepsilon > 0$ *such that for all* $x \in B_\varepsilon(\bar{x})$ *the equation*

$$-H(x) \in DH(x)(d) \tag{3.2}$$

admits a solution $d \in I\!\!R^n$. *Furthermore, the set* $S(x)$ *of solutions to* (3.2) *is computed by*

$$S(x) = \operatorname*{Lim\,sup}_{t\downarrow 0,\, h\to -H(x)} \frac{H^{-1}\big(H(x) + th\big) - x}{t} \neq \emptyset. \tag{3.3}$$

Proof. By the assumed metric regularity (2.11) of H we find a number $\mu > 0$ and neighborhoods U of $\bar{x}$ and V of $H(\bar{x})$ such that

$$\operatorname{dist}\big(x; H^{-1}(y)\big) \le \mu \operatorname{dist}\big(y; H(x)\big) \quad \text{for all} \quad x \in U \quad \text{and} \quad y \in V.$$

Pick now an arbitrary vector $x \in U$ and select sequences $h_k \to -H(x)$ and $t_k \downarrow 0$ as $k \to \infty$. Suppose with no loss of generality that $H(x) + t_k h_k \in V$ for all $k \in I\!N$. Then we have

$$\operatorname{dist}\big(x; H^{-1}(H(x) + t_k h_k)\big) \le \mu t_k \|h_k\|, \quad k \in I\!N,$$

and hence there is a vector $u_k \in H^{-1}(H(x) + t_k h_k)$ such that $\|u_k - x\| \le \mu t_k \|h_k\|$ for all $k \in I\!N$. This shows that the sequence $\{\|u_k - x\|/t_k\}$ is bounded, and thus it contains a subsequence that converges to some element $d \in I\!R^n$. Passing to the limit as $k \to \infty$ and recalling the definitions of the outer limit (1.6) and of the tangent cone (2.1), we arrive at

$$\big(d, -H(x)\big) \in \operatorname*{Lim\,sup}_{t \downarrow 0} \frac{\operatorname{gph} H - \big(x, H(x)\big)}{t} = T\big((x, H(x)); \operatorname{gph} H\big),$$

which justifies the desired inclusion (3.2). The solution representation (3.3) follows from (2.7) and Proposition 2.2(b) in the case of single-valued mappings, since

$$S(x) = DH(x)^{-1}\big(-H(x)\big)$$

due to (3.2). This completes the proof of the proposition. $\triangle$

3.2 Local Convergence

In this subsection we first formulate major assumptions of our generalized Newton method and then show that they ensure the superlinear local convergence of Algorithm 3.1.

(H1) There exist a constant $C > 0$, a neighborhood U of $\bar{x}$, and a neighborhood V of the origin in $I\!R^n$ such that the following holds:
For all $x \in U$, $z \in V$, and for any $d \in I\!R^n$ with $-H(x) \in DH(x)(d)$ there is a vector $w \in \widetilde{D}H(x)(z)$ such that

$$C\|d - z\| \le \|w + H(x)\| + o(\|x - \bar{x}\|).$$

(H2) There exists a neighborhood U of $\bar{x}$ such that for all $v \in \widetilde{D}H(x)(\bar{x} - x)$ we have

$$\|H(x) - H(\bar{x}) + v\| = o(\|x - \bar{x}\|).$$

A detailed discussion of these two assumptions and sufficient conditions for their fulfillment are given in Section 4. Note that assumption (H2) means, in the terminology of [8, Definition 7.2.2] focused on locally Lipschitzian mappings H, that the family $\{\widetilde{D}H(x)\}$ provides a *Newton approximation scheme* for H at $\bar{x}$.

Now we establish our principal local convergence result that makes use of the major assumptions (H1) and (H2) together with metric regularity.

Theorem 3.3 (superlinear local convergence of the generalized Newton method). *Let $\bar{x} \in I\!R^n$ be a solution to* (1.1) *for which the underlying mapping $H\colon I\!R^n \to I\!R^n$ is metrically regular around $\bar{x}$ and assumptions* (H1) *and* (H2) *are satisfied. Then there is a number $\varepsilon > 0$ such that for all $x^0 \in B_\varepsilon(\bar{x})$ the following assertions hold:*

(i) *Algorithm* 3.1 *is well defined and generates a sequence $\{x^k\}$ converging to $\bar{x}$.*

(ii) *The rate of convergence $x^k \to \bar{x}$ is at least superlinear.*

Proof. To justify (i), pick $\varepsilon > 0$ such that assumptions (H1) and (H2) hold with $U := B_\varepsilon(\bar{x})$ and $V := B_\varepsilon(0)$ and such that Proposition 3.2 can be applied. Then we choose a starting point $x^0 \in B_\varepsilon(\bar{x})$ and conclude by Proposition 3.2 that the subproblem

$$-H(x^0) \in DH(x^0)(d)$$

has a solution d^0. Thus the next iterate $x^1 := x^0 + d^0$ is well defined. Let further $z^0 := \bar{x} - x^0$ and get $\|z^0\| \le \varepsilon$ by the choice of the starting point x^0. By assumption (H1), find a vector $w^0 \in \widetilde{D}H(x^0)(z^0)$ such that

$$C\|x^1 - \bar{x}\| = C\|(x^1 - x^0) - (\bar{x} - x^0)\| = C\|d^0 - z^0\| \le \|w^0 + H(x^0)\| + o(\|x^0 - \bar{x}\|).$$

Taking this into account and employing assumption (H2), we get the relationships

$$\begin{array}{rcl} C\|x^1 - \bar{x}\| & \le & \|w^0 + H(x^0)\| + o(\|x^0 - \bar{x}\|) \\ & = & \|H(x^0) - H(\bar{x}) + w^0\| + o(\|x^0 - \bar{x}\|) \\ & = & o(\|x^0 - \bar{x}\|) \\ & \le & \frac{C}{2}\|x^0 - \bar{x}\|, \end{array}$$

which imply that $\|x^1 - \bar{x}\| \le \frac{1}{2}\|x^0 - \bar{x}\|$. The latter yields, in particular, that $x^1 \in B_\varepsilon(\bar{x})$. Now standard induction arguments allow us to conclude that the iterative sequence $\{x^k\}$ generated by Algorithm 3.1 is well defined and converges to the solution $\bar{x}$ of (1.1) with at least a linear rate. This justifies assertion (i) of the theorem.

Next we prove assertion (ii) showing that the convergence $x^k \to \bar{x}$ is in fact *superlinear* under the validity of assumption (H2). To proceed, we basically follow the proof of assertion (i) and construct by induction sequences $\{d^k\}$ satisfying

$$-H(x^k) \in DH(x^k)(d^k) \;\text{ for all }\; k \in I\!N,$$

$\{z^k\}$ with $z^k := \bar{x} - x^k$, and $\{w^k\}$ with $w^k \in \widetilde{D}H(x^k)(z^k)$ such that

$$C\|x^{k+1} - \bar{x}\| \le \|w^k + H(x^k)\| + o(\|x^k - \bar{x}\|), \quad k \in I\!N.$$

Applying then assumption (H2) gives us the relationships

$$C\|x^{k+1} - \bar{x}\| \le \|H(x^k) - H(\bar{x}) + w^k\| + o(\|x^k - \bar{x}\|) = o(\|x^k - \bar{x}\|),$$

which ensure the superlinear convergence of the iterative sequence $\{x^k\}$ to the solution $\bar{x}$ of (1.1) and thus complete the proof of the theorem. △

3.3 Global Convergence

Besides the local convergence in the classical Newton method based on suitable assumptions imposed at the (unknown) solution of the underlying system of equations, there are global (or semi-local) convergence results of the Kantorovich type [12] for smooth systems of equations which show that, under certain conditions at the starting point x^0 and a number of assumptions to hold in a suitable region around x^0, Newton's iterates are well defined and converge to a solution belonging to this region; see [7, 12] for more details and references. In the case of nonsmooth equations (1.1) results of the Kantorovich type were obtained in [31, 34] for the corresponding versions of Newton's method. Global convergence results of different types can be found in, e.g., [6, **?**, 9, 28] and their references.

