Deutsche Forschungsgemeinschaft
Priority Programme 1962

Newton Differentiability of Convex Functions in Normed Spaces and of a Class of Operators

Martin Brokate, Michael Ulbrich

SPP 1962 | Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-176

received on September 29, 2021

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

NEWTON DIFFERENTIABILITY OF CONVEX FUNCTIONS IN NORMED SPACES AND OF A CLASS OF OPERATORS

MARTIN BROKATE[*] AND MICHAEL ULBRICH[†]

Abstract. Newton differentiability is an important concept for analyzing generalized Newton methods for nonsmooth equations. In this work, for a convex function defined on an infinite-dimensional space, we discuss the relation between Newton and Bouligand differentiability and upper semicontinuity of its subdifferential. We also construct a Newton derivative of an operator of the form $(Fx)(p) = f(x,p)$ for general nonlinear operators f that possess a Newton derivative with respect to x and also for the case where f is convex in x.

AMS subject classifications. 49J52, 46G05, 47H04, 49M15

Key words. convex, subdifferential, semismooth, Newton derivative, Bouligand derivative, maximum functional, measurable selector

1. Introduction.

A standard approach to iteratively solve an equation

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

in a normed space is given by Newton's method. It is locally superlinearly or even quadratically convergent under suitable assumptions, namely, in roughly terms, sufficient smoothness of F and bounded invertibility of its derivative. It is known for more than three decades that, beyond the classical, smooth setting, superlinear convergence can still be achieved by a generalized version of Newton's method, the semismooth Newton method. It requires that F possesses a certain generalized derivative, the Newton derivative[1]. If F is even α-order Newton differentiable, $\alpha > 0$, then convergence with order $1 + \alpha$ can be achieved. Naturally, this close connection to Newton's method shows that Newton differentiability is an important and desirable property of nonlinear mappings. As will be reviewed below, the class of mappings currently known to be Newton differentiable is much larger in finite dimensions than in infinite dimensions.

In this paper, we make contributions to the infinite-dimensional branch of this theory. On the one hand, the Newton differentiability of convex functions in infinite-dimensional spaces and its relation to Bouligand differentiability and to the norm upper semicontinuity of their subdifferential are investigated. Secondly, we develop Newton differentiability results for a general class of operators of the form

$$(Fx)(p) = f(x,p)\,, \tag{1.2}$$

where $f(\cdot,p)$, $p \in P$, is a family of Newton differentiable operators. This work was initially motivation by the goal of finding a framework that, in particular, provides Newton differentiability results for functionals like the (convex) maximum functional $x \in X \subset C(K) \mapsto f(x) = \max_{s\in K} x(s) \in \mathbb{R}$, where K is a compact space and, in a next step, of the cumulated maximum operator $x \in X \subset C[a,b] \mapsto \max_{a\le s\le p} x(s) \in$

[*]Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85747 Garching, Germany; Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany; Faculty of Civil Engineering, Czech Technical University in Prague, Thákurova 7, 16629 Praha 6, Czech Republic (brokate@ma.tum.de).

[†]Chair of Mathematical Optimization, Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85747 Garching, Germany (mulbrich@ma.tum.de).

[1]This terminology is not used in a uniform way.

$C[a,b]$, which is of interest in the study of rate-independent systems. The cumulated maximum is a particular instance of an operator F as in (1.2), where f is chosen as the parametric maximum functional $f(x,p) = \max_{a\le s\le p} x(s)$, $p \in [a,b]$.

We now give a short summary of known results. For the theory on Newton differentiability, semismoothness, and generalized Newton methods in general spaces, we refer to [6, 12, 16, 20, 21, 29, 30, 31]. The literature in finite dimensions is comprehensive; we only refer to some early work [10, 23, 26, 27] and to [10] and the references therein. In order to apply this theory, it is therefore of interest to know whether a given operator F has a Newton derivative. As often F is constructed from simpler functions, this question naturally extends to those as well.

In finite dimensions, the term semismoothness is more common than Newton differentiability. A function between finite-dimensional normed spaces is called semismooth at a point if it is locally Lipschitz, directionally differentiable, and Newton differentiable at this point and the Newton derivative is given by Clarke's generalized Jacobian. In this setting, a large variety of functions has been shown to be semismooth, including piecewise C^1 functions [31], eigenvalues of symmetric matrices, singular values, projections onto the semidefinite and other cones [32], spectral functions [28] and spectral operators [8] of matrices induced by semismooth functions, and, very generally, tame functions [2]. This together with the fact that sums, composition, and vectorizations of semismooth functions are semismooth, allows for a comprehensive calculus. In infinite-dimensional spaces, the theory is significantly more difficult and much less is available. A particularly useful result is that superposition (i.e., Nemytskii) operators $F : u \in L^p(D) \to L^q(D)$, $F(u)(\omega) = f(u(\omega))$, $\omega \in D$, are Newton differentiable with respect to a canonical choice for the Newton derivative if $D \subset \mathbb{R}^d$ is bounded, $1 \le q < p \le \infty$, and $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and semismooth (some generalizations are possible) [12, 30, 29, 31]. Also, α-order Newton differentiability results for F, where the remainder term is $O(\|\cdot\|^{1+\alpha})$ instead of $o(\|\cdot\|)$, can be developed [12, 30, 29, 31].

As already outlined above, this paper studies the Newton differentiability of convex functionals on infinite-dimensional spaces, the relation to Bouligand differentiability and to norm upper semicontinuity properties of the subdifferential. Further, it develops Newton differentiability results for operators of the form (1.2). The rest of the paper is organized as follows:

In Section 2 we state some basic definitions and present a mean value inequality for Newton-differentiable functions.

Section 3 is concerned with the relation between Newton and Bouligand differentiability of convex functions and norm upper semicontinuity of their subdifferentials. Moreover, we show that Bouligand and Newton differentiability can be induced by a suitable change of the norms of the underlying spaces and use this to improve the estimate for the remainder term.

In Section 4 we investigate operators of the form (1.2). We show how to obtain Newton and Bouligand derivatives of F from corresponding derivatives of $f(\cdot,p)$ where f is a nonlinear vector-valued function. Then we apply this to the special case of parametric convex functions, that is, $f(\cdot,p)$ is convex for any given parameter value p.

In Section 5, the appendix, we list some results of set-valued analysis.

All vector spaces considered in this paper are spaces over the real numbers.

2. Newton and Bouligand derivatives. Let

$$F : U \to Y\,, \quad U \subset X\,,$$

where X and Y are normed spaces, and U is an open subset of X. By $\mathcal{L}(X, Y)$ we denote the Banach space of all linear continuous mappings from X to Y.

We deal with set-valued derivatives of F. For set-valued mappings Ψ from U to Y we write "$\Psi : U \rightrightarrows Y$" instead of "$\Psi : U \to \mathcal{P}(Y) \setminus \emptyset$".

Definition 2.1. *A mapping $G : U \rightrightarrows \mathcal{L}(X, Y)$ is called a* ***Newton derivative*** *of F at $x \in U$, if*

$$\lim_{h\to 0} \sup_{L\in G(x+h)} \frac{\|F(x+h) - F(x) - Lh\|}{\|h\|} = 0\,. \tag{2.1}$$

It is called a ***Newton derivative*** *of F in U, if (2.1) holds at every $x \in U$. If*

$$\sup_{L\in G(x+h)} \|F(x+h) - F(x) - Lh\| = O(\|h\|^{1+\alpha}) \tag{2.2}$$

with $\alpha > 0$, then G is said to be of ***order*** $\boldsymbol{\alpha}$*.*

The mapping F is called ***semismooth*** *at x (in U, resp.), if it is locally Lipschitz and there exists a Newton derivative of F at x (in U, resp.).*

It is well known that if F is continuously Fréchet differentiable in U, then $G(x) = \{DF(x)\}$ is a single-valued Newton derivative of F in U.

We may write (2.1) in remainder form,

$$\sup_{L\in G(x+h)} \|F(x+h) - F(x) - Lh\| \le \rho_x(\|h\|) \cdot \|h\|\,, \tag{2.3}$$

where $\rho_x(\delta) \downarrow 0$ as $\delta \downarrow 0$. Passing to $\tilde\rho_x(\delta) = \sup_{0<\eta\le\delta} \rho_x(\eta)$ if necessary we may always assume that ρ_x is nondecreasing.

If ℓ is a global Lipschitz constant for F and c_G is a global bound for the norms $\|L\|$ of the elements $L \in G(U)$, it directly follows from (2.3) that we may choose ρ_x to be globally bounded,

$$\rho_x \le \ell + c_G \qquad \forall\, x \in U\,. \tag{2.4}$$

If $G : U \rightrightarrows \mathcal{L}(X, Y)$ is a Newton derivative of F in U, then so is every $\tilde G : U \rightrightarrows \mathcal{L}(X, Y)$ satisfying $\tilde G(x) \subset G(x)$ for all $x \in U$. In particular, every **selector** $S : U \to L(X, Y)$ of G, that is, $S(x) \in G(x)$ for all $x \in U$, yields a single-valued Newton derivative of F in U.

The **directional derivative** of F at $x \in U$ in the direction $h \in X$ we denote by

$$F'(x; h) = \lim_{\lambda\downarrow 0} \frac{F(x+\lambda h) - F(x)}{\lambda}\,. \tag{2.5}$$

If the limit exists for all $h \in X$, F is called **directionally differentiable** at x.

Definition 2.2. *The function F is called* ***Bouligand differentiable*** *at $x \in U$, if it is directionally differentiable at x and if*

$$\lim_{h\to 0} \frac{\|F(x+h) - F(x) - F'(x; h)\|}{\|h\|} = 0\,. \tag{2.6}$$

Newton-differentiable functions satisfy a mean value inequality.

