Semismooth implicit functions

Florian Kruse[*]

August 18, 2016

Abstract

Semismoothness of implicit functions in infinite-dimensional spaces is investigated. We provide sufficient conditions for the semismoothness of given implicit functions and show how the existence of implicit functions that fulfill these conditions can be ensured. This approach also extends some finite-dimensional results.

Key words. Semismoothness, implicit function theorems, nonsmooth analysis, superposition operators, generalized derivatives, optimal control

AMS subject classifications. 47H30, 47J07, 49J20, 49J52

1 Introduction

This paper is devoted to semismooth implicit functions in infinite-dimensional settings. Yet, some of the results also extend the finite-dimensional theory on semismooth implicit functions.

Semismoothness was introduced by Mifflin in [Mif77] for real-valued functions on finite-dimensional vector spaces and generalized to vector-valued functions by Qi in [Qi93] and Qi and Sun in [QS93]. It is a concept of nonsmooth analysis that is tailored to obtain at least q-superlinear local convergence of Newton's method. As such, semismoothness has been applied with great success to many classes of nonsmooth optimization problems including infinite-dimensional problems. In particular, it is frequently used in PDE-constrained optimization, cf., e.g., the books [IK08, HPUU09, Ulb11, De 15]. In fact, we discovered the need for a semismooth implicit function theorem in an infinite-dimensional setting in the context of optimal control with PDEs and an L^1-constraint, cf. [KK16].

Implicit function theorems in the presence of semismoothness have been investigated in [Sun01, PSS03, Gow04, MSZ05, vK08]. However, only finite-dimensional mappings are considered there. Implicit function theorems for infinite-dimensional semismooth mappings are, to the best of the author's knowledge, not available. The present paper aims at closing this gap.

Many of the different finite-dimensional notions of semismoothness that exist in the literature demand local Lipschitz continuity as well as the use of a specific generalized derivative. For

[*]University of Graz, Institute of Mathematics and Scientific Computing, Heinrichstr. 36, 8010 Graz, Austria (florian.kruse@uni-graz.at).

instance, the original definition in [Mif77] involves Clarke's generalized gradient. In contrast to this, common semismoothness concepts in infinite-dimensional spaces such as the one by Chen, Nashed, Qi [CNQ00] and Ulbrich [Ulb11] neither prescribe a specific generalized derivative nor do they require or imply local Lipschitz continuity. It is a key difference to the existing literature on semismooth implicit function theorems that we use a notion of semismoothness in this paper that includes those of [CNQ00] and [Ulb11]. To point out what the benefits of this approach are let us sketch some of our results.

Implicit function theorems for $F(x, y) = 0$ can be viewed as having two aspects: First, they provide the existence of a (usually unique) implicit function g such that $F(x, g(x)) = 0$. Second, they state that g inherits smoothness properties from F. In the classical implicit function theorem of Hildebrandt and Graves, for instance, g is C^m provided that F is C^m, where $m \in \mathbb{N} := \{0, 1, 2, \ldots\}$, cf. [Zei93, Theorem 4.B]. Similarly, we prove that if g exists, then semismoothness of F induces semismoothness of g in a very broad setting. This setting includes infinite-dimensional spaces and does not involve concrete properties of the generalized derivative that is used. In consequence, we can use *any* implicit function theorem that ensures existence of g and couple it with our semismoothness result. For instance, it may be possible to deduce the existence of g from monotonicity properties of F—especially if $F(x, y) = 0$ represents the optimality conditions of a convex optimization problem—, while its semismoothness can then be inherited from F. Or, if F is semismooth with respect to a specific generalized derivative, then the existence of g can be deduced from an implicit function theorem for this generalized derivative. This may turn out to be a powerful tool since several generalized derivatives are available for mappings between infinite-dimensional spaces, for instance the ones of Ioffe [Iof81], Thibault [Thi82], Ralph [Ral90], Páles and Zeidan [PZ07]. Let us also mention that by the very same reasoning it is possible to recover and prove some of the known finite-dimensional semismooth implicit function theorems in a structurally very clear fashion. Last but not least, the approach to develop conditions for the semismoothness of g under the assumption that g exists allows to establish semismoothness in cases where g is not unique, which seems to be a new result even in finite dimensions.

This paper is organized as follows. In Section 2 we present our findings on how to obtain semismoothness of a given implicit function. In Section 3 we review some nonsmooth implicit function theorems that guarantee the existence of an implicit function in infinite-dimensional settings and demonstrate how these results can be applied to a class of nonlinear superposition operators that is particularly important for optimal control with PDEs. In Section 4 we comment on the achieved results and make concluding remarks.

Let us fix some notation. By W we usually denote the product space of the spaces X and Y, i.e., $W := X \times Y$. Moreover, X always represents the parameter space, i.e., we are looking for an implicit function from X to Y. This implicit function is denoted by $x \mapsto g(x)$ and solves $F(x, y) = 0$, i.e., $F(x, g(x)) = 0$ for all x of some subset $N_X \subset X$. When working with generalized derivatives we consider set-valued mappings. For instance, an implicit function $g : N_X \to Y$ will have a generalized derivative $\partial g : N_X \rightrightarrows \mathcal{L}(X, Y)$, $x \mapsto \partial g(x)$. However, we do not consider set-valued maps for F or g. The terms *set-valued* and *multivalued* are used interchangeably. Clarke's generalized Jacobian for a mapping $g : N_X \to Y$, where $N_X \subset X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, is denoted by $\partial^{\mathrm{Cl}} g$ throughout this paper. For normed linear spaces V_1 and V_2 we write $\mathcal{L}(V_1, V_2)$ to denote the space of bounded linear mappings from V_1 to V_2. If X is a topological space, then a *neighborhood* of $S \subset X$ is any set $N_X \subset X$ that contains an open set O with $S \subset O$. In particular, neighborhoods are not required to be open.

For clarity let us mention that, in the same setting, we write $N_X \setminus \{\bar{x}\}$ also in cases where $\bar{x} \notin N_X$.

2 Semismoothness of implicit functions

In this section we provide conditions that ensure semismoothness of a *given* implicit function. The main reason to consider the semismoothness of implicit functions separately from the question of their existence is that the concepts of semismoothness that are usually employed in infinite-dimensional spaces, see for instance Definition 2 below, are not strong enough to guarantee the existence of implicit functions. Moreover, since this approach allows to work out clearly which assumptions are required for the semismoothness and which for the existence of an implicit function, it may also be helpful in situations where the existence of an implicit function is already known and only its semismoothness is left to argue. Also, providing separate results for semismoothness and existence enables us to obtain semismoothness even if the implicit function is locally not unique, a situation in which many implicit function theorems cannot be applied at all since they assert existence of a locally unique implicit function.

We stress that the following considerations are valid in infinite-dimensional spaces.

Standing assumption. *In Section 2 we suppose that X and Y are normed linear spaces—sometimes with additional properties—and that $W := X \times Y$ is endowed with a norm that satisfies $\|(x, y)\|_W \leq C_W(\|x\|_X + \|y\|_Y)$ for all $(x, y) \in X \times Y$ and a constant $C_W > 0$.*

Remark. If W is finite-dimensional, then this assumption is always satisfied.

2.1 Generalized semismoothness

We introduce a concept for semismoothness that is more general than the ones usually found in the literature. It will allow us to obtain the main results of Section 2 in greater generality. We begin by recalling two classical notions of semismoothness (but let us stress that several different notions of semismoothness exist in the literature).