Here is a global convergence result for our generalized Newton method to solve (1.1).

Theorem 3.4 (global convergence of the generalized Newton method). *Let* x^0 *be a starting point of Algorithm* 3.1, *and let*

$$\Omega := \big\{x \in I\!R^n \big|\ \|x - x^0\| \le r\big\} \tag{3.4}$$

with some $r > 0$. *Impose the following assumptions:*

(a) *The mapping* $H\colon I\!R^n \to I\!R^n$ *in* (1.1) *is metrically regular on* Ω *with modulus* $\mu > 0$, *i.e., it is metrically regular around every point* $x \in \Omega$ *with the same modulus* μ.

(b) *The set-valued map* $DH(x)(z)$ *uniformly on* Ω *converges to* $\{0\}$ *as* $z \to 0$ *in the sense that: for all* $\varepsilon > 0$ *there is* $\delta > 0$ *such that*

$$\|w\| \le \varepsilon \ \text{ whenever } \ w \in DH(x)(z),\ \|z\| \le \delta, \ \text{ and } \ x \in \Omega.$$

(c) *There is* $\alpha \in (0, 1/\mu)$ *such that*

$$\mu\|H(x^0)\| \le r(1 - \alpha\mu) \tag{3.5}$$

and for all $x, y \in \Omega$ *we have the estimate*

$$\|H(x) - H(y) - v\| \le \alpha\|x - y\| \ \text{ whenever } \ v \in DH(x)(y - x). \tag{3.6}$$

Then Algorithm 3.1 *is well defined, the sequence of iterates* $\{x^k\}$ *remains in* Ω *and converges to a solution* $\bar{x} \in \Omega$ *of* (1.1). *Moreover, we have the error estimate*

$$\|x^k - \bar{x}\| \le \frac{\alpha\mu}{1 - \alpha\mu}\|x^k - x^{k-1}\| \ \text{ for all } \ k \in I\!N. \tag{3.7}$$

Proof. The metric regularity assumption (a) allows us to employ Proposition 3.2 and, for any $x \in \Omega$ and $d \in I\!R^n$ satisfying the inclusion $-H(x) \in DH(x)(d)$, to find sequences of $h_k \to -H(x)$ and $t_k \downarrow 0$ as $k \to \infty$ such that

$$\|d\| = \lim_{k\to\infty}\left\|\frac{H^{-1}\big(H(x) + t_k h_k\big) - x}{t_k}\right\| \le \lim_{k\to\infty}\mu\|h_k\| = \mu\|H(x)\|.$$

In view of assumption (3.5) in (c) and the iteration procedure of the algorithm, this implies

$$\|x^1 - x^0\| = \|d^0\| \le \mu\|H(x^0)\| \le r(1-\alpha\mu),$$

which ensures that $x^1 \in \Omega$ due the form of Ω in (3.4) and the choice of α. Proceeding further by induction, suppose that $x^1, \ldots, x^k \in \Omega$ and get the relationships

$$\begin{aligned}\|x^{k+1} - x^k\| = \|d^k\| &\le \mu\|H(x^k)\| \\ &\le \mu\|H(x^k) - H(x^{k-1}) + H(x^{k-1})\| \\ &\le \alpha\mu\|x^k - x^{k-1}\| \quad \Big(\text{using (3.6) and } - H(x^{k-1}) \in DH(x^{k-1})(x^k - x^{k-1})\Big) \\ &\le (\alpha\mu)^k\|x^1 - x^0\| \le r(\alpha\mu)^k(1-\alpha\mu),\end{aligned}$$

which imply the estimates

$$\|x^{k+1} - x^0\| \le \sum_{j=0}^{k} \|x^{j+1} - x^j\| \le \sum_{j=0}^{k} r(\alpha\mu)^j(1-\alpha\mu) \le r$$

and hence justify that $x^{k+1} \in \Omega$. Thus all the iterates generated by Algorithm 3.1 remain in Ω. Furthermore, for any natural numbers k and m, we have

$$\|x^{k+m+1} - x^k\| \le \sum_{j=k}^{k+m} \|x^{j+1} - x^j\| \le \sum_{j=k}^{k+m} r(\alpha\mu)^j(1-\alpha\mu) \le r(\alpha\mu)^k,$$

which shows that the generated sequence $\{x^k\}$ is a Cauchy sequence. Hence it converges to some point $\bar{x}$ that obviously belongs to the underlying closed set (3.4).

To show next that $\bar{x}$ is a solution to the original equation (1.1), we pass to the limit as $k \to \infty$ in the iterative inclusion

$$-H(x^k) \in DH(x^k)(x^{k+1} - x^k), \quad k \in I\!N. \tag{3.8}$$

It follows from assumption (b) that $\lim_{k\to\infty} H(x^k) = 0$. The continuity of H then implies that $H(\bar{x}) = 0$, i.e., $\bar{x}$ is a solution to (1.1).

It remains to justify the error estimate (3.7). To this end, first observe by (3.5) that

$$\|x^{k+m+1} - x^k\| \le \sum_{j=k}^{k+m} \|x^{j+1} - x^j\| \le \sum_{j=0}^{m} (\alpha\mu)^{j+1}\|x^k - x^{k-1}\| \le \frac{\alpha\mu}{1-\alpha\mu}\|x^k - x^{k-1}\|$$

for all $k, m \in I\!N$. Passing now to the limit as $m \to \infty$, we arrive at (3.7) thus completes the proof of the theorem. △

4 Discussion of Major Assumptions and Comparison with Semismooth Newton Methods

In this section we pursue a twofold goal: to discuss the major assumptions made in Section 3 and to compare our generalized Newton method based on graphical derivatives with the semismooth versions of the generalized Newton method developed in [30, 31]. As we will see from the discussions below, these two aims are largely interrelated.

Let us begin with sufficient conditions for metric regularity in terms of the constructions used in the semismooth versions of the generalized Newton method. Given a locally Lipschitz continuous vector-valued mapping $H\colon I\!R^n \to I\!R^m$, we have by the classical Rademacher theorem that the set of points

$$S_H := \{x \in I\!R^n \big|\ H \text{ is differentiable at } x\} \tag{4.1}$$

is of full Lebesgue measure in $I\!R^n$. Thus for any mapping $H\colon I\!R^n \to I\!R^m$ locally Lipschitzian around $\bar{x}$ the set

$$\partial_B H(\bar{x}) := \Big\{\lim_{k\to\infty} H'(x^k)\Big|\ \exists\,\{x^k\} \subset S_H \text{ with } x^k \to \bar{x}\Big\} \tag{4.2}$$

is nonempty and obviously compact in $I\!R^m$. It was introduced in [38] for $m = 1$ as the set of "almost-gradients" and then was called in [30] the *B-subdifferential* of H at $\bar{x}$. Clarke's *generalized Jacobian* [5] of H at $\bar{x}$ is defined by the convex hull

$$\partial_C H(\bar{x}) := \mathrm{co}\big\{\partial_B H(\bar{x})\big\}. \tag{4.3}$$

We also make use of the *Thibault derivative/limit set* [39] (called sometimes the "strict graphical derivative" [35]) of H at $\bar{x}$ defined by

$$D_T H(\bar{x})(z) := \mathop{\mathrm{Lim\,sup}}_{\substack{x\to\bar{x}\\ t\downarrow 0}} \frac{H(x+tz)-H(x)}{t}, \quad z \in I\!R^n. \tag{4.4}$$

Observe the known relationships [15, 39] between the above derivative sets

$$\partial_B H(\bar{x})z \subset D_T H(\bar{x})(z) \subset \partial_C H(\bar{x})z, \quad z \in I\!R^n. \tag{4.5}$$

The next result gives a sufficient condition for metric regularity of Lipschitzian mappings in terms of the Thibault derivative (4.4). It can be derived from the coderivative characterization of metric regularity (2.12), while we give here a direct independent proof.