PROPOSITION 2.3. *Let $G : U \rightrightarrows \mathcal{L}(X, Y)$ be a Newton derivative of $F : U \to Y$, let*

$$L = \sup_{x\in U} \sup_{M\in G(x)} \|M\| \,. \tag{2.7}$$

Then

$$\|F(x') - F(x)\| \le L\|x' - x\| \qquad \text{for all } x', x \in U. \tag{2.8}$$

Proof. Let $x', x \in U$ with $x' \neq x$, let $L' = \|F(x') - F(x)\|/\|x' - x\|$. It suffices to show that $L' \le L$.

For arbitrary $z \in [x, x'] = \{(1 - t)x + tx' : t \in [0, 1]\}$ we have

$$\|F(z) - F(x)\| \ge L'\|z - x\| \quad \text{or} \quad \|F(z) - F(x')\| \ge L'\|z - x'\| \,, \tag{2.9}$$

since

$$\begin{aligned} L'\|z - x\| + L'\|z - x'\| = L'\|x' - x\| &\le \|F(x') - F(x)\| \\ &\le \|F(z) - F(x)\| + \|F(z) - F(x')\| \,. \end{aligned}$$

Starting from $[x_0, x_0'] = [x, x']$ we construct a nested sequence of intervals $I_k = [x_k, x_k']$ as follows. We divide I_k by its midpoint $z = (x_k + x_k')/2$ and choose $I_{k+1} = [x_k, z]$ or $I_{k+1} = [z, x_k']$ so that $\|F(x_{k+1}) - F(x_{k+1}')\| \ge L'\|x_{k+1} - x_{k+1}'\|$. This is possible in view of (2.9), applied inductively with $[x_k, x_k']$ in place of $[x, x']$. Moreover, $\|x_k - x_k'\| = 2^{-k}\|x - x'\| \to 0$.

Let $\cap_{k\in\mathbb{N}} I_k = \{x\}$. Again by the argument leading to (2.9), now applied with $z = x$, we may choose $y_k \in \{x_k, x_k'\}$ such that

$$y_k \neq x \,, \quad \|F(y_k) - F(x)\| \ge L'\|y_k - x\| \qquad \text{for all } k \in \mathbb{N}.$$

It follows that

$$L'\|y_k - x\| \le \|F(y_k) - F(x) - M(y_k - x)\| + \|M(y_k - x)\| \tag{2.10}$$

holds for all k and all $M \in G(y_k)$. As G is a Newton derivative of F, taking the supremum over all such M yields

$$L'\|y_k - x\| \le o(\|y_k - x\|) + L\|y_k - x\| \,.$$

Passing to the limit $k \to \infty$ we arrive at $L' \le L$. ∎

3. Semismoothness of convex functions.

Let

$$f : U \to \mathbb{R} \,, \quad U \subset X \,,$$

be convex and continuous, where X is a Banach space and U is an open convex subset of X. As usual, we extend f to X by setting $f(x) = +\infty$ for $x \notin U$.

We first collect some well-known properties of convex functions.

PROPOSITION 3.1. *In the situation above,*
(i) f is locally Lipschitz continuous on U.
(ii) The directional derivative $f'(x; h)$ exists for all $x \in U$, $h \in X$, and

$$f'(x; h) \le f(x + h) - f(x) \,. \tag{3.1}$$

(iii) The mapping $f'(x; \cdot) : X \to \mathbb{R}$ is sublinear and continuous (hence, locally Lipschitz continuous by (i)) at every $x \in U$. In particular, for all $x \in U$ and all $h \in X$

$$f'(x; h) \ge -f'(x; -h) \,. \tag{3.2}$$

Proof. See [25], Proposition 1.6 for (i), Lemma 1.2 and Corollary 1.7 for (ii) and (iii). ◻

Let

$$\partial f(x) = \{x^* \in X^* : f(y) - f(x) \geq \langle x^*, y - x\rangle \text{ for all } y \in X\} \tag{3.3}$$

be the subdifferential of f at $x \in U$.

PROPOSITION 3.2. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X. Then*
(i) $\partial f(x)$ is nonempty, convex and w^*-compact for all $x \in U$.
(ii) $\partial f : U \rightrightarrows X^*$ is locally bounded. Every local Lipschitz constant for f at x is a bound for ∂f near x.
(iii) $\mathrm{graph}(\partial f)$ *is closed in $X \times (X^*, w^*)$.*
(iv) $\partial f : U \rightrightarrows (X^*, w^*)$ is upper semicontinuous (usc).
(v) For every $x \in U$ and $h \in X$,

$$f'(x; h) = \max_{x^* \in \partial f(x)} \langle x^*, h\rangle\,. \tag{3.4}$$

Proof. See [25], Proposition 1.11 for (i) and (ii), Proposition 7.3 for (iii), Proposition 2.5 for (iv). For (v), see [25, Proposition 2.24] or [3, Proposition 2.125]. ◻

Bouligand and Newton differentiability of convex functions. When X has finite dimension, every continuous convex function on an open subset U of X is Bouligand differentiable on U (see the proof of Proposition 3.1.3 in [10]), and ∂f is a Newton derivative of f in U (see Proposition 7.4.5 in [10] or Proposition 3.7 below). For infinite-dimensional X this is no longer the case, not even in Hilbert space, see the examples below.

Given $x \in U$, let us define the **Bouligand remainder function**

$$r^B(h) = f(x + h) - f(x) - f'(x; h)\,, \quad h \in U - x\,. \tag{3.5}$$

We have $r^B \geq 0$ by (3.1). By definition, f is Bouligand differentiable at x if and only if $r^B(h) = o(\|h\|)$. We also define the **Newton remainder function**

$$r^N(h) = \sup_{y^* \in \partial f(x+h)} |f(x + h) - f(x) - \langle y^*, h\rangle|\,. \tag{3.6}$$

Then ∂f is a Newton derivative of f in x if and only if $r^N(h) = o(\|h\|)$.

LEMMA 3.3. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X, let $x \in U$. Then for all $h \in U - x$*

$$r^N(h) = f'(x + h; h) - f(x + h) + f(x) \geq 0\,. \tag{3.7}$$

Proof. Let $h \in U - x$. Since

$$f(x) - f(x + h) \geq \langle y^*, -h\rangle \qquad \text{for all } y^* \in \partial f(x + h),$$

we have

$$\begin{aligned} 0 \leq r^N(h) &= \sup_{y^* \in \partial f(x+h)} (\langle y^*, h\rangle - f(x + h) + f(x)) \\ &= f'(x + h; h) - f(x + h) + f(x) \end{aligned}$$

by Proposition 3.2(v). ◻

LEMMA 3.4. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X, let $x \in U$. Then for all $h \in U - x$*

$$r^N(h) \le r^B(2h) - 2r^B(h) \le r^N(2h)\,. \tag{3.8}$$

Proof. We have

$$\begin{aligned} r^B(2h) - 2r^B(h) &= (f(x+2h) - f(x) - f'(x;2h)) - 2(f(x+h) - f(x) - f'(x;h)) \\ &= f(x+2h) - 2f(x+h) + f(x)\,. \end{aligned} \tag{3.9}$$

It follows from (3.1) and (3.7) that

$$f(x+2h) \ge f(x+h) + f'(x+h;h) = r^N(h) + 2f(x+h) - f(x)\,.$$

In view of (3.9) the left inequality in (3.8) follows. In order to prove the right inequality we observe that, due to (3.7) and (3.2),

$$\begin{aligned} r^N(2h) &= f'(x+2h;2h) - f(x+2h) + f(x) \\ &\ge -2f'(x+2h;-h) - f(x+2h) + f(x)\,. \end{aligned}$$

It follows from (3.9) and (3.1) that

$$\begin{aligned} r^N(2h) - (r^B(2h) - 2r^B(h)) &\ge 2(f(x+h) - f(x+2h) - f'(x+2h;-h)) \\ &\ge 0\,. \end{aligned}$$

∎

REMARK 3.5. If f is smooth and $c > 0$, then

$$\begin{aligned} r^B(ch) &= \int_0^c \int_0^t f''(x+sh)(h,h)\,ds\,dt\,, \\ r^N(ch) &= \int_0^c \int_t^c f''(x+sh)(h,h)\,ds\,dt\,. \end{aligned}$$

As the function $s \mapsto f''(x+sh)(h,h)$ is nonnegative, the inequalities in (3.8) alternatively can be derived from inclusions of certain corresponding subsets of $[0,2] \times [0,2]$.

Example 3.11 below shows that there exist continuous convex functions f for which an estimate $r^B(h)/\|h\| \le \eta(r^N(h)/\|h\|)$ for some function η with $\eta(\delta) \to 0$ as $\delta \to 0$ is not possible. Nevertheless one has the following result.