Definition 1 (Semismoothness in finite-dimensional spaces)**.** Let $X := \mathbb{R}^n$ and $Y := \mathbb{R}^m$. Let $N_X \subset X$ be a neighborhood of $\bar{x}$. Let $g : N_X \to Y$ be Lipschitz in N_X and directionally differentiable at $\bar{x}$.

(a) We say that g is *semismooth at $\bar{x}$* iff

$$\sup_{M \in \partial^{\mathrm{Cl}} g(\bar{x}+h)} \|g(\bar{x} + h) - g(\bar{x}) - Mh\|_Y = o(\|h\|_X) \qquad \text{for } \|h\|_X \to 0.$$

(b) We say that g is *semismooth at $\bar{x}$ with order α* iff there exists $\alpha \in (0, 1]$ such that g is α-order B-differentiable at $\bar{x}$ and satisfies

$$\sup_{M \in \partial^{\mathrm{Cl}} g(\bar{x}+h)} \|g(\bar{x} + h) - g(\bar{x}) - Mh\|_Y = O(\|h\|_X^{1+\alpha}) \qquad \text{for } \|h\|_X \to 0.$$

Remark. The preceding definition of (α-order) semismoothness is equivalent to the definition of (α-order) semismoothness for vector-valued functions introduced by Qi and Sun in [QS93]. The proof of equivalence can be found in [Ulb11, Section 2].

A generally weaker notion of semismoothness is the following.

Definition 2 (Semismoothness in the sense of Ulbrich, cf. [HPUU09, Section 2.4.4])**.** Let X, Y be Banach spaces. Let $N_X \subset X$ be a neighborhood of $\bar{x}$. Consider a continuous mapping $g : N_X \to Y$ together with a multifunction $\partial g : N_X \rightrightarrows \mathcal{L}(X, Y)$ that satisfies $\partial g(x) \neq \emptyset$ for all $x \in N_X$.

(a) We say that g is *∂g-semismooth at $\bar{x}$* iff

$$\sup_{M \in \partial g(\bar{x}+h)} \|g(\bar{x} + h) - g(\bar{x}) - Mh\|_Y = o(\|h\|_X) \qquad \text{for } \|h\|_X \to 0.$$

(b) We say that g is *∂g-semismooth at $\bar{x}$ with order α* iff there exists $\alpha \in (0, 1]$ such that

$$\sup_{M \in \partial g(\bar{x}+h)} \|g(\bar{x} + h) - g(\bar{x}) - Mh\|_Y = O(\|h\|_X^{1+\alpha}) \qquad \text{for } \|h\|_X \to 0.$$

Remark. 1) For the sake of completeness we remark that the original definition in [HPUU09, Section 2.4.4] is not formulated for a neighborhood N_X but considers $g : X \to Y$.

2) More on these and other notions of semismoothness can be found in the monograph of Ulbrich, [Ulb11].

Semismoothness allows to establish q-superlinear convergence of Newton's method for non-smooth functions. Let us consider a function $g : X \to Y$ acting on Banach spaces and a set-valued mapping $\partial g : X \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images. Algorithm SSN provides a basic semismooth Newton method for the solution of $g(\bar{x}) = 0$, while Theorem 1 states the corresponding convergence result.

Algorithm SSN: Semismooth Newton method

Data: $x_0 \in X$.
for $k = 0, 1, 2, \ldots$ **do**
  if $g(x_k) = 0$ **then terminate** with x_k;
  Choose $M_k \in \partial g(x_k)$;
  Solve $M_k d_k = -g(x_k)$;
  Set $x_{k+1} := x_k + d_k$;
end

Theorem 1. *Let X, Y be Banach spaces. Let $N_X \subset X$ be a neighborhood of $\bar{x}$. Let $g : N_X \to Y$ be ∂g-semismooth at $\bar{x}$ and $g(\bar{x}) = 0$. Furthermore, suppose that there is $C > 0$ such that for each $x \in N_X$ every $M \in \partial g(x)$ is invertible with $\|M^{-1}\|_{\mathcal{L}(Y,X)} \leq C$. Then Algorithm SSN converges locally to $\bar{x}$ at a q-superlinear rate. If g is even semismooth of order α at $\bar{x}$, then the convergence rate is $1 + \alpha$.*

Proof. See [HPUU09, Theorem 2.12]. □

The following notion of semismoothness generalizes Definition 1 and 2 as well as many other concepts of semismoothness that appear in the literature. It also comprises *slant differentiability* (in particular, semismoothness) in the sense of [CNQ00]. We will use it to obtain the main result of this section in greater generality.

Definition 3 (Generalized semismoothness, generalized derivative, generalized differential)**.** Let X, Y be normed linear spaces. Let $N_X \subset X$ and $\bar{x} \in N_X$ be given. Consider a mapping $g : N_X \to Y$ together with a multifunction $\partial g : N_X \rightrightarrows \mathcal{L}(X,Y)$ that satisfies $\partial g(x) \neq \emptyset$ for all $x \in N_X$. In the sequel we will call such a multifunction a *generalized derivative of* g, while for $x \in N_X$ every $M \in \partial g(x)$ is called a *generalized differential of* g *at* x.

(a) We say that g is *generalized semismooth at* $\bar{x}$ *with respect to* ∂g iff

$$\lim_{\substack{x \to \bar{x}, \\ x \subset N_X \setminus \{\bar{x}\}}} \frac{\sup_{M \in \partial g(x)} \|g(x) - g(\bar{x}) - M(x - \bar{x})\|_Y}{\|x - \bar{x}\|_X} = 0.$$

(b) We say that g is *generalized semismooth of order* α *at* $\bar{x}$ *with respect to* ∂g iff there exist $\alpha > 0$ and $C > 0$ such that

$$\limsup_{\substack{x \to \bar{x}, \\ x \subset N_X \setminus \{\bar{x}\}}} \frac{\sup_{M \in \partial g(x)} \|g(x) - g(\bar{x}) - M(x - \bar{x})\|_Y}{\|x - \bar{x}\|_X^{1+\alpha}} \leq C.$$

Remark. 1) Note that N_X is not required to be a neighborhood of $\bar{x}$.

2) Note further that for g to be generalized semismooth at $\bar{x}$ the image of ∂g at $\bar{x}$ can be arbitrary.

3) Only the case where $\bar{x}$ is an accumulation point of $N_X \setminus \{\bar{x}\}$ is interesting.

4) If ∂g is bounded near $\bar{x}$, then generalized semismoothness implies continuity of g at $\bar{x}$.

5) In many of the common notions of semismoothness, $\partial g(x)$ is chosen point-based, in particular not making use of $\bar{x}$ (except for $x = \bar{x}$, of course). This is due to the fact that semismoothness at $\bar{x}$ is employed to prove superlinear convergence of Newton's method, where $\bar{x}$ represents a root of g and is therefore unknown. That is, a choice of $\partial g(x)$ that requires to know $\bar{x}$ is usually not computable, which renders such a choice meaningless for practical purposes. However, we allow choices that are not point-based since this helps to unify the theory. For instance, we are thereby able to include the case of Fréchet differentiability, cf. the following example.

Example 1. Here are three situations in which we have generalized semismoothness.