Proposition 4.1 (sufficient condition for metric regularity in terms of Thibault's derivative). *Let $H\colon I\!R^n \to I\!R^n$ be locally Lipschitzian around $\bar{x}$, and let*

$$0 \notin D_T H(\bar{x})(z) \ \text{ whenever } \ z \neq 0. \tag{4.6}$$

Then the mapping H is metrically regular around $\bar{x}$.

Proof. Kummer's inverse function theorem [17, Theorem 1.1] ensures that condition (4.6) implies (actually is equivalent to) the fact that there are neighborhoods U of $\bar{x}$ and V of $H(\bar{x})$ such that the mapping $H\colon U \to V$ is one-to-one with a locally Lipschitzian inverse $H^{-1}\colon V \to U$. Let $\mu > 0$ be a Lipschitz constant of H^{-1} on V. Then for all $x \in U$ and $y \in V$ we have the relationships

$$\begin{aligned}
\mathrm{dist}\big(x; H^{-1}(y)\big) &= \|x - H^{-1}(y)\|\\
&= \|H^{-1}\big(H(x)\big) - H^{-1}(y)\|\\
&\le \mu\|H(x) - y\|\\
&= \mu\,\mathrm{dist}\big(y; H(x)\big),
\end{aligned}$$

which thus justify the metric regularity of H around $\bar{x}$. $\triangle$

To proceed further with sufficient conditions for the validity of our assumption (H1), we first introduce the notion of directional boundedness.

Definition 4.2 (directional boundedness). *A mapping $H\colon I\!R^n \to I\!R^m$ is said to be* DIRECTIONALLY BOUNDED *around $\bar{x}$ if*

$$\limsup_{t\downarrow 0} \left\| \frac{H(x+tz)-H(x)}{t} \right\| < \infty \tag{4.7}$$

for all x near $\bar{x}$ and for all $z \in I\!R^n$.

It is easy to see that if H is either directionally differentiable around $\bar{x}$ or locally Lipschitzian around this point, then it is directionally bounded around $\bar{x}$. The following example shows that the converse does not hold in general.

Example 4.3 (directional bounded mappings may not be directionally differentiable). Define a real function $H\colon I\!R \to I\!R$ by

$$H(x) := \begin{cases} x\sin\left(\frac{1}{x}\right) & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

It is easy to see that this function is not directionally differentiable at $\bar{x}=0$. However, it is directionally bounded around $\bar{x}$. Indeed, for any $x \neq 0$ near $\bar{x}$ condition (4.7) holds because H is simply differentiable at $x \neq 0$. For $x = 0$ we have

$$\limsup_{t\downarrow 0} \left| \frac{H(tz)-H(0)}{t} \right| = \limsup_{t\downarrow 0} \frac{|H(tz)|}{t} = \limsup_{t\downarrow 0} \left| z\sin\left(\frac{1}{tz}\right) \right| = |z| < \infty.$$

The next proposition and its corollary present verifiable sufficient conditions for the fulfillment of assumption (H1).

Proposition 4.4 (assumption (H1) from metric regularity). *Let $H\colon I\!R^n \to I\!R^n$, and let $\bar{x}$ be a solution to* (1.1)*. Suppose that H is metrically regular around $\bar{x}$ (i.e.,* $\ker D^*H(\bar{x}) = 0$*), that it is directionally bounded and one-to-one around this point. Then assumption* (H1) *is satisfied.*

Proof. Recall that the metric regularity of H around $\bar{x}$ is equivalent to the condition $\ker D^*H(\bar{x}) = \{0\}$ by the coderivative criterion (2.12). Let $U \subset I\!R^n$ be a neighborhood of $\bar{x}$ such that H is metrically regular and one-to-one on U. Choose further a neighborhood $V \subset I\!R^n$ of $H(\bar{x}) = 0$ from the definition of metric regularity of H around $\bar{x}$. Then pick $x \in U$, $z \in V$ and an arbitrary direction $d \in I\!R^n$ satisfying $-H(x) \in DH(x)(d)$. Employing now Proposition 3.2, we get

$$d \in \operatorname*{Lim\,sup}_{h\to -H(x),\, t\downarrow 0} \frac{H^{-1}\big(H(x)+th\big)-x}{t}.$$

By the local single-valuedness of H^{-1} and the metric regularity of H around $\bar{x}$ there exists a number $\mu > 0$ such that

$$\left\| \frac{H^{-1}(H(x)+th)-x}{t} - z \right\| \le \mu \left\| \frac{H(x)+th-H(x+tz)}{t} \right\| = \mu \left\| \frac{H(x+tz)-H(x)}{t} - h \right\|$$

for all $t > 0$ sufficiently small. It follows that

$$\|d-z\| \le \limsup_{\substack{t\downarrow 0 \\ h\to -H(x)}} \left\| \frac{H^{-1}\big(H(x)+th\big)-x}{t} - z \right\| \le \mu \limsup_{\substack{t\downarrow 0 \\ h\to -H(x)}} \left\| \frac{H(x+tz)-H(x)}{t} - h \right\| < \infty$$

by the directional boundedness of H around $\bar{x}$. The boundedness of the family

$$\Big\{ v(t) := \frac{H(x+tz)-H(x)}{t} \Big\}, \quad t \downarrow 0,$$

allows us to select a sequence $t_k \downarrow 0$ such that $v(t_k) \to w$ for some $w \in I\!R^n$. By passing to the limit above as $k \to \infty$ and employing Definition 2.1 we get that

$$w \in \widetilde{D}H(x)(z) \quad \text{and} \quad \frac{1}{\mu}\|d-z\| \le \|w+H(x)\|,$$

which completes the proof of the proposition. △

Corollary 4.5 (sufficient conditions for (H1) via Thibault's derivative). *Let $\bar{x}$ be a solution to* (1.1)*, where $H\colon I\!R^n \to I\!R^n$ is locally Lipschitzian around $\bar{x}$ and such that condition* (4.6) *holds, which is automatic when* $\det A \ne 0$ *for all $A \in \partial_C H(\bar{x})$. Then* (H1) *is satisfied with H being both metrically regular and one-to-one around $\bar{x}$.*

Proof. Indeed, both metric regularity and bijectivity of H around $\bar{x}$ assumed in Proposition 4.4 follow from Proposition 4.1 and its proof. Nonsingularity of all $A \in \partial_C H(\bar{x})$ clearly implies (4.6) by the second inclusion in (4.5). △

Note that other conditions ensuring the fulfillment of assumption (H1) for Lipschitzian and non-Lipschitzian mappings $H\colon I\!R^n \to I\!R^n$ can be formulated in terms of Warga's *derivate containers* by [40, Theorems 1 and 2] on "fat homeomorphisms" that also imply the metric regularity and one-to-one properties of H.

Next we proceed with the discussion of assumption (H2) and present, in particular, sufficient conditions for their fulfillment via semismoothness. First observe the following.