LEMMA 3.6. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X, let $x \in U$. Then for all $h \in U - x$*

$$r^B(h) \le \sum_{k=0}^{\infty} 2^k r^N(2^{-k}h)\,. \tag{3.10}$$

Proof. Let $h \in U - x$. We define

$$a_k(h) = 2^k(r^B(2^{-k}h) - 2r^B(2^{-(k+1)}h))\,, \quad s_m(h) = \sum_{k=0}^{m} a_k(h)\,.$$

Then $a_k(h) \ge 0$ by (3.8). We have

$$0 \le r^B(h) - s_m(h) = 2^{m+1} r^B(2^{-(m+1)}h) \to 0 \quad \text{as } m \to \infty, \tag{3.11}$$

since $r^B(th) = o(t)$ as f is directionally differentiable. From (3.8) it now follows that

$$r^B(h) = \sum_{k=0}^{\infty} a_k(h) \le \sum_{k=0}^{\infty} 2^k r^N(2^{-k}h)\,. \qquad \square$$

PROPOSITION 3.7. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X, let $x \in U$.*
(i) If f is Bouligand differentiable at x, then ∂f is a Newton derivative of f at x.
(ii) If ∂f is a Newton derivative of order $\alpha > 0$ of f at x, then f is Bouligand differentiable at x with

$$f(x+h) - f(x) - f'(x;h) = O(\|h\|^{1+\alpha})\,. \tag{3.12}$$

Proof. Part (i) follows directly from Lemma 3.4, as $r^B(h) = o(\|h\|)$ implies $r^N(h) = o(\|h\|)$ by (3.8). Concerning (ii), let $r^N(h) \le c\|h\|^{1+\alpha}$ for some $c > 0$. From Lemma 3.6 we obtain that

$$r^B(h) \le \sum_{k=0}^{\infty} 2^k r^N(2^{-k}h) \le \sum_{k=0}^{\infty} 2^k c\|2^{-k}h\|^{1+\alpha} = c\|h\|^{1+\alpha} \sum_{k=0}^{\infty} 2^{-\alpha k}$$
$$= c\|h\|^{1+\alpha} \frac{1}{1-2^{-\alpha}}\,. \qquad \square$$

In Example 3.11 we present a function where $r^N(h) = O(\|h\|\,|\ln(\|h\|)|^{-1/2})$ at 0, but which is not Bouligand differentiable at 0.

Examples. We provide several examples of continuous convex functions whose subdifferential is not a Newton derivative as well as an example of a continuous convex function which is not Bouligand differentiable at 0 but whose subdifferential is a Newton derivative at 0.

EXAMPLE 3.8. Let $X = U = \ell^p$, $1 \le p \le \infty$. (This includes the Hilbert space case $p = 2$.) We consider

$$f : \ell^p \to \mathbb{R}\,, \quad f(x) = \sup_{k\in\mathbb{N}} f_k(x_k)\,, \quad f_k(t) = \max\left\{t - \frac{1}{k}, 0\right\}.$$

Being the supremum of convex functions, f is convex. It is well-defined since $0 \le f(x) \le \|x\|_\infty \le \|x\|_p$ for all $x \in \ell^p$. Moreover, f is Lipschitz continuous: For arbitrary real-valued sequences (a_k) and (b_k) we have

$$\sup_{k\in\mathbb{N}} a_k \le \sup_{k\in\mathbb{N}}(a_k - b_k) + \sup_{k\in\mathbb{N}} b_k$$

and therefore

$$\big|\sup_{k\in\mathbb{N}} a_k - \sup_{k\in\mathbb{N}} b_k\big| \le \sup_{k\in\mathbb{N}} |a_k - b_k|\,.$$

It follows that for all $x, y \in \ell^p$

$$|f(x) - f(y)| \le \sup_{k\in\mathbb{N}} |f_k(x_k) - f_k(y_k)| \le \sup_{k\in\mathbb{N}} |x_k - y_k| \le \|x - y\|_p\,. \tag{3.13}$$

Let e_n be the n-th unit vector, let $t \ge 0$. Then $f(te_n) = f_n(t)$, $f'(te_n; e_n) = f_n'(t+)$. The remainder functions satisfy, see (3.7) and (3.5),

$$r^N(te_n) = f'(te_n; te_n) - f(te_n) + f(0) = tf_n'(t+) - f_n(t)$$
$$r^B(te_n) = f(te_n) - f(0) - f'(0; te_n) = f_n(t)\,.$$

For $h^n = (2/n)e_n$ we obtain

$$r^N(h^n) = \frac{2}{n} \cdot 1 - \frac{1}{n} = \frac{1}{2}\|h^n\|_p\,, \quad r^B(h^n) = \frac{1}{n} = \frac{1}{2}\|h^n\|_p\,.$$

Since $h^n \to 0$ as $n \to \infty$, ∂f is not a Newton derivative of f at 0 (and f is not Bouligand differentiable at 0).

EXAMPLE 3.9. Let $U = X = C(K)$, where K is a compact space. Let $f : X \to \mathbb{R}$ be the maximum functional

$$f(x) = \max_{s\in K} x(s)\,. \tag{3.14}$$

The dual space $C(K)^*$ consists of all signed regular Borel measures on K. It is not difficult to check that (see also Theorem 5.3.39 in [7] or (2.236) in [3])

$$\partial f(x) = \{\mu : \mu \in C(K)^*,\ \mathrm{supp}(\mu) \subset M(x),\ \mu \geq 0,\ \|\mu\| = 1\} \tag{3.15}$$

where

$$M(x) = \{t \in K : x(t) = f(x)\} \tag{3.16}$$

denotes the set on which x attains its maximum. It follows from Proposition 3.2(v), or it can be proved directly, that

$$f'(x;h) = \max_{t\in M(x)} h(t)\,. \tag{3.17}$$

If x is a constant function then f is Bouligand differentiable at x since $M(x) = K$ and $f'(x;h) = f(h) = f(x+h) - f(x)$, therefore $r^B(h) = 0$. We claim that ∂f is not a Newton derivative of f at x if x is not constant. Let $t \in M(x)$, set $x_M = x(t) = \max_K x$. We define $h_\lambda \in X$ by

$$h_\lambda(s) = 2\min\{x_M - x(s), \lambda\}\,, \quad \lambda > 0\,. \tag{3.18}$$

Then $0 \leq h_\lambda \leq 2\lambda$ since $t \in M(x)$. Moreover, for every $s \in K$ we have

$$\begin{aligned} (x + h_\lambda)(s) &= \min\{2x_M - x(s), 2\lambda + x(s)\} \\ &= x_M + \lambda + \min\{x_M - x(s) - \lambda, -x_M + x(s) + \lambda\} \\ &= x_M + \lambda - |x_M - x(s) - \lambda|\,. \end{aligned} \tag{3.19}$$

Assuming λ small enough, there exists $t_\lambda \in K$ with $x(t_\lambda) = x_M - \lambda$. For such a t_λ,

$$h_\lambda(t_\lambda) = 2\lambda = \|h_\lambda\|_\infty\,, \quad f(x + h_\lambda) = x_M + \lambda = (x + h_\lambda)(t_\lambda)\,. \tag{3.20}$$

Therefore, $t_\lambda \in M(x + h_\lambda)$, and $\mu = \delta_{t_\lambda}$ (the Dirac measure at t_λ) belongs to $\partial f(x + h_\lambda)$. It follows that

$$\frac{|f(x + h_\lambda) - f(x) - \langle \mu, h_\lambda\rangle|}{\|h_\lambda\|_\infty} = \frac{|x_M + \lambda - x_M - 2\lambda|}{2\lambda} = \frac{1}{2}\,. \tag{3.21}$$

As $h_\lambda \to 0$ for $\lambda \to 0$, we see that ∂f is not a Newton derivative of f at x. By Proposition 3.7, f is not Bouligand differentiable at x.

EXAMPLE 3.10. (i) For $X = \ell^1$, $f(x) = \|x\|_1 = \sum_{k=1}^\infty |x_k|$, we claim that ∂f is not a Newton derivative of f on X. The dual of ℓ^1 is isometrically isomorphic to ℓ^∞, and $x^* = (x_k^*) \in \partial f(x)$ if and only if $x_k^* \in \operatorname{sign}(x_k)$ for all k. Let $x = (x_k) \in \ell^1$ with $x_k \neq 0$ for infinitely many k, let

$$h_n = -2\sum_{k=n}^{\infty} x_k e_k\,, \quad n \in \mathbb{N}\,,$$

e_k being the unit vectors. Then for all n and all $x^* \in \partial f(x+h_n)$ we have $f(x+h_n) = f(x)$ and

$$f(x+h_n) - f(x) - \langle x^*, h_n\rangle = \sum_{k=n}^{\infty} \operatorname{sign}(-x_k)\cdot(-2x_k) = 2\sum_{k=n}^{\infty}|x_k| = -\|h_n\|_1\,.$$

As $\|h_n\|_1 \to 0$ as $n \to \infty$ and $\|h_n\|_1 \neq 0$ for all n, (2.1) does not hold.
(ii) A similar reasoning shows that for $X = L^1(\Omega)$, $\Omega \subset \mathbb{R}^n$ open, ∂f is not a Newton derivative of $f(x) = \|x\|_1 = \int_\Omega |x|$ on X. For $x \in L^1(\Omega)$ with $x \neq 0$ we choose $A_\varepsilon \subset \Omega$ with $0 < \int_{A_\varepsilon} |x| \to 0$ as $\varepsilon \to 0$ and set $h_\varepsilon = -2x$ on A_ε and $h_\varepsilon = 0$ elsewhere. Then $\|x + h_\varepsilon\|_1 = \|x\|_1$ and $\|h_\varepsilon\|_1 \to 0$ as $\varepsilon \to 0$. The elements $x^* \in \partial f(x+h_\varepsilon)$ satisfy

$$\langle x^*, h_\varepsilon\rangle = \int_{A_\varepsilon} \operatorname{sign}(-x)\cdot(-2x) = 2\int_{A_\varepsilon}|x| = \|h_\varepsilon\|_1\,,$$

so again (2.1) does not hold. □

Finally, we present a continuous convex function on the spaces ℓ^p for $p < \infty$ (which includes the Hilbert space ℓ^2) whose subdifferential is a Newton derivative at 0 but which is not Bouligand differentiable at 0. It is obtained by adapting the function presented in [19] which has a Newton derivative at 0 but is not directionally differentiable at 0.