- If $g : X \to Y$ is Fréchet differentiable at $\bar{x}$, then it is generalized semismooth at $\bar{x}$ with respect to $\partial g(x) = \{g'(\bar{x})\}$.
- If $g : N_X \to Y$ is semismooth at $\bar{x}$ in the sense of Definition 1 or 2, then it is generalized semismooth at $\bar{x}$.
- We consider Algorithm SSN for $g : X \to Y$ and $\partial g : X \rightrightarrows \mathcal{L}(X,Y)$. Let $\bar{x} \in X$ with $g(\bar{x}) = 0$ and let $x_0 \in X$. Assume that this method generates a sequence $(x_k)_{k \geq 0}$ that accumulates at $\bar{x}$. It can be proven that (x_k) converges q-superlinearly to $\bar{x}$ provided $g|_{N_X}$ is generalized semismooth at $\bar{x}$ with respect to ∂g, where $N_X := \{x_k \in X : k \geq 0\}$, and provided there exist $\delta > 0$ and $C > 0$ such that for every k with $\|x_k - \bar{x}\|_X \leq \delta$ every $M \in \partial g(x_k)$ is invertible and satisfies $\|M^{-1}\|_{\mathcal{L}(Y,X)} \leq C$, cf. [HPUU09, Section 2.4.2].

2.2 Generalized semismoothness of implicit functions

The following notion is crucial to obtain generalized semismoothness of implicit functions. It is inspired by the *consistency condition* used in [Sch08].

Definition 4 (Uniform consistency)**.** Let X, Y be normed linear spaces and $W := X \times Y$. Let $\bar{w} := (\bar{x}, \bar{y}) \in W$ and $N_W \subset W$. Consider a mapping $G : N_W \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images.

(a) We say that G is *uniformly consistent at* $\bar{w}$ iff

$$\lim_{\substack{w \to \bar{w}, \\ w \subset N_W \setminus \{\bar{w}\}}} \frac{\sup_{M \in G(w)} \|M(x - \bar{x}) - (y - \bar{y})\|_Y}{\|w - \bar{w}\|_W} = 0,$$

where $w = (x, y)$.

(b) We say that G is *uniformly consistent of order* α *at* $\bar{w}$ iff there exist $\alpha > 0$ and $C > 0$ such that

$$\limsup_{\substack{w \to \bar{w}, \\ w \subset N_W \setminus \{\bar{w}\}}} \frac{\sup_{M \in G(w)} \|M(x - \bar{x}) - (y - \bar{y})\|_Y}{\|w - \bar{w}\|_W^{1+\alpha}} \leq C.$$

The following result shows how to obtain semismoothness of a given implicit function.

Theorem 2 (Generalized semismoothness of implicit functions I)**.** *Let X, Y be normed linear spaces and $W := X \times Y$. Let $N_X \subset X$, $N_W \subset W$, $g : N_X \to Y$ and $G : N_W \rightrightarrows \mathcal{L}(X, Y)$ with nonempty images be given. Let $\bar{w} := (\bar{x}, \bar{y}) \in N_W$ with $\bar{x} \in N_X$. Suppose that*

- *g satisfies $g(\bar{x}) = \bar{y}$ and $(x, g(x)) \in N_W$ for all $x \in N_X$;*
- *G is uniformly consistent (of order α) at $\bar{w}$.*

Define the generalized derivative $\partial g : N_X \rightrightarrows \mathcal{L}(X, Y)$ by $\partial g(x) := G(x, g(x))$ and assume that

- *there exists $L > 0$ such that either*

$$\|g(x) - g(\bar{x})\|_Y \leq L \|x - \bar{x}\|_X \qquad \text{or} \qquad \sup_{M \in \partial g(x)} \|M(x - \bar{x})\|_Y \leq L \|x - \bar{x}\|_X$$

  holds for all $x \in N_X$. In the second case assume additionally that g is continuous at $\bar{x}$.

Then g is generalized semismooth (of order α) at $\bar{x}$ with respect to ∂g.

Moreover, if N_X is a neighborhood of $\bar{x}$ and g is continuous in all of N_X, then g is semismooth (of order $\tilde{\alpha}$) at $\bar{x}$ in the sense of Ulbrich, where $\tilde{\alpha} := \min\{\alpha, 1\}$.

Proof. It follows directly from the definitions that if N_X is a neighborhood of $\bar{x}$ and g is continuous in N_X as well as generalized semismooth (of order α) at $\bar{x}$, then g is semismooth (of order $\tilde{\alpha}$) at $\bar{x}$ in the sense of Ulbrich. Hence, it suffices to establish the asserted generalized semismoothness. We conduct the proof for the case that G is consistent of order $\alpha > 0$. To argue in the case where G is only consistent, repeat the following proof with $\alpha = 0$ in Step 2). By definition there exist $\alpha > 0$ and $C > 0$ such that

$$\limsup_{\substack{w\to\bar{w},\\ w\subset N_W\setminus\{\bar{w}\}}} \frac{\sup_{M\in G(w)} \|M(x-\bar{x})-(y-\bar{y})\|_Y}{\|w-\bar{w}\|_W^{1+\alpha}} \le C. \tag{1}$$

Define

$$w : N_X \to N_W, \qquad w(x) := (x, g(x)) .$$

Step 1): We demonstrate the existence of $\hat{C} > 0$ such that $\|w(x)-w(\bar{x})\|_W \le \hat{C}\|x-\bar{x}\|_X$ holds for all $x \in N_X$ sufficiently close to $\bar{x}$. Since $\|w(x)-w(\bar{x})\|_W \le C_W(\|x-\bar{x}\|_X+\|g(x)-g(\bar{x})\|_Y)$ for all $x \in N_X$, it suffices to establish that there is $L > 0$ that satisfies $\|g(x)-g(\bar{x})\|_Y \le L\|x-\bar{x}\|_X$ for all $x \in N_X$ sufficiently close to $\bar{x}$. To prove this we may assume that

$$\sup_{M\in\partial g(x)} \|M(x-\bar{x})\|_Y \le \hat{L}\,\|x-\bar{x}\|_X$$

holds for all $x \in N_X$ and that g is continuous at $\bar{x}$. The definition of ∂g and the triangle inequality therefore yield

$$\|g(x)-g(\bar{x})\|_Y \le \sup_{M\in G(w(x))} \|g(x)-g(\bar{x})-M(x-\bar{x})\|_Y + \hat{L}\,\|x-\bar{x}\|_X\,. \tag{2}$$

The continuity of g at $\bar{x}$ and $\|w(x)-\bar{w}\|_W \le C_W(\|x-\bar{x}\|_X + \|g(x)-g(\bar{x})\|_Y)$ imply that $N_W \ni w(x) \to \bar{w}$ for $N_X \ni x \to \bar{x}$. Hence, it follows from (1) that

$$\sup_{M\in G(w(x))} \|g(x)-g(\bar{x})-M(x-\bar{x})\|_Y \le \frac{1}{2C_W}\,\|w(x)-w(\bar{x})\|_W \le \frac{1}{2}\Big(\|x-\bar{x}\|_X+\|g(x)-g(\bar{x})\|_Y\Big)$$

holds for all $x \in N_X$ sufficiently close to $\bar{x}$. In conclusion, we obtain from (2) the estimate

$$\|g(x)-g(\bar{x})\|_Y \le \left(\hat{L}+\frac{1}{2}\right)\|x-\bar{x}\|_X + \frac{1}{2}\,\|g(x)-g(\bar{x})\|_Y$$

for these x. Subtracting $\frac{1}{2}\|g(x)-g(\bar{x})\|_Y$ this establishes $\|g(x)-g(\bar{x})\|_Y \le L\|x-\bar{x}\|_X$ with $L := 2\hat{L}+1$.