Proposition 4.6 (relationship between graphical derivative and generalized Jacobian). *Let $H\colon I\!R^n \to I\!R^m$ be locally Lipschitzian around $\bar{x}$. Then we have*

$$DH(\bar{x})(z) \subset \partial_C H(\bar{x})z \ \ \textit{for all} \ \ z \in I\!R^n. \tag{4.8}$$

Proof. Pick $w \in DH(\bar{x})(z)$ and get by Proposition 2.2(c) and Definition 2.1 a sequence of $t_k \downarrow 0$ as $k \to \infty$ such that

$$w = \lim_{k\to\infty} \frac{H(\bar{x}+t_k z)-H(\bar{x})}{t_k}. \tag{4.9}$$

It follows from [5, Proposition 2.6.5] that

$$\frac{H(\bar{x}+t_k z)-H(\bar{x})}{t_k}\in \mathrm{co}\big\{\partial_C H[\bar{x},\bar{x}+t_k z]\big\}z \;\text{ for all }\; k\in I\!N.$$

Applying to the latter the classical Carathéodory theorem, we find scalars $\gamma_i^k\in[0,t_k]$, $\lambda_i^k\in[0,1]$ and matrices $A_i^k\in\partial_C H(\bar{x}+\gamma_i^k z)$ for $i=1,\ldots,m+1$ such that

$$\frac{H(\bar{x}+t_k z)-H(\bar{x})}{t_k}=\Big[\sum_{i=1}^{m+1}\lambda_i^k A_i^k\Big]z \quad\text{and}\quad \sum_{i=1}^{m+1}\lambda_i^k=1 \;\text{ for all }\; k\in I\!N.$$

Due to the boundedness of the sequences $\{\lambda_i^k\}_{k\in I\!N}$, the convergence $\bar{x}+\gamma_i^k z\to\bar{x}$ as $k\to\infty$ for all $i=1,\ldots,m+1$, and the outer/upper semicontinuity of the mapping $x\mapsto\partial_C H(x)$ proved in [5, Proposition 2.6.2] we have that the sequences $\{A_i^k\}$ are bounded as well. Hence there are subsequences of these sequences (without relabelling), scalars $\lambda_i\in[0,1]$, and matrices A_i as $i=1,\ldots,m+1$ such that

$$\lambda_i^k\to\lambda_i,\quad \sum_{i=1}^{m+1}\lambda_i=1,\;\text{ and }\; A_i^k\to A_i\in\partial_C H(\bar{x}) \;\text{ as }\; k\to\infty.$$

By (4.9) and the subsequent relationships therein, we get

$$w=\lim_{k\to\infty}\Big[\sum_{i=1}^{m+1}\lambda_i^k A_i^k\Big]z=\Big[\sum_{i=1}^{m+1}\lambda_i A_i\Big]z\in\mathrm{co}\big\{\partial_C H(\bar{x})\big\}z=\partial_C H(\bar{x})z$$

and thus complete the proof of the proposition. $\triangle$

Inclusion (4.8)—which may be strict as illustrated by Example 4.7 below—shows that our generalized Newton Algorithm 3.1 based on the graphical derivative provides in the case of Lipschitz equations (1.1) a more accurate choice of the iterative direction d^k via (3.1) in comparison with the iterative relationship

$$-H(x^k)\in\partial_C H(x^k)d^k,\quad k=0,1,2,\ldots, \tag{4.10}$$

used in the semismooth Newton method [31] and related developments [15, 16] based on the generalized Jacobian. If in addition to the assumptions of Proposition 4.6 the mapping H is directionally differentiable at $\bar{x}$, then $DH(\bar{x})(z)=\{H'(\bar{x};z)\}$ by Proposition 2.2(c,d). Thus in this case we have from Proposition 4.6 that for any $z\in I\!R^n$ there is $A\in\partial_C H(\bar{x})$ such that $H'(\bar{x};z)=Az$, which recovers a well-known result from [31, Lemma 2.2].

The following example shows that the converse inclusion in Proposition 4.6 is not satisfied in general even with the replacement of the set $DH(\bar{x})(z)$ in (4.8) by its convex hull $\mathrm{co}DH(\bar{x})(z)$ in the case of real functions. Furthermore, the same holds if we replace the generalized Jacobian in (4.8) by the smaller B-subdifferential $\partial_B H(\bar{x})$ from (4.2).

Example 4.7 (graphical derivative is strictly smaller than B-subdifferential and generalized Jacobian). Consider the simplest nonsmooth convex function $H(x)=|x|$ on $I\!R$. In this case $\partial_B H(0)=\{-1,1\}$ and $\partial_C H(0)=[-1,1]$. Thus

$$\partial_B H(0)z=\{-1,1\}\;\text{ and }\;\partial_C H(0)z=[-1,1]\;\text{ for }\; z=1.$$

Since $H(x) = |x|$ is locally Lipschitzian and directionally differentiable, we have

$$DH(0)(z) = \big\{H'(0; z)\big\} = |z| = \{1\} \;\text{ for }\; z = 1.$$

Hence it gives the relationships

$$DH(0)(z) = \mathrm{co}\big\{DH(0)(z)\big\} \subset \partial_B H(0)z \subset \partial_C H(0)z,$$

where both inclusions are strict. Observe also the difference between the convexification of the graphical derivative and of the coderivative; in the latter case we have equality (2.8).

As mentioned in Section 1, there is an improvement [30] of the iterative procedure (4.10) with the replacement the generalized Jacobian therein by the B-subdifferential

$$-H(x^k) \in \partial_B H(x^k)d^k, \quad k = 0, 1, 2, \ldots. \tag{4.11}$$

Note that, along with obvious advantages of version (4.11) over the one in (4.10), in some settings it is easier to deal with the generalized Jacobian than with its B-subdifferential counterpart due to much better calculus and convenient representations for $\partial_C H(\bar{x})$ in comparison with the case of $\partial_B H(\bar{x})$, which does not even reduce to the classical subdifferential of convex analysis for simple convex functions as, e.g., $H(x) = |x|$. A remarkable common feature for both versions in (4.10) and (4.11) is the efficient semismoothness assumption imposed on the underlying mapping H to ensure its local superlinear convergence. This assumption, which unifies and labels versions (4.10) and (4.11) as the "semismooth Newton method", is replaced in our generalized Newton method by assumption (H2). Let us now recall the notion of semismoothness and compare it with (H2).

A mapping $H\colon I\!\!R^n \to I\!\!R^m$, locally Lipschitzian and directionally differentiable around $\bar{x}$, is *semismooth* at this point if the limit

$$\lim_{\substack{h \to z,\, t \downarrow 0 \\ A \in \partial_C H(\bar{x}+th)}} \big\{Ah\big\} \tag{4.12}$$

exists for all $z \in I\!\!R^n$; see [8, Definition 7.4.2]. This notion was introduced in [20] for real-valued functions and then extended in [31] to the vector mappings for the purpose of applications to a nonsmooth Newton's method. It is not hard to check [31, Proposition 2.1] that the existence of the limit in (4.12) implies the directional differentiability of H at $\bar{x}$ (but may not around this point) with

$$H'(\bar{x}; z) = \lim_{\substack{h \to z,\, t \downarrow 0 \\ A \in \partial_C H(\bar{x}+th)}} \big\{Ah\big\} \;\text{ for all }\; z \in I\!\!R^n.$$

One of the most useful properties of semismooth mappings is the following representation for them obtained in [29, Proposition 1]:

$$\|H(\bar{x} + z) - H(\bar{x}) - Az\| = o(\|z\|) \;\text{ for all }\; z \to 0 \;\text{ and }\; A \in \partial_C H(\bar{x} + z), \tag{4.13}$$

which we exploit now to relate semismoothness to our assumption (H2).