EXAMPLE 3.11. Let $X = U = \ell^p$, $1 \le p < \infty$. As in Example 3.8 we consider a function $f : \ell^p \to \mathbb{R}$ of the form

$$f(x) = \sup_{k\in\mathbb{N}} f_k(x_k)\,.$$

Let $a_1 = 1$, $b_k = e^{-k}a_k$ and $a_{k+1} = e^{-k}b_k$ for $k \ge 1$; using induction we see that $b_k = e^{-k^2}$. We define $f_k : \mathbb{R} \to \mathbb{R}$ by

$$f_k(t) = \begin{cases} \frac{k+1}{k}t - \frac{a_k}{k}\,, & t \ge a_k\,,\\ \frac{t}{k}\ln(t/b_k)\,, & b_k < t < a_k\,,\\ 0\,, & t \le b_k \end{cases} \tag{3.22}$$

We have $f_k \ge 0$, $f_k(0) = 0 = f_k(b_k) = f_k(b_k+)$ and $f_k(a_k-) = f_k(a_k) = a_k$. Thus, the functions f_k are continuous.

We immediately see that f is not Bouligand differentiable at $x = 0$ since for $h^n = a_n e_n$ the Bouligand remainder satisfies

$$r^B(h^n) = f(h^n) - f(0) - f'(0; h_n) = f(h^n) = f_n(a_n) = a_n = \|h^n\|_p$$

and $h^n \to 0$ as $n \to \infty$.

The derivative of f_k is given by

$$f_k'(t) = \begin{cases} \frac{k+1}{k}\,, & t > a_k\,, \\ \frac{1}{k}(\ln(t/b_k)+1)\,, & b_k < t < a_k\,, \\ 0\,, & t < b_k\,. \end{cases} \tag{3.23}$$

The branches of f_k' are monotonically increasing and there holds $f_k'(b_k+) = \frac{1}{k} > 0 = f_k'(b_k-)$, $f_k'(a_k+) = \frac{k+1}{k} = f_k'(a_k-)$. Hence, all f_k are convex, and therefore so is f. Also, f_k is Lipschitz with constant $(k+1)/k \le 2$. Thus, all f_k are 2-Lipschitz. By the same reasoning as above in (3.13),

$$|f(x) - f(y)| \le \sup_{k\in\mathbb{N}} |f_k(x_k) - f_k(y_k)| \le \sup_{k\in\mathbb{N}} 2|x_k - y_k| \le 2\|x-y\|_p\,.$$

Thus, f is Lipschitz continuous.

It remains to show that the Newton remainder r^N at $x = 0$ given by (3.7) as

$$r^N(h) = f'(h;h) - f(h) + f(0) = f'(h;h) - f(h) \tag{3.24}$$

satisfies $r^N(h) = o(\|h\|_p)$. It suffices to prove that this holds on the subsets $S_1 = \{h : f(h) > 0, \|h\|_p \le 1\}$ and $S_2 = \{h : f(h) = 0, \|h\|_p \le 1\}$ of ℓ^p.

(1) Let $h \in S_1$. Since $p < \infty$, $h_k \to 0$ as $k \to \infty$. Therefore there exist indices $\ell \le L$ with $f_\ell(h_\ell) > \max_{k>L} f_k(h_k)$. Since the family $\{f_k\}_{k\in\mathbb{N}}$ is equicontinuous, it follows that f can be written in some open neighbourhood of h as the maximum $f = \max_{k\le L} f_k$ of finitely many functions f_k. By well-known results,

$$f'(h;h) = \max_{j\in J} f_j'(h_j;h_j)\,,$$

where $J = \{j : f(h) = f_j(h_j)\}$.

Let $k \in J$ be an index with $f'(h;h) = f_k'(h_k;h_k)$. Then $f_k(h_k) = f(h) > 0$ and thus $h_k > b_k$. From (3.24) we get

$$r^N(h) = f_k'(h_k;h_k) - f_k(h_k)\,.$$

Using (3.22) and (3.23) we obtain that

$$r^N(h) = \frac{k+1}{k}h_k - \frac{k+1}{k}h_k + \frac{a_k}{k} = \frac{a_k}{k} \qquad \text{if } h_k \ge a_k,$$

and

$$r^N(h) = \frac{h_k}{k}\Big(\ln\Big(\frac{h_k}{b_k}\Big)+1\Big) - \frac{h_k}{k}\ln\Big(\frac{h_k}{b_k}\Big) = \frac{h_k}{k} \qquad \text{if } b_k < h_k < a_k.$$

Thus $r^N(h) \le h_k/k$. Since $b_k = e^{-k^2}$ we have $k = \sqrt{\ln(1/b_k)}$. As r^N is nonnegative by definition and $h_k > b_k$, we conclude that

$$0 \le r^N(h) \le \frac{h_k}{k} = \frac{h_k}{\sqrt{|\ln b_k|}} \le \frac{h_k}{\sqrt{|\ln h_k|}} \le \frac{\|h\|_p}{\sqrt{|\ln(\|h\|_p)|}} \tag{3.25}$$

on S_1.

(2) Let $h \in S_2$. Then $f_k(h_k) = 0$ and $h_k \le b_k \le 1$ for all k, and

$$r^N(h) = f'(h;h) = \lim_{t\downarrow 0} \frac{f(h+th)}{t} = \lim_{t\downarrow 0} \frac{1}{t}\sup_{k\in\mathbb{N}} f_k(h_k + th_k)\,. \tag{3.26}$$

by (3.24). Our goal now is to estimate $f_k(h_k + th_k)$. Let $t \in (0,1]$, $k \geq 1$. If $h_k \leq b_k/2$ we have $f_k(h_k + th_k) = 0$. If $b_k/2 < h_k \leq b_k$ we use the mean value inequality

$$f_k(h_k + th_k) = f_k(h_k + th_k) - f_k(h_k) \leq t \sup_{0\leq s\leq t} h_k f'(h_k + sh_k)\,.$$

As $h_k < h_k + th_k \leq 2h_k \leq 2b_k < a_k$ and f' is nondecreasing,

$$f_k(h_k + th_k) \leq th_k f_k'(2b_k) = th_k \frac{1}{k}(1 + \ln 2)\,. \tag{3.27}$$

As in (1) above we use the formula $k = \sqrt{\ln(1/b_k)}$ and obtain

$$k = \sqrt{\ln(2/b_k) - \ln 2} \geq \sqrt{\ln(1/h_k) - \ln 2} \qquad \text{if } b_k/2 < h_k \leq b_k.$$

In view of (3.26) and (3.27) and since $h_k \leq b_k \leq 1$ we arrive at

$$0 \leq r^N(h) \leq (1 + \ln 2)\|h\|_p \,|\ln \|h\|_p + \ln 2|^{-1/2} \tag{3.28}$$

on S_2. Since (3.25) and (3.28) imply that

$$r^N(h) = O\big(\|h\|_p \,|\ln(\|h\|_p)|^{-1/2}\big) = o(\|h\|_p) \tag{3.29}$$

on ℓ^p, we have shown that ∂f is a Newton derivative of f at 0. □

Relation to upper semicontinuity. It is known that a continuous convex function on a normed space X is Bouligand differentiable at a point x if and only if its subdifferential is upper semicontinuous w.r.t. the norm topology in X^*. In fact, this is a consequence of Theorem 3.1 in [11], a more general result in the framework of locally convex spaces. For the convenience of the reader we present the special case used here with a full proof.

PROPOSITION 3.12. *Let $f : U \to \mathbb{R}$ be convex and continuous on an open convex subset U of a Banach space X. Then there are equivalent:*
(i) f is Bouligand differentiable at $x \in U$ (in U, resp.).
(ii) ∂f is norm-usc at x, that is, $\partial f : U \rightrightarrows (X^*, \|\cdot\|_{X^*})$ is upper semicontinuous at $x \in U$ (in U, resp.).
Moreover, (i) and (ii) imply that ∂f is a Newton derivative for f at $x \in U$ (in U, resp.).

Proof. The last claim is a direct consequence of Proposition 3.7(i).
"(ii)⇒(i)": Let $x \in U$ be arbitrary. For every $y \in U$ and $y^* \in \partial f(y)$ we get, using (3.1) with $h = y - x$ as well as (3.3),

$$0 \leq f(y) - f(x) - f'(x; y - x) \leq -\langle y^*, x - y\rangle - f'(x; y - x)\,. \tag{3.30}$$

It follows from (3.4) that for every $x^* \in \partial f(x)$

$$\langle y^*, y - x\rangle - f'(x; y - x) \leq \langle y^* - x^*, y - x\rangle \leq \|y^* - x^*\| \cdot \|y - x\|\,.$$

Passing to the infimum with respect to x^* yields

$$\langle y^*, y - x\rangle - f'(x; y - x) \leq \|y - x\| \cdot \operatorname{dist}(y^*, \partial f(x))\,.$$

In view of (3.30) we arrive at

$$|f(y) - f(x) - f'(x; y - x)| \leq \|y - x\| \cdot \operatorname{dist}(y^*, \partial f(x))\,. \tag{3.31}$$

By upper semicontinuity (see (5.2)),

$$\sup_{y^*\in\partial f(y)} \operatorname{dist}(y^*, \partial f(x)) \to 0 \quad \text{as} \quad y \to x\,.$$

Thus, (i) follows from (3.31).