Step 2): We know from Step 1) that $\|w(x)-\bar{w}\|_W \le \hat{C}\|x-\bar{x}\|_X$ holds for all $x \in N_X\setminus\{\bar{x}\}$ sufficiently close to $\bar{x}$. Hence, for these x we have

$$\sup_{M\in\partial g(x)} \frac{\|g(x)-g(\bar{x})-M(x-\bar{x})\|_Y}{\|x-\bar{x}\|_X^{1+\alpha}} \le \hat{C}^{1+\alpha} \sup_{M\in G(w(x))} \frac{\|g(x)-g(\bar{x})-M(x-\bar{x})\|_Y}{\|w(x)-\bar{w}\|_W^{1+\alpha}}$$

Since $w(x) \to \bar{w}$ for $x \to \bar{x}$ and $w(x) \in N_W$ for $x \in N_X$, the assertion follows from (1). □

Remark. 1) The property of $g : N_X \to Y$ that there exists $L > 0$ with

$$\|g(x)-g(\bar{x})\|_Y \le L\,\|x-\bar{x}\|_X$$

for all $x \in N_X$ is called *calmness of g at $\bar{x}$*. We point out that Step 1) in the proof of Theorem 2 is nothing else but to demonstrate the calmness of g at $\bar{x}$. Calmness is connected to stability of generalized equations and fixed points, cf. [Rob79] and [ADT14]. For further discussions of calmness let us also refer to the books [RW98] and [DR14].

2) If N_X is a neighborhood of $\bar{x}$, then the property $\sup_{M\in\partial g(x)}\|M(x-\bar{x})\|_Y \le L\|x-\bar{x}\|_X$ for all $x \in N_X$ is equivalent to $\sup_{M\in\partial g(x)}\|M\|_{\mathcal{L}(X,Y)} \le L$ for all $x \in N_X \setminus \{\bar{x}\}$.

3) If $\sup_{M\in\partial g(x)}\|M(x-\bar{x})\|_Y \le L\|x-\bar{x}\|_X$ is satisfied for all $x \in N_X$ and g is generalized semismooth at $\bar{x}$, then g is calm at $\bar{x}$, hence in particular continuous at $\bar{x}$. Therefore, the continuity of g at $\bar{x}$ required in the above theorem is a necessary condition to obtain generalized semismoothness of g.

4) Finally, let us point out that if there exists $C_X > 0$ such that $\|x\|_X \le C_X\|w\|_W$ holds for all $w = (x, y) \in W$ (which is always satisfied if W is finite-dimensional), then we can sharpen Theorem 2 by including the consistency assumption on G into its main statement as follows:

   Then g is generalized semismooth (of order α) at $\bar{x}$ with respect to ∂g if and only if $G|_{\tilde{N}_W}$ is uniformly consistent (of order α) at $\bar{w}$, where $\tilde{N}_W \subset N_W$ is given by $\tilde{N}_W := \{(x, g(x)) : x \in N_X\}$.

   That is, if all other prerequisites of Theorem 2 are fulfilled, then the uniform consistency of $G|_{\tilde{N}_W}$ is both necessary and sufficient for generalized semismoothness of g. To establish this, note that Theorem 2 already contains the *if* part of the new assertion since N_W can be replaced by $\tilde{N}_W$. To prove the *only if* part, assume that g is generalized semismooth of order α at $\bar{x}$ with respect to ∂g as defined in Theorem 2. Then $w \in \tilde{N}_W \setminus \{\bar{w}\}$ implies $w = w(x) = (x, g(x))$ with $x \in N_X \setminus \{\bar{x}\}$ and there holds

$$\frac{\sup_{M\in G(w)}\|M(x-\bar{x})-(y-\bar{y})\|_Y}{\|w-\bar{w}\|_W^{1+\alpha}} = \frac{\sup_{M\in G(w(x))}\|M(x-\bar{x})-(g(x)-g(\bar{x}))\|_Y}{\|w(x)-\bar{w}\|_W^{1+\alpha}} \le C_X^{1+\alpha}\sup_{M\in\partial g(x)}\frac{\|g(x)-g(\bar{x})-M(x-\bar{x})\|_Y}{\|x-\bar{x}\|_X^{1+\alpha}}$$

   for all these w. Since any sequence $(w_k) \subset \tilde{N}_W \setminus \{\bar{w}\}$ with $w_k \to \bar{w}$ has the form $w_k = (x_k, g(x_k))$ with $x_k \neq \bar{x}$ and since the inequality between $\|\cdot\|_X$ and $\|\cdot\|_W$ implies $x_k \to \bar{x}$, the assertion follows from the above inequality. Analogously for $\alpha = 0$.

We are now going to derive a result from the previous theorem in which the consistency assumption on G is replaced by assumptions that are easier to verify. To conveniently state this result we require the following definitions.

Definition 5. Let X, Y, Z be normed linear spaces and $W := X \times Y$. Let $\bar{w} \in N_W \subset W$ and consider a mapping $F : N_W \to Z$ together with a generalized derivative ∂F. For each generalized differential $M(w) \in \partial F(w)$ define the projected differentials $M_X(w) \in \mathcal{L}(X, Z)$ and $M_Y(w) \in \mathcal{L}(Y, Z)$ by

$$M_X(w)h_X := M(w)(h_X, 0) \quad \text{and} \quad M_Y(w)h_Y := M(w)(0, h_Y)$$

for all $h = (h_X, h_Y) \in W$.

Remark. The boundedness of the projected differentials follows from the Standing Assumption.

Definition 6. For normed linear spaces X, Y, Z and linear mappings $A \in \mathcal{L}(X, Z)$ and $B \in \mathcal{L}(Y, Z)$ we write (A, B) to denote the linear mapping $M \in \mathcal{L}(X \times Y, Z)$ given by $M(h_X, h_Y) := Ah_X + Bh_Y$ for all $(h_X, h_Y) \in X \times Y$.

Remark. The projected differentials satisfy $M(w) = (M_X(w), M_Y(w))$ by definition.

Definition 7. Under the assumptions of Definition 5 we set

$$\pi_X \partial F(w) := \{A \in \mathcal{L}(X,Z) : \ \exists B \in \mathcal{L}(Y,Z) : \ (A,B) \in \partial F(w)\};$$

analogously for $\pi_Y \partial F(w)$.

Generalized semismoothness in two variables simultaneously implies generalized semismoothness in each of these two variables separately.

Lemma 1. *Let X, Y, Z be normed linear spaces and $W := X \times Y$. Let $\bar{x} \in N_X \subset X$ and $\bar{y} \in N_Y \subset Y$. Set $N_W := N_X \times N_Y$ and consider a mapping $F : N_W \to Z$ that is generalized semismooth at $\bar{w} := (\bar{x}, \bar{y})$ (with order α) with respect to ∂F. Then $g : N_X \to Z$, $g(x) := F(x, \bar{y})$ is generalized semismooth at $\bar{x}$ (with order α) with respect to $\partial g(x) := \pi_X \partial F(x, \bar{y})$ and $h : N_Y \to Z$, $h(y) := F(\bar{x}, y)$ is generalized semismooth at $\bar{y}$ (with order α) with respect to $\partial h(y) := \pi_Y \partial F(\bar{x}, y)$.*

Proof. It follows from the generalized semismoothness of F at $\bar{w}$ (with order α) with respect to ∂F that g is generalized semismooth at $\bar{x}$ (with order α) with respect to ∂g as defined in the statement of the lemma, provided there is $C_X > 0$ such that $\|(x,0)\|_W \leq C_X \|x\|_X$ holds for all $x \in X$. Analogously, h is generalized semismooth at $\bar{y}$ (with order α) with respect to ∂h, provided there is $C_Y > 0$ such that $\|(0,y)\|_W \leq C_Y \|y\|_Y$ holds for all $y \in Y$. Due to the Standing Assumption this is satisfied with $C_X = C_Y = C_W$. □

Remark. Lemma 1 shows that generalized semismoothness implies partial generalized semismoothness with respect to the corresponding projected differential. However, already the single-valued case of Fréchet differentiability shows that partial generalized semismoothness does generally not imply generalized semismoothness. Moreover, depending on the definition of the generalized derivative at hand it may or may not be true that the projected generalized derivative and the generalized partial derivative coincide.