Proposition 4.8 (semismoothness implies assumption (H2)). *Let $H\colon I\!R^n \to I\!R^m$ be semismooth at $\bar{x}$. Then assumption* (H2) *is satisfied.*

Proof. Since any semismooth mapping is Lipschitz continuous on a neighborhood U of $\bar{x}$, we have by Proposition 2.2(c) that

$$\widetilde{D}H(x)(\bar{x}-x) = DH(x)(\bar{x}-x) \text{ for all } x \in U.$$

Proposition 4.6 yields therefore that

$$\widetilde{D}H(x)(\bar{x}-x) \subset \partial_C H(x)(\bar{x}-x) \text{ whenever } x \in U.$$

Given any $v \in \widetilde{D}H(x)(\bar{x}-x)$ and using the latter inclusion, find a matrix $A \in \partial_C H(x)$ such that $v = A(\bar{x}-x)$. Applying finally property (4.13) of semismooth mappings, we get

$$\|H(x) - H(\bar{x}) + v\| = \|H(x) - H(\bar{x}) - A(x-\bar{x})\| = o(\|x-\bar{x}\|) \text{ for all } x \in U,$$

which thus verifies (H2) and completes the proof of the proposition. △

Note that the previous proposition actually shows that condition (4.13) implies (H2). The next result states that the converse is also true, i.e., we have that assumption (H2) is completely equivalent to (4.13) for locally Lipschitzian mappings.

Proposition 4.9 (equivalent description of (H2)). *Let $H\colon I\!R^n \to I\!R^m$ be locally Lipschitzian around $\bar{x}$, and let assumption* (H2) *hold with some neighborhood U therein. Then*

$$\|H(x) - H(\bar{x}) - A(x-\bar{x})\| = o(\|\bar{x}-x\|) \text{ for all } x \in U \text{ and } A \in \partial_B H(x). \tag{4.14}$$

Therefore assumption (H2) *is equivalent to* (4.13)*.*

Proof. Arguing by contradiction, suppose that (4.14) is violated and find sequences $x_k \to \bar{x}$, $A_k \in \partial_B H(x_k)$ and a constant $\gamma > 0$ such that

$$\|H(x_k) - H(\bar{x}) - A_k(x_k-\bar{x})\| \ge \gamma\|\bar{x}-x_k\|, \quad k \in I\!N.$$

By the Lipschitz property of H and by construction (4.2) of the B-subdifferential there are points of differentiability $u_k \in S_H$ close to x_k with $H'(u_k)$ sufficiently close to A_k satisfying

$$\|H(u_k) - H(\bar{x}) - H'(u_k)(u_k-\bar{x})\| \ge \tfrac{\gamma}{2}\|\bar{x}-u_k\|, \quad k \in I\!N.$$

Then Proposition 2.2(c,f) gives us the representations

$$\widetilde{D}H(u_k)(\bar{x}-u_k) = DH(u_k)(\bar{x}-u_k) = -H'(u_k)(u_k-\bar{x})$$

for all $k \in I\!N$, which imply therefore that

$$\|H(u_k) - H(\bar{x}) + v\| \ge \tfrac{\gamma}{2}\|\bar{x}-u_k\| \text{ whenever } v \in \widetilde{D}H(u_k)(\bar{x}-u_k), \quad k \in I\!N.$$

This clearly contradicts assumption (H2) for k sufficiently large and thus ensures property (4.14). The equivalence between (H2) and (4.13) follows now from the implication (H2)$\Longrightarrow$(4.14) and the proof of Proposition 4.8. $\triangle$

It is well known that, for the class of locally Lipschitzian and directionally differentiable mappings, condition (4.13) is equivalent to the original definition of semismoothness; see, e.g., [8, Theorem 7.4.3]. Proposition 4.9 above establishes the equivalence of (4.13) to our major assumption (H2) provided that H is locally Lipschitzian around the reference point while it may *not* be directionally differentiable therein. In fact, it follows from Example 4.11 that assumption (H2) may hold for locally Lipschitzian functions, which are not directionally differentiable and hence not semismooth. Let us now illustrate that (H2) may also be satisfied for non-Lipschitzian mappings, in which case it is *not* equivalent to property (4.13).

Example 4.10 (assumption (H2) holds for non-Lipschitzian one-to-one mappings). Consider the mapping $H\colon I\!R^2 \to I\!R^2$ defined by

$$H(x_1,x_2) := \Big(x_2\sqrt{|x_1|+|x_2|^3},x_1\Big) \quad \text{for} \quad x_1,x_2\in I\!R. \tag{4.15}$$

It is easy to check that this mapping is one-to-one around $(0,0)$. Focusing for definiteness on the nonnegative branch of the mapping H, observe that at any point $(x_1,x_2)\in I\!R^2$ with either $x_1,x_2>0$, the classical Jacobian $JH(x_1,x_2)$ is computed by

$$JH(x_1,x_2) = \begin{bmatrix} \dfrac{x_2}{2\sqrt{x_1+x_2^3}} & \sqrt{x_1+x_2^3}+\dfrac{3x_2^3}{2\sqrt{x_1+x_2^3}} \\ 1 & 0 \end{bmatrix}.$$

Setting $x_1 = x_2^3$, we see that the first component

$$\frac{x_2}{2\sqrt{x_1+x_2^3}} = \frac{x_2}{2\sqrt{x_2^3+x_2^3}}$$

is unbounded when $x_1,x_2\downarrow 0$. This implies that the Jacobian $JH(x_1,x_2)$ is unbounded around $(\bar{x}_1,\bar{x}_2)=(0,0)$, and hence H is not locally Lipschitzian around the origin.

Let us finally verify that the underlying assumption (H2) is satisfied for the mapping H in (4.15). First assume that $x_1,x_2>0$. Then we need to check that

$$\begin{aligned} &\|H(x_1,x_2)-H(\bar{x}_1,\bar{x}_2)+JH(x_1,x_2)(-x_1,-x_2)\| \\ &= \left|x_2\sqrt{x_1+x_2^3}-\frac{x_1x_2}{2\sqrt{x_1+x_2^3}}-x_2\sqrt{x_1+x_2^3}-\frac{3x_2^4}{2\sqrt{x_1+x_2^3}}\right| \\ &= \left|\frac{x_1x_2}{2\sqrt{x_1+x_2^3}}+\frac{3x_2^4}{2\sqrt{x_1+x_2^3}}\right| = o\big(\sqrt{x_1^2+x_2^2}\big). \end{aligned}$$

The latter surely holds as $(x_1,x_2)\to(0,0)$ due to the estimates

$$\begin{aligned} \frac{x_1x_2}{2\sqrt{x_1+x_2^3}\sqrt{x_1^2+x_2^2}} &\le \frac{x_1}{\sqrt{x_1+x_2^3}} \le \sqrt{x_1}, \\ \frac{3x_2^4}{2\sqrt{x_1+x_2^3}\sqrt{x_1^2+x_2^2}} &\le \frac{3x_2^3}{2\sqrt{x_1+x_2^3}} \le 3x_2, \end{aligned}$$

which thus justify the fulfillment of assumption (H2) in this case. The other cases where $x_1 > 0, x_2 \le 0$ or $x_1 < 0, x_2 > 0$ or $x_1 < 0, x_2 \le 0$ or, finally, $x_1 = 0, x_2$ arbitrary (here H is not differentiable) can be treated in a similar way.