"(i)⇒(ii)": Let $x \in U$ and $\varepsilon > 0$ be arbitrary. It suffices to prove that there exists $\delta > 0$ such that

$$\partial f(B_\delta(x)) \subset B_\varepsilon(\partial f(x))\,. \tag{3.32}$$

(Here, $B_\delta(x)$ denotes the open ball centered at x with radius δ.) By Proposition 3.1(i), f is locally Lipschitz. Let $\gamma > 0$ be such that, with $\delta = \frac{\gamma\varepsilon}{8L}$, $B_{\gamma+\delta}(x) \subset U$, that f is Lipschitz on $B_{\gamma+\delta}(x)$ with modulus L, and that

$$r^B(h) = f(x+h) - f(x) - f'(x;h) \le \frac{\varepsilon}{4}\|h\| \qquad \text{for all } h \in B_\gamma(0). \tag{3.33}$$

We abbreviate $\eta := 2L\delta = \frac{\gamma\varepsilon}{4}$. Let $y \in B_\delta(x)$ and $y^* \in \partial f(y)$. Then for all $h \in B_\gamma(0)$ there holds

$$\begin{aligned}\langle y^*, h\rangle &\le f(y+h) - f(y) \le f(x+h) - f(x) + 2L\|y-x\| \\ &\le f(x+h) - f(x) + 2L\delta = f(x+h) - f(x) + \eta.\end{aligned}$$

Thus, for all $x^* \in \partial f(x)$ and all $h \in B_\gamma(0)$ we have

$$\langle y^* - x^*, h\rangle \le f(x+h) - f(x) + \eta - \langle x^*, h\rangle\,.$$

It follows from (3.4) and (3.33) that for all $h \in B_\gamma(0)$

$$\begin{aligned}\inf_{x^*\in\partial f(x)} \langle y^* - x^*, h\rangle &\le f(x+h) - f(x) + \eta - f'(x;h) \\ &\le \frac{\varepsilon}{4}\|h\| + \eta < \frac{1}{2}\gamma\varepsilon\,.\end{aligned} \tag{3.34}$$

Let $K = \{z^* \in X^* : \|z^*\| \le (3/4)\varepsilon\}$. It suffices to prove that

$$(y^* - \partial f(x)) \cap K \ne \emptyset\,, \tag{3.35}$$

since then there exists $x^* \in \partial f(x)$ with $\|y^* - x^*\| < \varepsilon$.

Let us assume that (3.35) does not hold. Since the sets $y^* - \partial f(x)$ and K are w^*-compact subsets of X^*, by strict separation in (X^*, w^*) we find $h \in X$ such that

$$\inf_{x^*\in\partial f(x)} \langle y^* - x^*, h\rangle > \sup_{z^*\in K} \langle z^*, h\rangle = \frac{3}{4}\varepsilon\|h\|\,, \tag{3.36}$$

see e.g. paragraph 11F on p.64 in [14]. As we may choose h in (3.36) such that $\|h\| = (2/3)\gamma$, we arrive at a contradiction to (3.34). Thus (3.35) holds and the proof is complete. □

REMARK 3.13. By a slight modification of the proof of the implication "(ii)⇒(i)" one obtains a direct proof that ∂f is a Newton derivative if it is norm-usc. One replaces (3.30) with

$$0 \le f(x) - f(y) - \langle y^*, x-y\rangle \le -f'(x; y-x) - \langle y^*, x-y\rangle$$

and then uses this estimate instead of (3.30) to replace (3.31) with

$$|f(y) - f(x) - \langle y^*, y-x\rangle| \le \|y-x\| \cdot \operatorname{dist}(y^*, \partial f(x))\,.$$

REMARK 3.14. Proposition 3.12 provides an alternative proof that if X has finite dimension then f is Bouligand differentiable and ∂f is a Newton derivative of f. Indeed, in that case the norm topology and the weak star topology coincide on X^*, so ∂f is norm-usc by Proposition 3.2(iv),

Newton derivatives via compact embeddings. In Example 3.9 we have seen that the subdifferential of the maximum functional is not a Newton derivative w.r.t. the standard norm of $C[a,b]$. Nevertheless it turns out that on a space compactly embedded into $C[a,b]$ the subdifferential actually becomes a Newton derivative of the maximum functional. In view of Proposition 3.12 this will be a consequence of Proposition 3.17 below where we prove that $f = f_0 \circ A$ is norm-usc if A is a compact operator and f_0 is convex and continuous.

As a preparation we need a lemma.

Let X, Y be Banach spaces, $U \subset X$ convex and open, $A \in \mathcal{L}(X, Y)$, $\Psi : U \rightrightarrows Y^*$. By $A^* \in \mathcal{L}(Y^*, X^*)$ we denote the dual of A defined by $A^*y^* = y^* \circ A$ for $y^* \in Y^*$.

LEMMA 3.15. *In the situation above, let A be compact, Ψ locally bounded with w^*-compact values, and let $\Psi : U \rightrightarrows (Y^*, w^*)$ be usc. Then $A^*\Psi : U \rightrightarrows X^*$ is norm-usc.*

Proof. Let M be an arbitrary closed subset of X^*. It suffices to show (see Definition 5.1) that

$$(A^*\Psi)^{-1}(M) = \{x \in X : (A^*\Psi)(x) \cap M \neq \emptyset\} \tag{3.37}$$

is a closed subset of X. Let $x_n \in (A^*\Psi)^{-1}(M)$ with $x_n \to x \in X$. We have to show that $(A^*\Psi)(x) \cap M$ is not empty. By the choice of x_n there exist $y_n^* \in \Psi(x_n)$ with $x_n^* = A^*y_n^* \in M$. As Ψ is locally bounded, $\{y_n^*\}$ is bounded in Y^*. Since A^* is compact by a general result of functional analysis, passing to a subsequence we obtain $x_n^* \to x^*$ for some $x^* \in X^*$. As M is closed, $x^* \in M$. It now suffices to show that $x^* = A^*y^*$ for some $y^* \in \Psi(x)$. Since $\{y_n^*\}$ is bounded, $y_\beta^* \overset{*}{\rightharpoonup} y^*$ for some subnet of $\{y_n^*\}$ and some $y^* \in Y^*$. Consequently, $x_\beta^* = A^*y_\beta^* \overset{*}{\rightharpoonup} A^*y^*$, since $A^* : (Y^*, w^*) \to (X^*, w^*)$ is continuous As limits are unique, $x^* = A^*y^*$. Now $y_\beta^* \in \Psi(x_\beta)$, $x_\beta \to x$ in X, $y_\beta^* \overset{*}{\rightharpoonup} y_*$ in Y^*. By Proposition 5.4(i), the graph of Ψ is closed in $U \times (Y^*, w^*)$, so $y^* \in \Psi(x^*)$. The proof is complete. ∎

REMARK 3.16. We refer to [17] for the definition and properties of nets in general topology. If X is separable, one can replace subnets by subsequences in the proof above, since bounded subsets of X^* are sequentially weak star compact in that case.

Let X and Y be Banach spaces, $V \subset Y$ open and convex, $A \in \mathcal{L}(X, Y)$ compact and $f_0 : V \to \mathbb{R}$ convex and continuous. We consider the function

$$f : U \to \mathbb{R}, \quad f = f_0 \circ A, \quad U = A^{-1}(V). \tag{3.38}$$

PROPOSITION 3.17. *In the situation above, $\partial f : U \rightrightarrows X^*$ is a Newton derivative of f in U, and f is Bouligand differentiable in U.*

Proof. The function f is convex and continuous on U. By [9, Ch.I,Prop.5.7], $\partial f(x) = A^*\partial f_0(Ax)$. The set-valued mapping $\Psi = (\partial f_0) \circ A$ satisfies the assumptions of Lemma 3.15 by virtue of Proposition 3.2(i), (ii) and (iv). Therefore, ∂f is norm-usc. The result now follows from Propositions 3.12 and 3.7. ∎

EXAMPLE 3.18. Let X be a Banach space which is compactly embedded into $Y = C[a,b]$. Let $f_0 : C[a,b] \to \mathbb{R}$ be the maximum functional, $f_0(x) = \max_s x(s)$. Applying Proposition 3.17 to the restriction $f = f_0|X$ of f_0, we obtain that $\partial f : X \rightrightarrows X^*$ is a Newton derivative of f in X. Since for all $x \in X$

$$\partial f(x) = \{y^*|X : y^* \in \partial f_0(x)\},$$

formula (3.15) remains valid. In particular, the maximum functional is semismooth on the Hölder spaces $C^{0,\alpha}[a,b]$ for $0 < \alpha \le 1$ and on the Sobolev spaces $W^{1,p}(a,b)$ for $1 < p \le \infty$. □

Improving the remainder estimate. In addition to (3.38) let X_0 be another Banach space such that $X \subset X_0$ continuously. As we will see in Proposition 3.20 below, we can then improve the remainder estimate (2.3) for f to

$$\sup_{x^* \in \partial f(x+h)} |f(x+h) - f(x) - \langle x^*, h\rangle| \le \rho_x(\|h\|_0) \cdot \|h\|, \tag{3.39}$$

provided it is possible to extend the compact operator $A \in \mathcal{L}(X,Y)$ to an operator $A_0 \in \mathcal{L}(X_0,Y)$ which however is not required to be compact. Our proof is based on interpolation theory; the relevant facts are collected in the following lemma.

LEMMA 3.19. *Let $(X_0, \|\cdot\|_0)$ and $(X_1, \|\cdot\|_1)$ be Banach spaces such that $X_1 \subset X_0$ continuously, let $\theta \in (0,1)$.*
(i) There exists a Banach space $(X_\theta, \|\cdot\|_\theta)$ and a constant $c_\theta > 0$ such that $X_1 \subset X_\theta \subset X_0$ continuously and

$$\|x\|_\theta \le c_\theta \|x\|_0^{1-\theta} \|x\|_1^\theta \quad \text{for all } x \in X_1. \tag{3.40}$$

(ii) Let Y be another Banach space, let $A_1 \in \mathcal{L}(X_1, Y)$ and $A_0 \in \mathcal{L}(X_0, Y)$ such that $A_0|X_1 = A_1$. If A_1 is compact, then so is $A_\theta = A_0|X_\theta : X_\theta \to Y$.

Proof. For X_θ one may choose any space $[X_0, X_1]_{\theta,p}$ constructed in real interpolation theory. See [22], Proposition 1.5 and Corollary 1.7 for (i), and [1], Theorem 3.8.1 for (ii). □

Again, let

$$f : U \to \mathbb{R}, \quad f = f_0 \circ A, \quad U = A^{-1}(V), \tag{3.41}$$

where X and Y are Banach spaces, $V \subset Y$ open and convex, $A \in \mathcal{L}(X,Y)$ compact and $f_0 : V \to \mathbb{R}$ is convex and continuous.