Theorem 2 has the following consequence.

Corollary 1 (Generalized semismoothness of implicit functions II)**.** *Let X, Y, Z be normed linear spaces and $W := X \times Y$. Let $N_X \subset X$, $N_W \subset W$, $g : N_X \to Y$ and $F : N_W \to Z$ be given. Let $\bar{w} := (\bar{x}, \bar{y}) \in N_W$ with $\bar{x} \in N_X$. Suppose that*

- *$g(\bar{x}) = \bar{y}$ as well as $(x, g(x)) \in N_W$ and $F(x, g(x)) = 0$ for all $x \in N_X$;*
- *F is generalized semismooth (of order α) at $\bar{w}$ with respect to ∂F;*
- *there exists $C > 0$ such that for all $w \in N_W \setminus \{\bar{w}\}$ every $B \in \pi_Y \partial F(w)$ is invertible with $\|B^{-1}\|_{\mathcal{L}(Z,Y)} \leq C$.*

Define the generalized derivative $\partial g : N_X \rightrightarrows \mathcal{L}(X,Y)$ by

$$\partial g(x) := \left\{ M \in \mathcal{L}(X,Y) : M = -B^{-1}A, \ \text{where} \ (A,B) \in \partial F(x, g(x)) \right\}$$

for every $x \in N_X \setminus \{\bar{x}\}$ and arbitrarily for $x = \bar{x}$. Assume that

- *there exists $L > 0$ such that either*

$$\|g(x) - g(\bar{x})\|_Y \le L\,\|x - \bar{x}\|_X \qquad \text{or} \qquad \sup_{M \in \partial g(x)} \|M(x - \bar{x})\|_Y \le L\,\|x - \bar{x}\|_X$$

holds for all $x \in N_X$. In the second case assume additionally that g is continuous at $\bar{x}$.

Then g is generalized semismooth (of order α) at $\bar{x}$ with respect to ∂g.

Moreover, if N_X is a neighborhood of $\bar{x}$ and g is continuous in all of N_X, then g is semismooth (of order $\tilde{\alpha}$) at $\bar{x}$ in the sense of Ulbrich, where $\tilde{\alpha} := \min\{\alpha, 1\}$.

Proof. It is enough to establish the generalized semismoothness (of order α) of g. To this end, define $\tilde{N}_W := \{(x, g(x)) \in N_W : x \in N_X\} \subset N_W$ and $w : N_X \to \tilde{N}_W$, $w(x) := (x, g(x))$. Applying Theorem 2 with $\tilde{N}_W$ and using the definition of $\tilde{N}_W$ it suffices to argue that

$$G : \tilde{N}_W \rightrightarrows \mathcal{L}(X, Y), \quad G(w) := \Big\{ M \in \mathcal{L}(X, Y) : M = -B^{-1}A, \text{ where } \ (A, B) \in \partial F(w) \Big\},$$

with $G(\bar{w})$ defined arbitrarily, is uniformly consistent (of order α) at $\bar{w}$. We argue only uniform consistency, since uniform consistency of order α can be established analogously. Thus, we have to demonstrate

$$\lim_{\substack{w \to \bar{w},\\ w \subset \tilde{N}_W \setminus \{\bar{w}\}}} \frac{\sup_{M \in G(w)} \|M(x - \bar{x}) - (y - \bar{y})\|_Y}{\|w - \bar{w}\|_W} = 0.$$

The definition of $\tilde{N}_W$, $g(\bar{x}) = \bar{y}$, and the continuity of g at $\bar{x}$ imply that this is the same as to show

$$\lim_{\substack{x \to \bar{x},\\ x \subset N_X \setminus \{\bar{x}\}}} \frac{\sup_{M \in G(w(x))} \|M(x - \bar{x}) - (g(x) - g(\bar{x}))\|_Y}{\|w(x) - \bar{w}\|_W} = 0.$$

The definition of G, $F(w(x)) = F(\bar{w}) = 0$, and the uniform boundedness of B^{-1} for all $B \in \pi_Y \partial F(w)$ with $w \in \tilde{N}_W \setminus \{\bar{w}\}$ imply that for all $x \in N_X \setminus \{\bar{x}\}$ there holds

$$\begin{aligned}
&\sup_{M \in G(w(x))} \frac{\|M(x - \bar{x}) - (g(x) - g(\bar{x}))\|_Y}{\|w(x) - \bar{w}\|_W} \\
&= \sup_{(A,B) \in \partial F(w(x))} \frac{\Big\|B^{-1}\Big[-A(x - \bar{x}) - B(g(x) - g(\bar{x}))\Big]\Big\|_Y}{\|w(x) - \bar{w}\|_W} \\
&= \sup_{(A,B) \in \partial F(w(x))} \frac{\Big\|B^{-1}\Big[F(w(x)) - F(\bar{w}) - A(x - \bar{x}) - B(g(x) - g(\bar{x}))\Big]\Big\|_Y}{\|w(x) - \bar{w}\|_W} \\
&\le C \sup_{M \in \partial F(w(x))} \frac{\|F(w(x)) - F(\bar{w}) - M(w(x) - \bar{w})\|_Z}{\|w(x) - \bar{w}\|_W}.
\end{aligned}$$

where we used that $A(x - \bar{x}) + B(g(x) - g(\bar{x})) = M(w(x) - \bar{w})$ for $M = (A, B) \in \partial F(w(x))$. Taking the limit for $N_X \ni x \to \bar{x}$ we obtain the consistency of G at $\bar{w}$ from the semismoothness of F at $\bar{w}$. □

Remark. 1) If N_X and N_W are neighborhoods of $\bar{x}$ and $\bar{w}$, respectively, then the requirement $(x, g(x)) \in N_W$ can be dropped, as it already follows from the continuity of g at $\bar{x}$ (at least after replacing N_X by a smaller neighborhood of $\bar{x}$).

2) Let us point out that if N_W is a neighborhood of $\bar{w}$ and F is ∂F-semismooth (of order $\alpha \in (0,1]$) at $\bar{w}$ in the sense of Ulbrich, then F is, in particular, generalized semismooth (of order α) at $\bar{w}$ with respect to ∂F. That is, the semismoothness requirement on F in the above corollary is satisfied in this case.

3) Notice that to achieve semismoothness in the sense of Ulbrich it suffices to have generalized semismoothness at $\bar{w}$ of F *restricted to* $\{(x, g(x)) \in W : x \in N_X\}$. In general, this is a significantly smaller set than a full neighborhood of $\bar{w}$. In particular, it is enough if the restriction of F to the set $\{w \in W : F(w) = 0\}$ is generalized semismooth at $\bar{w}$.

4) In fact, the observation of 3) allows us to slightly sharpen the existing finite-dimensional semismooth implicit function theorems. However, since our focus is on the infinite-dimensional case, we do not work this out in more detail here.

5) If there exists $\hat{C} > 0$ such that for all $w \in N_W \setminus \{\bar{w}\}$ there holds $\|A\|_{\mathcal{L}(X,Z)} \le \hat{C}$ for every $A \in \pi_X \partial F(w)$, then the uniform invertibility assumption on the elements of $\pi_Y \partial F(w)$ implies that there is $L > 0$ with $\sup_{M \in \partial g(x)} \|M(x - \bar{x})\|_Y \le L \|x - \bar{x}\|_X$ for all $x \in N_X$.