To complete our discussion on the major assumptions in this section, let us present an example of a locally Lipschitzian function, which satisfies assumptions (H1) and (H2) being locally one-to-one and metrically regular around the point in question while not being directionally differentiable and hence not semismooth at this point.

Example 4.11 (non-semismooth but metrically regular, Lipschitzian, and one-to-one functions satisfying (H1) and (H2)). We construct a function $H\colon [-1,1] \to I\!R$ in the following way. First set $H(\bar{x}) := 0$ at $\bar{x} = 0$. Then define H on the interval $(1/2, 1]$ staying between two lines

$$\Big(1 - \frac{1}{2}\Big)x + \frac{1}{4} \le H(x) \le x$$

in the following way: start from $(1,1)$ and let H be continuous piecewise linear when x goes from 1 to $1/2$ with the slope $1{+}1/4$ and then with the slope $1/2 - 1/4$ alternatively until x reaches $1/2$. Consider further each interval $(2^{-k}, 2^{-(k-1)}]$ for $k = 2, 3, \ldots$ and, starting from the point $\big(2^{-(k-1)}, 2^{-(k-1)}\big)$, define H to be continuous piecewise linear with the corresponding slopes of either $1 + 2^{-2k}$ or $1 - 2^{-k} - 2^{-2k}$ staying between the two lines

$$\Big(1 - \frac{1}{2^k}\Big)x + \frac{1}{2^{2k}} \le H(x) \le x. \tag{4.16}$$

Thus we have constructed H on the whole interval $[0,1]$; see Figure 1 for illustration. On the interval $[-1, 0]$, define the function H symmetrically with respect to the origin. Then it is easy to see that H in continuous on $[-1,1]$ and satisfies the following properties:

- H is clearly Lipschitz continuous around $\bar{x} = 0$.
- Since H is continuous and monotone with a positive uniform slope, it is one-to-one and metrically regular around $\bar{x}$, which directly follows, e.g., from the coderivative criterion (2.12). This ensures the fulfillment of assumption (H1) by Proposition 4.4.
- To verify assumption (H2), fix $k \in I\!N$ and $x \in (2^{-k}, 2^{-(k-1)}]$ and then pick any

  $$v \in DH(x)(\bar{x} - x) \subset \Big[1 - \frac{1}{2^k} - \frac{1}{2^{2k}}, 1 + \frac{1}{2^{2k}}\Big](\bar{x} - x).$$

  Since $\bar{x} = 0$, the latter implies that

  $$-\Big(1 + \frac{1}{2^{2k}}\Big)x \le v \le \Big(1 - \frac{1}{2^k} - \frac{1}{2^{2k}}\Big)x$$

  Thus we have by (4.16) and simple computations that

  $$|H(x) - H(\bar{x}) + v| \le \frac{1}{2^k}|x| + \frac{1}{2^{2k}} + \frac{1}{2^{2k}} = o\Big(\frac{1}{2^k}\Big) = o(|x - \bar{x}|),$$

  which shows that assumption (H2) is satisfied. In fact, it follows from above that the latter value is $O(2^{-2k}) = O(\|x - \bar{x}\|^2)$.

- Let us finally check that H is not directionally differentiable at $x_k = 2^{-k}$ for any $k \in I\!N$; therefore it is not directionally differentiable around the reference point $\bar{x} = 0$ and hence not semismooth at $\bar{x}$. Indeed, this follows directly from computing the graphical derivative by

$$DH(x_k)(1) = \Big[1 - \frac{1}{2^k}, 1\Big], \quad k \in I\!N,$$

which is not single-valued at x_k, and thus H is not directionally differentiable at x_k due to Proposition 2.2(c,d).

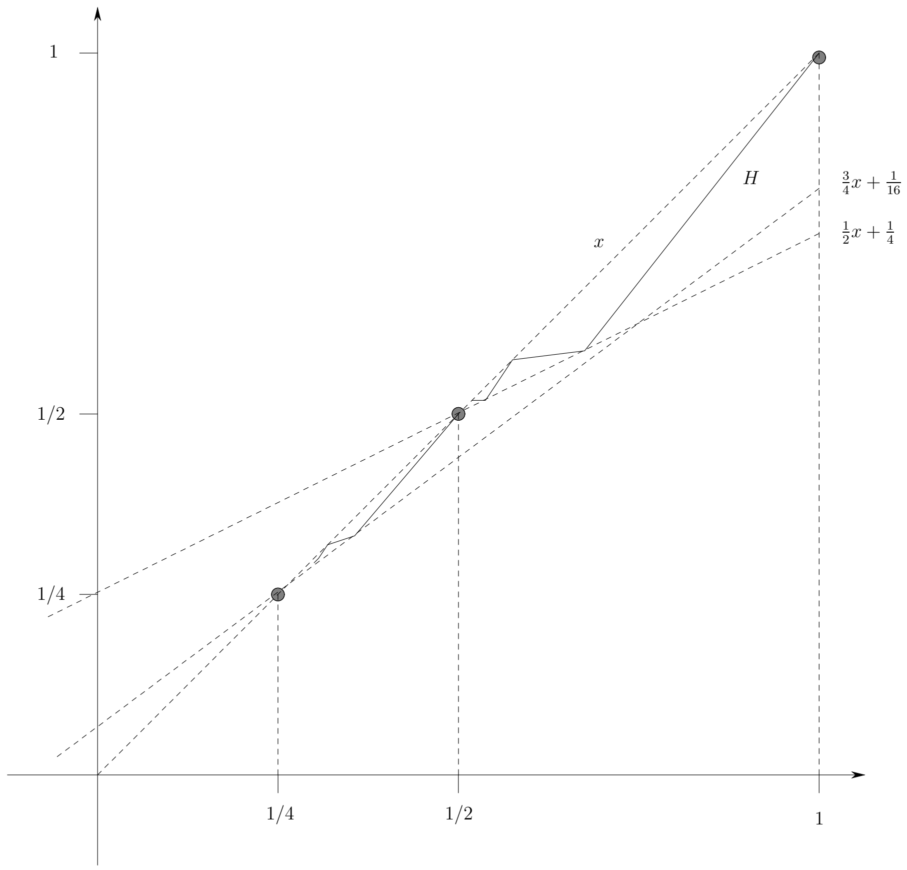

Figure 1: Construction of the mapping from Example 4.11: Illustration

5 Application to the B-differentiable Newton Method

In this section we present applications of the graphical derivate-based generalized Newton method developed above to the B-differentiable Newton method for nonsmooth equations (1.1) originated by Pang [27].

Throughout this section, suppose that $H\colon I\!R^n \to I\!R^n$ is locally Lipschitzian and directionally differentiable around the reference solution $\bar{x}$ to (1.1). Proposition 2.2(c,d) yields in this setting that the generalized Newton equation (3.1) in our Algorithm 3.1 reduces to

$$-H(x^k) = H'(x^k; d^k) \tag{5.1}$$

with respect to the new search direction d^k and that the new iterate x^{k+1} is computed by

$$x^{k+1} := x^k + d^k, \quad k = 0, 1, 2, \ldots. \tag{5.2}$$

Note that Pang's B-differentiable Newton method and its further developments (see, e.g., [8, 9, 28, 30, 31]) are based on Robinson's notion of the B(ouligand)-derivative [32] for nonsmooth mappings; hence the name. As was then shown in [37], the B-derivative of a locally Lipschitzian mapping agrees with the classical directional derivative. Thus the iteration scheme in Pang's B-differentiable method reduces to (5.1) and (5.2) in the Lipschitzian and directionally differentiable case, and so we keep the original name of [27].