PROPOSITION 3.20. *In the situation above, let $X \subset X_0$ continuously for some Banach space $(X_0, \|\cdot\|_0)$, and let A admit a linear continuous extension $A_0 : X_0 \to Y$. Then*

$$\sup_{x^* \in \partial f(x+h)} |f(x+h) - f(x) - \langle x^*, h\rangle| \le \rho_x(\|h\|_0) \cdot \|h\|, \tag{3.42}$$

as well as

$$|f(x+h) - f(x) - f'(x;h)| \le \rho_x(\|h\|_0) \cdot \|h\|, \tag{3.43}$$

for some nondecreasing function $\rho_x : (0, \delta_0) \to \mathbb{R}$ with $\rho_x(\delta) \to 0$ as $\delta \to 0$.

Proof. Setting $X_1 = X$, fix $\theta \in (0,1)$, and let X_θ be a Banach space according to Lemma 3.19(i). By part (ii) of this lemma, $A_\theta = A_0|X_\theta$ is compact. Then $U \subset U_\theta := A_\theta^{-1}(V)$. By Proposition 3.17, $f_\theta = f_0 \circ A_\theta : U_\theta \to \mathbb{R}$ is Bouligand differentiable, and $\partial f_\theta : U_\theta \rightrightarrows X_\theta^*$ is a Newton derivative of f_θ. Therefore,

$$|f_\theta(x+h) - f_\theta(x) - f'(x;h)| \le \rho_{\theta,x}(\|h\|_\theta)\|h\|_\theta \tag{3.44}$$

holds for all $h \in X_\theta$ with $\|h\|_\theta$ small enough, where $\rho_{\theta,x}$ is nondecreasing and satisfies $\rho_{\theta,x}(\delta) \to 0$ as $\delta \to 0$. As $\|h\|_\theta \le c\|h\|$ for all $h \in X$ and some constant c, due to (3.40) we may estimate the remainder term as

$$\rho_{\theta,x}(\|h\|_\theta)\|h\|_\theta \le \rho_{\theta,x}(c_\theta \|h\|_0^{1-\theta}\|h\|^\theta)c\|h\| \,. \tag{3.45}$$

Consequently, for all $\|h\| < \delta_0$, δ_0 small enough, setting

$$\rho_x(\delta) = c\rho_{\theta,x}(c_\theta \delta^{1-\theta}\delta_0^\theta)$$

we arrive at (3.43). Replacing the left side of (3.44) by the corresponding expression for the Newton derivative, we also obtain (3.42). ∎

EXAMPLE 3.21. Continuing Example 3.18, again let X be a Banach space which is compactly embedded via $A : X \to Y = C[a,b]$. Let $X_0 = Y$, $A_0 = id$, $f_0 : X_0 \to \mathbb{R}$ the maximum functional and $f = f_0|X$. From Proposition 3.20 we obtain the improved remainder estimate

$$\sup_{x^* \in \partial f(x+h)} |f(x+h) - f(x) - \langle x^*, h\rangle| \le \rho_x(\|h\|_\infty) \cdot \|h\|_X \,.$$

This applies to $X = C^{0,\alpha}[a,b]$ for $\alpha > 0$ and to $X = W^{1,p}(a,b)$ for $p > 1$. □

4. Semismoothness of parametric functions. Let X, Z be normed spaces, let $U \subset X$ be open and let $(P, \mathcal{A}, \pi)$ be a measure space equipped with a $\sigma-$algebra $\mathcal{A}$ and a measure π. For a given function

$$f : U \times P \to Z \tag{4.1}$$

we want to investigate the semismoothness of the operator F defined by

$$(Fx)(p) = f(x,p) \,, \quad F : U \to L^r(P, Z) \,, \quad 1 \le r < \infty \,. \tag{4.2}$$

Here, $L^r(P,Z)$ denotes the space of Bochner measurable functions $z : P \to Z$ which are Bochner integrable to the power r,

$$\int_P \|z(p)\|_Z^r \, d\pi < \infty \,.$$

Our goal is to construct a Newton derivative of F from Newton derivatives $x \mapsto g(x,p)$ of the functions $x \mapsto f(x,p)$,

$$g : U \times P \rightrightarrows \mathcal{L}(X, Z) \,. \tag{4.3}$$

We assume that

$$p \mapsto f(x,p) \text{ is Bochner measurable for all } x \in U \tag{4.4}$$

and that for all $x \in U$ there exist a neighborhood $N_x = \{y : \|y - x\|_X < \delta_x\}$ of x and functions $c_x, C_x \in L^r(P)$ such that

$$\|f(x,p)\|_Z \le c_x(p) \qquad \text{a.e. in } p \tag{4.5}$$

as well as

$$\sup_{M \in g(y,p), y \in N_x} \|M\|_{\mathcal{L}(X,Z)} \le C_x(p) \qquad \text{a.e. in } p. \tag{4.6}$$

By Proposition 2.3 (the mean value inequality), (4.6) implies that for all $x \in U$

$$\|f(y,p) - f(y',p)\|_Z \le C_x(p)\|y - y'\|_X \qquad \forall\, y, y' \in N_x \tag{4.7}$$

holds a.e. in p. It then follows that for every $x \in U$,

$$\begin{aligned} \|Fx\|_{L^r(P,Z)} &\le \|c_x\|_{L^r(P)} \\ \|Fy - Fy'\|_{L^r(P,Z)} &\le \|C_x\|_{L^r(P)}\|y - y'\|_X \qquad \forall\, y, y' \in N_x\,. \end{aligned} \tag{4.8}$$

This ensures that F in (4.2) is well-defined and locally Lipschitz continuous. We moreover assume that for all $x \in U$

$$\sup_{M \in g(x+h,p)} \|f(x+h,p) - f(x,p) - Mh\|_Z \le \rho_{x,p}(v(h))\|h\|_X \tag{4.9}$$

for all $h \in X$ with $\|h\|_X < \delta_x$. Here, $\rho_{x,p} : \mathbb{R}_+ \to \mathbb{R}_+$ are nondecreasing functions with $\rho_{x,p}(\delta) \to 0$ as $\delta \to 0$, and $v : X \to \mathbb{R}_+$ is a function with $v(h) \to 0$ as $h \to 0$ in X. For the choice $v(h) = \|h\|_X$ this means that indeed $x \mapsto g(x,p)$ is a Newton derivative for $x \mapsto f(x,p)$; the choice $v(h) = \|h\|_{X_0}$ generalizes the setting of Section 3 concerning the refined remainder estimate (3.39) from the scalar-valued to the vector-valued case.

A Newton derivative G of F is a set-valued mapping

$$G : U \rightrightarrows \mathcal{L}(X, L^r(P,Z))\,. \tag{4.10}$$

In order to construct G from g, we require that elements $L \in G(x) \subset \mathcal{L}(X, L^r(P,Z))$, $x \in U$, satisfy for a.a. $p \in P$

$$(Ld)(p) \in g(x,p)d \qquad \text{for all } d \in X. \tag{4.11}$$

Accordingly we define G by

$$G(x) = \{L : L \in \mathcal{L}(X, L^r(P,Z)), (4.11) \text{ holds a.e. in } p\}\,, \quad x \in U\,. \tag{4.12}$$

In order that G becomes a Newton derivative of F, we want to obtain the remainder estimate

$$\sup_{L \in G(x+h)} \|F(x+h) - F(x) - Lh\|_{L^r(P,Z)} \le \rho_x(v(h))\|h\|_X$$

for some function $\rho_x : \mathbb{R}_+ \to \mathbb{R}_+$ with $\rho_x(\delta) \to 0$ as $\delta \to 0$. We define

$$\rho_x(\delta) = \sup_{\substack{0 < \|h\|_X < \delta_x \\ v(h) \le \delta}} \sup_{L \in G(x+h)} \frac{\|F(x+h) - F(x) - Lh\|_{L^r(P,Z)}}{\|h\|_X}\,. \tag{4.13}$$

PROPOSITION 4.1. *Let $f : U \times P \to Z$ and $g : U \times P \rightrightarrows \mathcal{L}(X, Z)$ satisfy (4.4) – (4.6) and (4.9) with functions c_x, C_x, $\rho_{x,p}$ and v as described above. Then:*
(i) $F : U \to L^r(P, Z)$ defined by $(Fx)(p) = f(x, p)$ is well-defined and locally Lipschitz continuous.
(ii) For $G : U \rightrightarrows \mathcal{L}(X, L^r(P, Z))$ from (4.12), the function ρ_x from (4.13) satisfies $\rho_x(\delta) \to 0$ as $\delta \to 0$ for all $x \in U$, and the remainder estimate

$$\sup_{L\in G(x+h)} \|F(x + h) - F(x) - Lh\|_{L^r(P,Z)} \le \rho_x(v(h))\|h\|_X \tag{4.14}$$

holds for all $x \in U$ and all $h \in X$ with $\|h\|_X < \delta_x$.

Proof. The claims in (i) immediately follow from (4.4) – (4.8). Concerning (ii), estimate (4.14) is a direct consequence of the definition of ρ_x in (4.13), since in the evaluation of $\rho_x(v(h))$ the element h belongs to the set over which the first supremum is taken.