6) Let us apply Corollary 1 to the case where F is Fréchet differentiable at $\bar{w}$. This is possible since F is generalized semismooth in this case, cf. Example 1. We can use Lemma 1 to infer that in this setting there hold $\pi_X \partial F(w) = \{\partial_x F(\bar{w})\}$ and $\pi_Y \partial F(w) = \{\partial_y F(\bar{w})\}$ for all $w \in N_W$, where ∂_x and ∂_y denote the usual partial derivatives. Using 5) we deduce that in this setting Corollary 1 states that if $\partial_y F(\bar{w})$ is invertible and g is continuous at $\bar{x}$, then g is Fréchet differentiable at $\bar{x}$.

7) It follows from Banach's lemma, cf. [Cia13, Theorem 3.6-3], that if $w \mapsto \pi_Y \partial F(w)$ is upper semicontinuous at $\bar{w}$ and Y is a Banach space, then the uniform invertibility assumption on $B \in \pi_Y \partial F(w)$ can be replaced by requiring only the existence of $C > 0$ such that every $B \in \pi_Y \partial F(\bar{w})$ is invertible with $\|B^{-1}\|_{\mathcal{L}(Z,Y)} \le C$. If $\pi_Y \partial F(\bar{w})$ is compact, then the existence of such a C follows already if every $B \in \pi_Y \partial F(\bar{w})$ is invertible. Since π_Y is continuous, we obtain that $w \mapsto \pi_Y \partial F(w)$ is in particular upper semicontinuous, respectively, compact-valued at $\bar{w}$ if $w \mapsto \partial F(w)$ has this property. Let us, however, stress that while several generalized derivatives in finite dimensions are indeed compact-valued and upper semicontinuous, neither of these properties is usually satisfied in infinite-dimensional settings.

8) Theorem 2 and Corollary 1 do neither require nor assert that g is locally unique. In fact, they can be applied even in situations where g is not locally unique, as Example 2 below shows. Since most implicit function theorems guarantee existence and smoothness of a locally unique implicit function, it is clear that the prerequisites of such theorems cannot be satisfied in situations where the implicit function is not locally unique. Therefore, in these situations it is not possible to obtain smoothness results from standard implicit function theorems, not even if the existence of a (non-unique) implicit function is already known. In this respect Theorem 2 and Corollary 1 also extend finite-dimensional results.

9) We point out that Theorem 2 and Corollary 1 are more flexible in the choice of the generalized derivative for ∂F than most of the existing results on semismooth implicit functions. Example 3 presents a situation where this is beneficial.

10) Results that demonstrate how to obtain continuity, calmness, or Lipschitz continuity of g are contained in Section 3.

The following example shows, in particular, that Corollary 1 can be applied in situations where the implicit function is locally not unique. Let us stress that this example is chosen in such a way that we can easily check that the conclusions obtained from Corollary 1 are true.

Example 2. Using the notation of Corollary 1 let $X, Y, Z = \mathbb{R}$ and set $W := \mathbb{R}^2$. We endow all these spaces with the Euclidean norm. Consider $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) := |x| - |y|$. Since F is even semismooth in the sense of Definition 1 at every $w \in \mathbb{R}^2$, it is in particular generalized semismooth at every $w \in \mathbb{R}^2$. Clarke's generalized Jacobian of F at $w = (x,y)$ is given by $\partial^{\mathrm{Cl}} F(x,y) = (\widetilde{\operatorname{sign}}(x), -\widetilde{\operatorname{sign}}(y))^T$, where

$$\widetilde{\operatorname{sign}} : \mathbb{R} \rightrightarrows \mathbb{R}, \qquad \widetilde{\operatorname{sign}}(t) := \begin{cases} -1 : & \text{if } \ t<0, \\ [-1,1] : & \text{if } \ t=0, \\ +1 : & \text{if } \ t>0. \end{cases}$$

Hence, $\pi_Y \partial^{\mathrm{Cl}} F(x,y) \in \{\pm 1\}$ if $y \neq 0$. Since our focus is on the case where the implicit function is not locally unique, we consider $\bar{w} := (\bar{x}, \bar{y}) := (0,0)$. But let us mention that $\bar{w} \neq (0,0)$ with $F(\bar{w}) = 0$ can also be treated by Corollary 1, and that in this case locally unique existence of an implicit function can be established, too, for instance by using Corollary 2 from Section 3. Using $N_W := \{(x,y) \in \mathbb{R}^2 : y \neq 0\} \cup \{\bar{w}\}$ we obtain $\{w \in W : F(w) = 0\} \subset N_W$ and $\|B^{-1}\|_{\mathcal{L}(Z,Y)} = 1$ for every $B \in \pi_Y \partial^{\mathrm{Cl}} F(w) \in \{\pm 1\}$ with $w \in N_W \setminus \{\bar{w}\}$. Note it is crucial here that $\bar{w}$ can be excluded. Let $N_X \subset \mathbb{R}$ be a set that contains $\bar{x} = 0$. Then Corollary 1 implies that any implicit function $g : N_X \to Y$ with $g(0) = 0$ is generalized semismooth at 0 provided it is continuous at 0, where we have used that any such function satisfies $\sup_{M \in \partial g(x)} \|M\|_{\mathcal{L}(X,Y)} \leq 1$ for all $x \in N_X \setminus \{\bar{x}\}$ due to $\|A\|_{\mathcal{L}(X,Z)} \leq 1$ for every $A \in \pi_X \partial^{\mathrm{Cl}} F(w)$ with $w \in W$. So far, this includes the possibility for g to be continuous *only* at $\bar{x}$, for instance if we consider $N_X = \mathbb{R}$ and

$$g(x) := \begin{cases} x : & \text{if } x \in \mathbb{Q}, \\ -x : & \text{otherwise.} \end{cases}$$

Corollary 1 implies further that if we choose for N_X a neighborhood of $\bar{x} = 0$ and g is continuous in N_X, then g is ∂g-semismooth at $\bar{x}$ in the sense of Ulbrich. We remark that there are implicit function theorems that assert the existence of an only continuous implicit function, cf., e.g., [Jit78, Theorem 2.1], [War78, Theorem 1] and [Kum80, Theorem 1.1]. The combination of such results with Corollary 1 yields a new class of semismooth implicit function theorems for functions which are semismooth in the sense of Ulbrich. This may be helpful even in finite-dimensional settings since semismoothness in the sense of Ulbrich still allows to show local q-superlinear convergence of Newton's method, cf. Theorem 1.

Remark. A similar finite-dimensional example shows that in contrast to Theorem 2 the conditions of Corollary 1 are only sufficient to obtain generalized semismoothness of the implicit function. In fact, consider $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) := x^2 - y^2$. This mapping has the same roots as F in the example above, but $\pi_Y \partial F(x,y) = -2y$ yields that the uniform invertibility assumption of Corollary 1 is violated close to $\bar{w} = (0,0)$.

The following example makes use of the flexibility of Corollary 1 with respect to the choice of the generalized derivative ∂F.