The next theorem shows what we get from applying our local convergence result from Theorem 3.3 and the subsequent analysis developed in Sections 3 and 4 to the B-differentiable Newton method. This theorem employs an equivalent description of assumption (H2) held in the setting under consideration and the coderivative criterion (2.12) for metric regularity of the underlying Lipschitzian mapping H ensuring the validity of assumption (H1).

Theorem 5.1 (solvability and local convergence of the B-differentiable Newton method via metric regularity). *Let $H\colon I\!R^n \to I\!R^n$ be semismooth, one-to-one, and metrically regular around a reference solution $\bar{x}$ to* (1.1)*, i.e.,*

$$0 \in \partial\langle z, H\rangle(\bar{x}) \Longrightarrow z = 0. \tag{5.3}$$

Then the B-differentiable Newton method (5.1), (5.2) *is well defined* (*meaning that equation* (5.1) *is solvable for d^k as $k \in I\!N$*) *and converges at least superlinearly to the solution $\bar{x}$.*

Proof. Since H is locally Lipschitzian around $\bar{x}$, the coderivative criterion (2.12) is equivalently written in form (5.3) via the limiting subdifferential (2.10) due to the scalarization formula (2.9). Applying Theorem 3.3 to the B-differentiable Newton method, we need to check that assumptions (H1) and (H2) are satisfied in the setting under consideration. Indeed, it follows from Proposition 4.9 and the discussion right after it that (H2) is equivalent to the semismoothness for locally Lipschitzian and directionally differentiable mappings. The fulfillment of assumption (H1) is guaranteed by Proposition 4.4. △

More specific sufficient conditions for the well-posedness and superlinear convergence of the B-differentiable Newton method are formulated via of the Thibault derivative (4.4).

Corollary 5.2 (B-differentiable Newton method via Thibault's derivative). *Let $H\colon I\!R^n \to I\!R^n$ be semismooth at the reference solution point $\bar{x}$ of equation* (1.1)*, and let condition* (4.6) *be satisfied. Then the B-subdifferential Newton method* (5.1), (5.2) *is well defined and converges superlinearly to the solution $\bar{x}$.*

Proof. Follows from Theorem 5.1 and Proposition 4.5. △

Observe by the second inclusion in (4.5) that the assumptions of Corollary 5.2 are satisfied when all the matrices from the generalized Jacobian $\partial_C H(\bar{x})$ are nonsingular. In the latter case the solvability of subproblem (5.1) and the superlinear convergence of the B-differentiable Newton method follow from the results of [31] that in turn improve the original ones in [27], where H is assumed to be strongly Fréchet differentiable at the solution point.

Further, it is shown in [30] that the B-differentiable method for semismooth equations (1.1) converges superlinearly to the solution $\bar{x}$ if just matrices $A \in \partial_B H(\bar{x})$ are nonsingular while *assuming* in addition that subproblem (5.1) is *solvable*. As illustrated by the example presented on pp. 243–244 of [30], without the latter assumption the B-differentiable Newton method may not be well defined for semismooth mappings H on the plane with all the nonsingular matrices from $\partial_B H(\bar{x})$. We want to emphasize that the solvability assumption for (5.1) is not imposed in Theorem 5.1—it is *ensured* by *metric regularity*.

Let us now discuss interconnections between the metric regularity property of locally Lipschitzian mappings $H\colon I\!\!R^n \to I\!\!R^n$ via its coderivative characterization (5.3) and the nonsingularity of the generalized Jacobian and B-subdifferential of H at the reference point. To this end, observe the following relationships between the corresponding constructions.

Proposition 5.3 (relationships between the B-subdifferential, generalized Jacobian, and coderivative of Lipschitzian mappings). *Let $H\colon I\!\!R^n \to I\!\!R^m$ be locally Lipschitzian around $\bar{x}$. Then we have*

$$\partial_B H(\bar{x})^T z \subset \partial\langle z, H\rangle(\bar{x}) \subset \partial_C H(\bar{x})^T z \quad \text{for all} \quad z \in I\!\!R^m, \tag{5.4}$$

where both inclusions in (5.4) *are generally strict.*

Proof. Recall that the middle term in (5.4) expressed via the limiting subdifferential (2.10) is exactly the coderivative $D^*H(\bar{x})(z)$ due to the scalarization formula (2.9) for locally Lipschitzian mappings. Thus the second inclusion in (5.4) follows immediately from the well-known equality (2.8) involving convexification, and it is strict as a rule due to the usual nonconvexity of the limiting subdifferential; see [24, 35].

To justify the first inclusion in (5.4), observe that the limiting subdifferential $\partial f(\bar{x})$ of every function $f\colon I\!\!R^n \to I\!\!R$ continuous around $\bar{x}$ admits the representation

$$\partial f(\bar{x}) = \operatorname*{Lim\,sup}_{x \to \bar{x}} \widehat{\partial} f(x) \tag{5.5}$$

via the outer limit (1.6) of the Fréchet/regular subdifferentials

$$\widehat{\partial} f(x) := \Big\{ p \in I\!\!R^n \Big| \liminf_{u \to x} \frac{f(u) - f(x) - \langle p, u - x\rangle}{\|u - x\|} \ge 0 \Big\} \tag{5.6}$$

of f at x; see, e.g., [24, Theorem 1.89]. We obviously have from (5.6) that $\widehat{\partial} f(\bar{x}) = \{f'(\bar{x})\}$ if f is (Fréchet) differentiable at $\bar{x}$ with its derivative/gradient $f'(\bar{x})$.

Having the mapping $H = (h_1, \ldots, h_m)\colon I\!\!R^n \to I\!\!R^m$ in the proposition and fixing an arbitrary vector $\bar{z} = (\bar{z}_1, \ldots, \bar{z}_m) \in I\!\!R^m$, form now a scalar function $f_{\bar{z}}\colon I\!\!R^n \to I\!\!R$ by

$$f_{\bar{z}}(x) := \sum_{i=1}^m \bar{z}_i h_i(x), \quad x \in I\!\!R^n. \tag{5.7}$$

Then the first inclusion in (5.4) amounts to say that

$$\partial_B H(\bar{x})^T \bar{z} \subset \partial f_{\bar{z}}(\bar{x}). \tag{5.8}$$

To proceed with proving (5.8), pick any matrix $A \in \partial_B H(\bar{x})^T \bar{z}$ and denote by $a_i \in I\!R^n$, $i = 1, \ldots, n$, its vector rows. By definition (4.2) of the B-subdifferential $\partial_B H(\bar{x})$ there is a sequence $\{x^k\} \subset S_H$ from the set of differentiability (4.1) such that $x^k \to \bar{x}$ and $H'(x^k) \to A$ as $k \to \infty$. It is clear from (5.7) that the function $f_{\bar{z}}$ is differentiable at each x^k with

$$f'_{\bar{z}}(x^k) = \sum_{i=1}^{m} \bar{z}_i h'_i(x^k) \to \sum_{i=1}^{m} \bar{z}_i a_i = A^T \bar{z} \ \text{ as } \ k \to \infty.$$

Since $\widehat{\partial} f_{\bar{z}}(x^k) = \{f'_{\bar{z}}(x^k)\}$ at all the points of differentiability, we arrive at (5.8) by representation (5.5) of the limiting subdifferential and thus justify the first inclusion in (5.4).