Let $x \in U$ be arbitrary. It remains to prove that $\rho_x(\delta) \to 0$ as $\delta \to 0$. If this is not the case, there exist $\gamma > 0$ and $\delta_k > 0$ with $\rho_x(\delta_k) > \gamma$ for all $k \in \mathbb{N}$ and $\delta_k \to 0$. By (4.13), for all $k \in \mathbb{N}$ there exist $h_k \in X$ and $L_k \in G(x + h_k)$ with $0 < \|h_k\|_X < \delta_x$, $v(h_k) \le \delta_k$ and

$$\|F(x + h_k) - F(x) - L_k h_k\|_{L^r(P,Z)} \ge \gamma\|h_k\|_X \,. \tag{4.15}$$

Let $k \in \mathbb{N}$. Setting

$$r_k = F(x + h_k) - F(x) - L_k h_k$$

there holds

$$r_k(p) = f(x + h_k, p) - f(x, p) - M_k(p)h_k \qquad \text{a.e. in } p$$

for some $M_k(p) \in g(x + h_k, p)$ by (4.11). It follows that

$$\|r_k(p)\|_Z \le 2C_x(p)\|h_k\|_X \qquad \text{a.e. in } p \tag{4.16}$$

by (4.6) and (4.7), as well as

$$\|r_k(p)\|_Z \le \rho_{x,p}(v(h_k))\|h_k\|_X \qquad \text{a.e. in } p. \tag{4.17}$$

by (4.9). As $\delta_k \to 0$ it follows that $v(h_k) \to 0$ and hence $\rho_{x,p}(v(h_k)) \to 0$ for all $p \in P$. The function

$$d_k(p) = \frac{\|r_k(p)\|_Z}{\|h_k\|_X}$$

therefore converges to 0 a.e. in p by (4.17). It follows that $d_k \to 0$ in $L^r(P)$ by (4.16) and dominated convergence. On the other hand, $\|d_k\|_{L^r(P)} \ge \gamma$ for all k by (4.15), a contradiction. ∎

COROLLARY 4.2. (Newton derivative of F)
(i) Let the assumptions of Proposition 4.1 hold with $v(h) = \|h\|_X$, Then G as defined in (4.12) is a Newton derivative of $F : U \to L^r(P, Z)$, $(Fx)(p) = f(x, p)$.
(ii) Let X be continuously embedded in some normed space X_0, let the assumptions of Proposition 4.1 hold with $v(h) = \|h\|_{X_0}$. Then G as defined in (4.12) is a Newton derivative of $F : U \to L^r(P, Z)$, $(Fx)(p) = f(x, p)$ with the refined remainder estimate

$$\sup_{L\in G(x+h)} \|F(x + h) - F(x) - Lh\|_{L^r(P,Z)} \le \rho_x(\|h\|_{X_0})\|h\|_X \,. \tag{4.18}$$

Next, we want to obtain subsets $G_S(x)$ of $G(x)$ whose elements L can be described more explicitly in terms of selectors $\ell : U \times P \to \mathcal{L}(X,Z)$ of g. For this purpose, we equip $\mathcal{L}(X,Z)$ with the weak operator topology. This topology is generated by the family of seminorms $M \mapsto z^*(Md)$ where d ranges over X and z^* over Z^*, and makes $\mathcal{L}(X,Z)$ into a locally convex space.

LEMMA 4.3. *Let X be a normed space, $x \in U \subset X$, let Z be a separable Banach space, assume that (4.6) holds for $g : U \times P \rightrightarrows \mathcal{L}(X,Z)$. Let ℓ be a selector of g such that*

$$\ell_x : P \to (\mathcal{L}(X,Z), w)\,, \quad \ell_x(p) = \ell(x,p)\,, \tag{4.19}$$

is measurable. Then the mapping L defined by

$$(Ld)(p) = \ell(x,p)d \tag{4.20}$$

belongs to $G(x)$ as defined in (4.12).

Proof. Let $d \in X$, $z^* \in Z^*$. Since $M \mapsto z^*(Md)$ is a continuous functional on $(\mathcal{L}(X,Z), w)$, the mapping $p \mapsto z^*((Ld)(p)) = z^*(\ell(x,p)d)$ is a measurable mapping from P to $\mathbb{R}$ by assumption on ℓ. As this is true for all $z^* \in Z^*$, and as Z is assumed to be separable, the mapping $Ld : P \to Z$ is measurable by the measurability theorem of Pettis. Since $\|(Ld)(p)\|_Z \le C_x(p)\|d\|_X$ a.e. in P by (4.6), it follows that Ld belongs to $L^r(P,Z)$, that $d \mapsto Ld$ is linear and that $L \in \mathcal{L}(X, L^r(P,Z))$. ∎

Parametric convex functions. We apply the foregoing results to the case where U is an open convex subset of a separable Banach space X and

$$f : U \times P \to \mathbb{R} \tag{4.21}$$

is convex with respect to $x \in U$. Then $Z = \mathbb{R}$ is separable, and $(\mathcal{L}(X,Z), w)$ coincides with (X^*, w^*). We assume that f satisfies the Carathéodory conditions

$$\begin{aligned} &x \mapsto f(x,p) \quad \text{is convex and continuous for all } p \in P,\\ &p \mapsto f(x,p) \quad \text{is measurable for all } x \in U. \end{aligned} \tag{4.22}$$

By a standard result, (4.22) implies that $f : U \times P \to \mathbb{R}$ (and hence, its extension to $X \times P$ by $+\infty$) is measurable.

Let

$$g(x,p) = \partial_x f(x,p)\,, \quad g : U \times P \rightrightarrows (X^*, w^*)\,. \tag{4.23}$$

In order to obtain a large subset $G_S(x)$ of the Newton derivative $G(x)$ of F at $x \in U$ via the construction in Lemma 4.3, we apply certain results from set-valued analysis. For this purpose, we assume that the measure space $(P, \mathcal{A}, \pi)$ is complete, that is, the σ-algebra $\mathcal{A}$ includes all subsets of every set $A \in \mathcal{A}$ with $\pi(A) = 0$, and that π is σ-finite.

LEMMA 4.4. *The graph of the mapping*

$$g(x,\cdot) : P \rightrightarrows (X^*, w^*)\,, \quad x \in U\,, \tag{4.24}$$

is measurable.

Proof. Since f is measurable on $X \times P$, it follows from Proposition 9.5 in Chapter 2 of [15] that $f^* : (X^*, w^*) \times P \to \mathbb{R} \cup \{+\infty\}$,

$$f^*(x^*, p) = \sup_{x \in X} (f(x,p) - \langle x^*, x \rangle) \,, \tag{4.25}$$

is measurable. The graph of $g(x, \cdot)$ has the representation (see e.g. Proposition 5.1 in [9])

$$\begin{aligned} \operatorname{Gr} g(x, \cdot) &= \{(p, x^*) : p \in P, \, x^* \in \partial_x f(x,p)\} \\ &= \{(p, x^*) \in P \times X^* : \, f(x,p) + f^*(x^*, p) = \langle x^*, x \rangle\} \end{aligned} \tag{4.26}$$

and thus is measurable since f and f^* are measurable. ∎

For $x \in U$, let

$$S_x = \{\ell_x | \, \ell_x : P \to (X^*, w^*) \text{ measurable}, \, \ell_x(p) \in g(x,p) \; \forall \, p \in P\} \tag{4.27}$$

be the set of all measurable selectors of the mapping $p \mapsto g(x,p)$.

Proposition 4.5. *Let X be a separable Banach space, $U \subset X$ open and convex, let $(P, \mathcal{A}, \pi)$ be a complete measure space with π being σ-finite, let (4.5), (4.6) and (4.22) hold for f and g.*
(i) We have

$$g(x,p) \subset \overline{\{\ell_x(p) : \ell_x \in S_x\}} \qquad \text{for all } x \in U, \, p \in P, \tag{4.28}$$

the closure being taken w.r.t. the w^*-topology.
(ii) For $x \in U$, let $G_S(x)$ be the set of all $L \in \mathcal{L}(X, L^r(P))$ which for some $\ell_x \in S_x$ satisfy

$$(Ld)(p) = \ell_x(p) d \,, \qquad \text{for all } d \in X, \, p \in P. \tag{4.29}$$

Then $G_S(x) \subset G(x)$ as defined in (4.12).

Proof. Part (i) is an immediate consequence of Lemma 4.4 and Proposition 5.8 since (X^*, w^*) is a Suslin space. Part (ii) follows from Lemma 4.3. ∎

Concerning Bouligand differentiability of F, the situation is much simpler, and separability of X is not required. We again assume that (4.5), (4.6) and (4.22) hold. Since $f : U \times P \to \mathbb{R}$ is jointly measurable, the partial directional derivatives

$$f'_x((x,p); h) = \lim_{\lambda \downarrow 0} \frac{f(x + \lambda h, p) - f(x,p)}{\lambda} \tag{4.30}$$

are measurable, too, and

$$(F'(x;h))(p) = f'_x((x,p); h) \tag{4.31}$$

defines an element of $L^r(P)$. One may use dominated convergence directly to prove that $F'(x;h)$ is indeed a directional derivative of $F : U \to L^r(P)$. As in the proof of Proposition 4.1 one shows that F is Bouligand differentiable if the functions $f(\cdot, p)$ are Bouligand differentiable; moreover, the refined remainder estimate carries over analogously as in Corollary 4.2.

EXAMPLE 4.6. The function $f : C[a,b] \times [a,b] \to \mathbb{R}$

$$f(x,p) = \max_{a \le s \le p} x(s) \tag{4.32}$$

yields the cumulated maximum $F : C[a,b] \to C[a,b]$. In order to obtain Newton and Bouligand differentiability of F, let $U = X$ be a Banach space which is compactly embedded into $C[a,b]$ and let $P = [a,b]$. Then the arguments immediately above and Corollary 4.2 imply that $F : X \to L^r(a,b)$ is Newton as well as Bouligand differentiable; their derivatives are obtained as described above, see also [4].

5. Appendix: set-valued mappings.

In this section, we recall some standard results from set-valued analysis, given e.g. in [24], and derive some consequences needed in this paper.

Let $\Psi : X \rightrightarrows Y$. We generally assume that $\Psi(x) \neq \emptyset$ for every $x \in X$.