Example 3. Let $X, Y, Z = \mathbb{R}$, set $W := \mathbb{R}^2$, and endow these spaces with the Euclidean norm. Consider $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) := (x-y)^+$. Since F is even semismooth in the sense of Definition 1 at every $w \in \mathbb{R}^2$, it is in particular generalized semismooth at every

$w \in \mathbb{R}^2$. For $(x,y) \in N_W := \{(x,y) : y \le x\}$ we have $\partial^{\mathrm{Cl}} F(x,y) = (1,-1)^T$ if $y < x$ and $\partial^{\mathrm{Cl}} F(x,y) = \{(\beta,-\beta)^T : \beta \in [0,1]\}$ if $y = x$. Let $\bar{w} = (\bar{x},\bar{y}) \in N_W$. We now establish semismoothness in the sense of Ulbrich at $\bar{x}$ for the nowhere(!) unique implicit function $g(x) := x$ defined in $N_X := \mathbb{R}$. Using $\partial F(w) := \partial^{\mathrm{Cl}} F(w) \setminus \{(\beta,-\beta)^T : \beta \in [0,1/2]\}$ we obtain that F is generalized semismooth at $\bar{w}$ with respect to $\partial F \subset \partial^{\mathrm{Cl}} F$, $\pi_Y \partial F$ satisfies the required uniform invertibility assumption, and $\sup_{M \in \partial g(x)} \|M\|_{\mathcal{L}(X,Y)} = 1$ for all $x \in N_X$ due to $\partial g(x) = \{1\}$ for all $x \in N_X$. Since g is continuous in N_X, we obtain its ∂g-semismoothness in the sense of Ulbrich at $\bar{x}$ as claimed. In fact, since ∂g is constant, the definition of generalized semismoothness implies that g is Fréchet differentiable at $\bar{x}$ with $g'(\bar{x}) = 1$. By repeating the whole argument we infer that g is Fréchet differentiable in N_X with $g' \equiv 1$, which shows that g' is even continuously differentiable in N_X. (Let us mention that the continuous differentiability of g can also be established in a more direct fashion by using $N_W := \{(x,x) : x \in \mathbb{R}\}$ and noticing that $F(x,y) = 0 = x - y =: \tilde{F}(x,y)$ for all $(x,y) \in N_W$. Since $\tilde{F}$ is Fréchet differentiable, $F : N_W \to \mathbb{R}$ is generalized semismooth at $\bar{w}$ with respect to $\partial F(w) := \{\tilde{F}'(\bar{w})\} = \{(1,-1)\}$. From this it is not hard to deduce that g is Fréchet differentiable at $\bar{x}$ with $g'(\bar{x}) = 1$.)

Example 4. $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x,y) := x - y^3$ shows that even if F is arbitrarily smooth and g is locally unique, then g may fail to be generalized semismooth with respect to ∂g as defined in Corollary 1.

3 Existence of implicit functions

In this section we collect results that can be used to ensure existence of implicit functions in the presence of semismoothness. As in the preceding section the following considerations are valid in infinite-dimensional spaces.

Standing assumption. *In Section 3 we suppose, unless stated otherwise, that product spaces are endowed with the product topology.*

Remark. This assumption only makes sense if the underlying spaces are at least topological spaces. In statements where this is not satisfied but a product space appears anyway, this product space is only considered as a set.

3.1 Existence through monotonicity

We require the following concepts of monotonicity.

Definition 8. Let X be a set, Y be a real Hilbert space with scalar product $(\cdot,\cdot)_Y$, and $W := X \times Y$. Let $N_W \subset W$ and $F : N_W \to Y$, $(x,y) \mapsto F(x,y)$ be given. Then F is called *coercive with respect to y uniformly in x* iff there exists an increasing function $\alpha : [0,\infty) \to [0,\infty)$ with

$$(F(x,y_1) - F(x,y_2), y_1 - y_2)_Y \ge \alpha(\|y_1 - y_2\|_Y) \, \|y_1 - y_2\|_Y$$

for all $(x,y_1),(x,y_2) \in N_W$.

Moreover, iff there even exists $\alpha > 0$ with

$$(F(x,y_1) - F(x,y_2), y_1 - y_2)_Y \ge \alpha \, \|y_1 - y_2\|_Y^2$$

for all $(x,y_1),(x,y_2) \in N_W$, then F is called *strongly monotone with respect to y uniformly in x*.

Definition 9. Let X, Y be metric spaces and $W := X \times Y$. Let $N_W \subset W$, and $F : N_W \to Y$, $(x, y) \mapsto F(x, y)$ be given. Then F is called *Lipschitz continuous with respect to x uniformly in y* iff there exists $L > 0$ with

$$d_Y(F(x_1, y), F(x_2, y)) \leq L d_X(x_1, x_2)$$

for all $(x_1, y), (x_2, y) \in N_W$.

Remark. The topology on W is arbitrary in Definition 9.

In Section 2 we have seen that calmness of an implicit function can be used to obtain semismoothness. The following relaxed versions of the above definitions are tailored to yield calm implicit functions.

Definition 10. Let X be a set, Y be a real Hilbert space with scalar product $(\cdot, \cdot)_Y$, and $W := X \times Y$. Let $(\bar{x}, \bar{y}) \in N_W \subset W$ and $F : N_W \to Y$, $(x, y) \mapsto F(x, y)$ be given. We say that $F(\bar{x}, \cdot)$ is *strongly monotone at* $\bar{y}$ iff there exists $\alpha > 0$ such that

$$(F(\bar{x}, y) - F(\bar{x}, \bar{y}), y - \bar{y})_Y \geq \alpha \|y - \bar{y}\|_Y^2$$

holds for all $y \in Y$ with $(\bar{x}, y) \in N_W$.

Definition 11. Let X, Y be metric spaces and $W := X \times Y$. Let $(\bar{x}, \bar{y}) \in N_W \subset W$ and $F : N_W \to Y$, $(x, y) \mapsto F(x, y)$ be given. We say that F is *calm at* $\bar{x}$ *uniformly in y* iff there exists $L > 0$ such that

$$d_Y(F(x, y), F(\bar{x}, y)) \leq L d_X(x, \bar{x})$$

holds for all $(x, y) \in N_W$ with $(\bar{x}, y) \in N_W$.

Remark. The topology on W is arbitrary in Definition 11.

The following theorem is a slightly sharpened version of a result due to Alt and Kolumbán. It ensures the existence of an implicit function in the presence of monotonicity.

Theorem 3 (Cf. Theorem 5.1 in [AK93])**.** *Let X be a topological space and Y be a real Hilbert space. Let $(\bar{x}, \bar{y}) \in X \times Y$ and a neighborhood N_Y of $\bar{y} \in Y$ be given. Suppose that the mapping $F : X \times N_Y \to Y$ satisfies*

- $F(\bar{x}, \bar{y}) = 0$*;*
- *F is continuous;*
- *F is coercive with respect to y uniformly in x.*

Then there exist neighborhoods $N_X \subset X$ of $\bar{x}$ and $\tilde{N}_Y \subset N_Y$ of $\bar{y}$ together with a unique mapping $g : N_X \to \tilde{N}_Y$ continuous in N_X such that $g(\bar{x}) = \bar{y}$ and $F(x, g(x)) = 0$ for all $x \in N_X$.

Proof. The main statement is Theorem 5.1 in [AK93], except that there only continuity of g at $\bar{x}$ is proven. To deduce the continuity of g in all of N_X, we may suppose without loss of generality that N_Y is open, hence N_Y is a neighborhood of every $y \in N_Y$. Without loss of generality we may, furthermore, assume that g as provided by Theorem 5.1 in [AK93] is uniquely determined in the *open* neighborhood N_X and continuous at $\bar{x}$. Thus, N_X is a neighborhood of every $x \in N_X$. We are now in the situation that Theorem 5.1 in [AK93] can be applied to $(x, g(x))$ for each $x \in N_X$. However, since g is uniquely determined in the open neighborhood N_X of $\bar{x}$, it follows that g is continuous at every $x \in N_X$. □

The following result can be used on its own or in combination with the preceding theorem to obtain calmness or Lipschitz continuity of an implicit function.