To illustrate that the latter inclusion may be strict, consider the function $H(x) := |x|$ on $I\!R$. Then $\partial_B H(0)z = \{-z, z\}$ for all $z \in I\!R$, while

$$\partial(zH)(0) = D^* H(0)(z) = \begin{cases} [-z, z] & \text{for } \ z \geq 0, \\ \{-z, z\} & \text{for } \ z < 0. \end{cases}$$

This completes the proof of the proposition. △

It follows from Proposition 5.3 in the case of Lipschitzian transformations $H\colon I\!R^n \to I\!R^n$ that the nonsingularity of all the matrices $A \in \partial_C H(\bar{x})$ is a sufficient condition for the metric regularity of H around $\bar{x}$ due to the coderivative criterion (5.3) while the nonsingularity of all $A \in \partial_B H(\bar{x})$ is a necessary condition for this property. Note however, as it has been discussed above, that the nonsingularity condition for $\partial_B H(\bar{x})$ alone does not ensure the solvability of subproblem (5.1) in the B-differentiable Newton method, and thus it cannot be used alone for the justification of algorithm (5.1), (5.2) in the B-differentiable semismooth case. Furthermore, we are not familiar with any verifiable condition to support the nonsingularity of $\partial_B H(\bar{x})$ in the full justification of the B-differentiable Newton method.

In contrast to this, the metric regularity itself—via its verifiable pointwise characterization (5.3)—ensures the solvability of (5.1) and fully justifies the B-differentiable Newton method with its superlinear convergence provided that the mapping H is semismooth and locally invertible around the reference solution point. Note that the nonsingularity of the generalized Jacobian $\partial_C H(\bar{x})$ implies not only the metric regularity but simultaneously the semismoothness and local invertibility of a Lipschitzian transformation $H\colon I\!R^n \to I\!R^n$. However, the latter condition fails to spot a number of important situations when all the assumptions of Theorem 5.1 are satisfied; see, in particular, Corollary 5.2 and the corresponding conditions in terms of Warga's derivate containers discussed right after Corollary 4.5. We refer the reader to the specific mappings $H\colon I\!R^2 \to I\!R^2$ from [17, Example 2.2] and [40, Example 3.3] that can be used to illustrate the above statement.

6 Concluding Remarks

In this paper we develop a new generalized Newton method for solving systems of nonsmooth equations $H(x) = 0$ with $H\colon I\!R^n \to I\!R^n$. Local superlinear convergence and global (of the Kantorovich type) convergence results are derived under relatively mild conditions. In particular, the local Lipschitz continuity and directional differentiability of H are not necessarily required. We show that the new method and its specifications have some advantages in comparison with previously known results on the semismooth and B-differentiable versions of the generalized Newton method for nonsmooth Lipschitz equations.

Our approach is heavily based on advanced tools of variational analysis and generalized differentiation. The algorithm itself is built by using the graphical/contingent derivative of H, while other graphical derivatives and coderivatives are employed in formulating appropriate assumptions and proving solvability and convergence results. The fundamental property of metric regularity and its pointwise coderivative characterization play a crucial role in the justification of the algorithm and its satisfactory performance.

In the other lines of developments, it seems appealing to develop an alternative Newton-type algorithm, which is constructed by using the basic coderivative instead of the graphical derivative. This requires certain symmetry assumptions for the given problem, since the coderivative is an extension of the adjoint derivative operator. Major advantages of a coderivative-based Newton method would be comprehensive calculus rules held for the coderivative in contrast to the contingent derivative, complete coderivative characterizations of Lipschitzian stability, and explicit calculations of the coderivative in a number of settings important for applications. The details of these ideas are part of our future research.

References

[1] F. J. Aragón Artacho and M. H. Goeffroy, *Uniformity and inexact version of a proximal method for metrically regular mappings*, J. Math. Anal. Appl., 135 (2007), pp. 168–183.

[2] J.-P. Aubin, *Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions*, in Mathematical Analysis and Applications, L. Nachbin, ed., Academic Press, New York, 1981, pp. 159–229.

[3] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser, Boston, 1990.

[4] J. V. Burke, A. S. Lewis and M. L. Overton, *Variational analysis of functions of the roots of polynomials*, Math. Program., 104 (2005), pp. 263–292.

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

[6] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Program., 75 (1996), pp. 407–439.

[7] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ, 1983.

[8] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalitites and Complementarity Problems, Volumes I and II*, Springer, New York, 2003.

[9] S. P. Han, J.-S. Pang and N. Rangaraj, *Globally convergent Newton methods for nonsmooth equations*, Math. Oper. Res., 17 (1992), pp. 586–607.

[10] A. D. Ioffe, *Metric regularity and subdifferential calculus*, Russian Math. Surveys, 55 (2000), pp. 501–558.

[11] N. H. Josephy, *Newton's method for generalized equations and the PIES energy model*, Ph.D. Dissertation, Department of Industrial Engineering, Univesity of Wisconsin–Madison, 1979.

[12] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Macmillan, New York, NY, 1964.

[13] C. Kanzow, I. Ferenczi and M. Fukushima, *On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity*, SIAM J. Optim., 20 (2009), pp. 297–320.

[14] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*, SIAM, Philadelphia, 2003.

[15] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications*, Kluwer, Dordrecht, The Netherlands, 2002.

[16] B. Kummer, *Newton's method fort non-differentiable functions*, in Advances in Mathematical Optimization, J. Guddat et al., eds., Akademi-Verlag, Berlin, 1988, pp. 114–125.

[17] B. Kummer, *Lipschitzian inverse functions, directional derivatives, and applications in $C^{1,1}$ optimization*, J. Optim. Theory Appl., 70 (1991), 561–582.

[18] A. S. Lewis, D. R. Luke and J. Malick, *Local linear convergence for alternative and averaged nonconvex projections*, Found. Comp. Math., 9 (2009), pp. 485–513.

[19] Z.-Q. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.

[20] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 957–972.

[21] B. S. Mordukhovich, *Maximum principle in problems of time optimal control with nonsmooth constraints*, J. Appl. Math. Mech., 40 (1976), pp. 960–960.

[22] B. S. Mordukhovich, *Metric approximations and necessary optimality conditions for general classes of extremal problems*, Soviet Math. Dokl., 22 (1980), pp. 526–530.

[23] B. S. Mordukhovich, Complete characterizations of openness, metric regularity, and Lipschitzian properties of multifunctions, Trans. Amer. Math. Soc., 340 (1993), pp. 1–35.

[24] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II; Applications*, Springer, Berlin, 2006.

[25] B. S. Mordukhovich, J. Peña and V. Roshchina, *Applying metric regularity to compute a condition measure of smoothing algorithm for matrix games*, http://arxiv.org/abs/1007.4458, to appear in SIAM J. Optim.

[26] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[27] J.-S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 311–341.

[28] J.-S. Pang, *A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Program., 51 (1991), pp. 101–131.

[29] J.-S. Pang and L. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.

[30] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

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

[32] S. M. Robinson, *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*, Math. Program. Studies, 30 (1987), pp. 45–66.

[33] S. M. Robinson, *An implicit function theorem for a class of nonsmooth functions*, Math. Oper. Res., 16 (1991), pp. 292–309.

[34] S. M. Robinson, *Newton's method for a class of nonsmooth functions*, Set-Valued Anal., 2 (1994), pp. 291–305.

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

[36] W. Schirotzek, *Nonsmooth Analysis*, Springer, Berlin, 2007.

[37] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl., 66 (1990), pp. 477–487.

[38] N. Shor, *On a class of almost-differentiable functions and on a minimization method for functions from this class*, Kibernetika, No. 4 (1972), pp. 65–70.

[39] L. Thibault, *Subdifferentials of compactly Lipschitz vector-valued functions*, Ann. Matem. Pure Appl., 4 (1980), pp. 157–192.

[40] J. Warga, *Fat homeomorphisms and unbounded derivate containers*, J. Math. Anal. Appl., 81 (1981), 545–560.