DEFINITION 5.1. *Let X, Y be Hausdorff topological spaces, let $\Psi : X \rightrightarrows Y$.*
(i) We say that Ψ is ***upper semicontinuous*** *(or* ***usc*** *for short), if*

$$\Psi^{-1}(A) := \{x : x \in X,\ \Psi(x) \cap A \neq \emptyset\} \tag{5.1}$$

is closed for every closed subset A of Y.
(ii) Let in particular X and Y be normed spaces. Ψ is called ***locally bounded*** *if for every $x \in X$ the sets $\{\|y\| : y \in \Psi(\xi), \|\xi - x\| \le \delta\}$ are bounded for some suitable $\delta = \delta(x)$. Ψ is called* ***globally bounded*** *if the bound can be chosen independently from x.*

LEMMA 5.2. *Let X, Y be Hausdorff topological spaces. A mapping $\Psi : X \rightrightarrows Y$ is upper semicontinuous (usc) if and only if for every $x \in X$ and every open set V with $\Psi(x) \subset V \subset Y$ there exists an open set $U \subset X$ with $x \in U$ and $\Psi(U) \subset V$.*

Proof. See Proposition 6.1.3 in [24]. □

If Y is a normed space, upper semicontinuity of Ψ at x is obviously equivalent to

$$\lim_{\xi \to x} \sup_{\eta \in \Psi(\xi)} \operatorname{dist}(\eta, \Psi(\mathrm{x})) = 0\,. \tag{5.2}$$

A single-valued mapping is continuous in the usual sense if and only if it is usc in the sense above.

The composition $\Psi_2 \circ \Psi_1$ of two set-valued mappings $\Psi_1 : X \rightrightarrows Y$ and $\Psi_2 : Y \rightrightarrows Z$ is defined as

$$(\Psi_2 \circ \Psi_1)(x) = \bigcup_{y \in \Psi_1(x)} \Psi_2(y)\,. \tag{5.3}$$

LEMMA 5.3. *Let X, Y, Z be Hausdorff topological spaces, let $\Psi_1 : X \rightrightarrows Y$ and $\Psi_2 : Y \rightrightarrows Z$ be usc. Then $\Psi_2 \circ \Psi_1$ is usc.*

Proof. This is straightforward, using Lemma 5.2. See Proposition 2.56 in [15]. □

We use Lemma 5.3 mainly for the special cases $\Psi \circ f$ and $f \circ \Psi$ where Ψ is usc and f is single-valued and continuous.

PROPOSITION 5.4. *Let X, Y be Hausdorff topological spaces, let $\Psi : X \rightrightarrows Y$. We assume that Ψ has compact values, that is, $\Psi(x)$ is compact for all $x \in X$.*
(i) If Ψ is usc, then the graph of Ψ,

$$\operatorname{Gr} \Psi = \{(x,y) : x \in X,\ y \in \Psi(x)\} \tag{5.4}$$

is closed in $X \times Y$.
(ii) If the graph of Ψ is closed in $X \times Y$ and if $\Psi(X)$ is relatively compact in Y, then Ψ is usc.

Proof. See Proposition 6.1.8, Remark 6.1.9 and Proposition 6.1.10 in [24]. ∎

DEFINITION 5.5. *Let X be a measurable space, Y a metric space. A set-valued mapping $\Psi : X \rightrightarrows Y$ is said to be* ***measurable*** *if $\Psi^{-1}(V)$ is measurable for all open $V \subset Y$. A mapping $\psi : X \to Y$ is called a* ***measurable selector*** *of Ψ if ψ is measurable and $\psi(x) \in \Psi(x)$ for every $x \in X$.*

LEMMA 5.6. *Let X be a Hausdorff space equipped with the Borel algebra, let Y be a metric space and $\Psi : X \rightrightarrows Y$ be usc. Then Ψ is measurable.*

Proof. This is a direct consequence of Proposition 6.2.3 in [24]. ∎

PROPOSITION 5.7. *Let X be a measurable space, Y a complete separable metric space, let $\Psi : X \rightrightarrows Y$ be such that $\Psi(x)$ is closed for all $x \in X$. Then the following are equivalent:*
(i) Ψ is measurable,
(ii) there exists a sequence $\{\psi_n\}_{n\in\mathbb{N}}$ of measurable selectors of Ψ such that for all $x \in X$

$$\Psi(x) = \overline{\{\psi_n(x) : n \in \mathbb{N}\}}\,. \tag{5.5}$$

Proof. See Proposition 4.3.3 in [7]. ∎

The proposition implies in particular that Ψ has a measurable selector if it is measurable (this is the Kuratowski–Ryll-Nardzewski selection theorem). It applies in particular if Y is a bounded subset of the dual of a separable Banach space, equipped with the w^*-topology.

When dealing in Section 4 with the parametric case, the range set Y will still be constructed as a subset of some dual space, but we do not want to assume that it is bounded. In that case we can use the following proposition since the dual of a separable Banach space, equipped with the w^*-topology, is a Suslin space. (A Suslin space is the image under a continuous mapping of a space homeomorphic to a complete metric space.)

PROPOSITION 5.8. *Let X be a complete measurable space, Y a Suslin space, let $\Psi : X \rightrightarrows Y$ be such that* $\operatorname{Gr}\Psi \subset X \times Y$ *is measurable. Then there exists a sequence $\{\psi_n\}_{n\in\mathbb{N}}$ of measurable selectors of Ψ such that for all $x \in X$*

$$\Psi(x) = \overline{\{\psi_n(x) : n \in \mathbb{N}\}}\,. \tag{5.6}$$

Proof. See Proposition 2.17 in Chapter 2 of [15]. ∎

The proposition implies in particular that Ψ has a measurable selector if its graph is measurable (this is the Yankov–von Neumann–Aumann selection theorem).

Acknowledgements. The authors would like to acknowledge the DFG priority program SPP 1962, which enabled interactions that were helpful for carrying out this joint work.

REFERENCES

[1] J. Bergh, J. Löfström, *Interpolation Theory. An Introduction*, Springer-Verlag Berlin Heidelberg, 1976.

[2] J. Bolte, A. Daniilidis, A. Lewis, Tame functions are semismooth, *Math. Programming* **117** (2009), 5–19.

[3] J.F. Bonnans, A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[4] M. Brokate, Newton and Bouligand derivatives of the scalar play and stop operator, *Math. Model. Nat. Phenom.* **15** (2020), 51. DOI: https://doi.org/10.1051/mmnp/2020013.

[5] M. Brokate, P. Krejci, Weak Differentiability of Scalar Hysteresis Operators, *Discrete Contin. Dyn. Syst.* **35** (2015), 2405–2421.

[6] X. Chen, Z. Nashed, L. Qi, Smoothing methods and semismooth methods for nondifferentiable operator equations, *SIAM J. Numer. Anal.*, **38** (2000), 1200–1216.

[7] Z. Denkowski, S. Migórski, N.S. Papageorgiou, *An Introduction to Nonlinear Analysis*, Springer-Verlag, New York, 2003.

[8] C. Ding, D. Sun, J. Sun, K.-C. Toh, *Spectral operators of matrices: Semismoothness and characterizations of the generalized Jacobian SIAM Journal on Optim.* **30** (2020), 630–659.

[9] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North Holland and American Elsevier, Amsterdam and New York, 1976; SIAM, Philadelphia, 1999.

[10] F. Facchinei, J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. I & II*, Springer Series in Operations Research, Springer, New York, 2003.

[11] D.A. Gregory, Upper semi-continuity of subdifferential mappings, *Canad. Math. Bull.* **23**, 11–19.

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

[13] M. Hintermüller, K. Kunisch, PDE-constrained optimization subject to pointwise constraints on the control, the state, and its derivatives, *SIAM J. Opt.* **20** (2009), 1133–1156.

[14] R.B. Holmes, *Geometric functional analysis and its applications*, Springer 1975.

[15] S. Hu, N.S. Papageorgiu, *Handbook of multivalued analysis, volume I: Theory*, Kluwer 1997.

[16] K. Ito, K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, SIAM Series Advances in Design and Control, SIAM, Philadelphia, 2008.

[17] J.L. Kelley, *General Topology*, Springer-Verlag, New York 1975.

[18] D. Klatte, B. Kummer, Approximations and generalized Newton methods, *Math. Program. Ser. B* **168** (2018), 673–716.

[19] D. Klatte, B. Kummer, Corrections to: Approximations and generalized Newton methods, *Math. Program. Ser. A* **179** (2020), 469–471.

[20] B. Kummer, Newton's method for nondifferentiable functions. In: Advances in Mathematical Optimization, 114–125, Math. Res., 45, Akademie-Verlag, Berlin, 1988.

[21] B. Kummer, Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. in: Advances in Optimization (Lambrecht, 1991), 171–194, Lecture Notes in Econom. and Math. Systems, 382, Springer, Berlin, 1992.

[22] A. Lunardi, *Interpolation Theory*, Edizione della Normale, 3rd ed. 2018.

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

[24] N.S. Papageorgiu, S.T. Kyritsi-Yiallourou, *Handbook of Applied Analysis*, Springer 2009.

[25] R.R. Phelps, *Convex Functions, Monotone Operators, and Differentiability*, Springer-Verlag, Berlin Heidelberg 1993.

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

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

[28] H. Qi, X. Yang, Semismoothness of spectral functions, *SIAM J. Matrix Anal. Appl.* **25** (2003), 766–783.

[29] A. Schiela, A simplified approach to semismooth Newton methods in function space, *SIAM J. Optim.* **19** (2008), 1417–1432.

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

[31] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, SIAM, Philadelphia, 2011.

[32] D. Sun, J. Sun, Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems, *SIAM J. Numer. Anal.*, **40** (2003), 2352–2367.