Lemma 2 (Cf. Corollary 5.2 in [AK93])**.** *Let X be a topological space, Y be a real Hilbert space, and $W := X \times Y$. Let $N_W \subset W$, $g : X \to Y$ and $F : N_W \to Y$ be given. Suppose that*

- *$g(\bar{x}) = \bar{y}$ as well as $(x, g(x)), (\bar{x}, g(x)) \in N_W$ and $F(x, g(x)) = 0$ for all $x \in X$;*
- *$F(\bar{x}, \cdot)$ is strongly monotone at $\bar{y}$ with constant α;*
- *$F(\cdot, y)$ is continuous for every fixed y.*

Then g is continuous.

If X is a metric space and instead of the continuity assumption on F there holds that F is calm at $\bar{x}$ uniformly in y with constant L, then g is calm at $\bar{x}$ with constant L/α.

If X is a metric space and F is even strongly monotone with respect to y uniformly in x with constant α as well as Lipschitz continuous with respect to x uniformly in y with constant L, then g is Lipschitz continuous with constant L/α.

Proof. The following proof is based on the proof of Corollary 5.2 in [AK93]. Let us demonstrate the calmness of g at $\bar{x}$. Using $F(\bar{x}, \bar{y}) = 0 = F(x, g(x))$ and the strong monotonicity of $F(\bar{x}, \cdot)$ at $\bar{y}$ as well as the Cauchy-Schwarz inequality we obtain for all $x \in X$

$$\alpha \|g(x) - \bar{y}\|_Y^2 \leq (F(\bar{x}, g(x)) - F(\bar{x}, \bar{y}), g(x) - \bar{y})_Y \leq \|F(\bar{x}, g(x)) - F(x, g(x))\|_Y \|g(x) - \bar{y}\|_Y ,$$

from which the assertion follows by the uniform calmness of F at $\bar{x}$. The continuity, respectively, Lipschitz continuity of g can be proven analogously. □

Remark. 1) If N_W is a neighborhood of $(\bar{x}, \bar{y})$ and g is continuous at $\bar{x}$, then the requirements $(x, g(x)), (\bar{x}, g(x)) \in N_W$ can be dropped, as they follow from the continuity of g at $\bar{x}$ (at least after replacing X by a smaller subset of X).

2) Note that N_W can be chosen as $\{(x, g(x)) \in W : x \in N_X$ with $F(x, g(x)) = 0\}$, which in general is significantly smaller than a full neighborhood of $(\bar{x}, \bar{y})$. In particular, it suffices in Lemma 2 if the restriction of F to the set $\{w \in W : F(w) = 0\}$ satisfies the assumptions imposed on F.

3.2 Existence through approximation

We discuss two existence results that do not rely on monotonicity but instead use certain approximation properties. To state these results conveniently we introduce some definitions.

Definition 12. Let V be a metric space and $\mathcal{A} \subset V$. The *measure of non-compactness of $\mathcal{A}$* is given by

$$\chi(\mathcal{A}) := \inf \left\{ r > 0 : \ \mathcal{A} \subset \bigcup_{i=1}^{m} \overline{B_r(a_i)} \text{ for some } m \in \mathbb{N} \text{ and } a_i \in \mathcal{A},\ 1 \leq i \leq m \right\},$$

where $\overline{B_r(v)}$ denotes the closed ball of radius $r > 0$ around v.

Remark. If $\mathcal{A}$ is compact, then there obviously holds $\chi(\mathcal{A}) = 0$.

Definition 13. Let X be a topological space and Y, Z be normed linear spaces. Let $(\bar{x}, \bar{y}) \in X \times Y$ and a neighborhood $N_Y \subset Y$ of $\bar{y}$ be given. Consider a mapping $F : X \times N_Y \to Z$ and a set $\mathcal{A}$ of homogeneous mappings from Y to Z. We say that $\mathcal{A}$ is a *uniform strict prederivative for F at $\bar{y}$ near $\bar{x}$* iff for every $\varepsilon > 0$ there exists a neighborhood $N_W \subset X \times N_Y$ of $(\bar{x}, \bar{y})$ such that

$$\inf_{A \in \mathcal{A}} \|F(x, y_2) - F(x, y_1) - A(y_2 - y_1)\|_Z \le \varepsilon \|y_2 - y_1\|_Y$$

is satisfied for all $(x, y_1), (x, y_2) \in N_W$.

Remark. The notion of a strict prederivative was introduced by Ioffe, cf. [Iof81].

We employ the following weakened concept of upper semicontinuity.

Definition 14. Let W be a topological space and V be a metric space. A set-valued mapping $H : W \rightrightarrows V$ is said to be *mildly upper semicontinuous at $\bar{w} \in W$* iff for every $\varepsilon > 0$ there exists a neighborhood N_W of $\bar{w}$ such that $H(w) \subset B_\varepsilon(H(\bar{w}))$ for every $w \in N_W$. Here, we have used the notation $B_\varepsilon(H(\bar{w})) := \cup_{v \in H(\bar{w})} B_\varepsilon(v)$ with $B_\varepsilon(v) := \{\hat{v} \in V : d_V(\hat{v}, v) < \varepsilon\}$.

Remark. Notice the difference to the definition of upper semicontinuity: Mild upper semicontinuity only requires $H(w) \subset B_\varepsilon(H(\bar{w}))$, while upper semicontinuity demands $H(w) \subset N_V$ with an *arbitrary* neighborhood N_V of $H(\bar{w})$. In particular, every upper semicontinuous mapping is mildly upper semicontinuous. The converse, however, is false as the examples $H_{1/2} : \mathbb{R} \rightrightarrows \mathbb{R}$ given by $H_1(t) := B_{|t|}(0)$ and $H_2(t) := \mathbb{N} + \overline{B_{|t|}(0)}$ show. For H_1 this is obvious since $H_1(t)$ is open for every t. For H_2 take $t = 0$ and consider the neighborhood $\cup_{m \in \mathbb{N}}(m + B_{1/m}(0))$ of $H_2(0) = \mathbb{N}$ to deduce that H_2 is not upper semicontinuous at $t = 0$. Furthermore, for single-valued mappings upper semicontinuity and mild upper semicontinuity coincide. More generally, if H is compact-valued at $\bar{w}$, then H is upper semicontinuous at $\bar{w}$ iff it is mildly upper semicontinuous at $\bar{w}$. In fact, without using the term mild upper semicontinuity this equivalence is already noted in [AF09, below Definition 1.4.1], but in the sequel only upper semicontinuity is investigated further there. In contrast to the compact-valued case we observe that for closed-valued mappings mild upper semicontinuity does not imply upper semicontinuity as H_2 shows.

Mild upper semicontinuity can be expressed through sequences.

Lemma 3. *Let V, W be metric spaces. A set-valued mapping $H : W \rightrightarrows V$ is mildly upper semicontinuous at $\bar{w} \in W$ iff for every sequence $(w_k) \subset W$ with $w_k \to \bar{w}$ the following is true: For every selection (H_k) of $(H(w_k))$ and every $\varepsilon > 0$ there exists $K \in \mathbb{N}$ such that for every $k \ge K$ there is $\bar{H}_k \in H(\bar{w})$ with $d_V(H_k, \bar{H}_k) < \varepsilon$. Here, we speak of a selection (H_k) of $(H(w_k))$ iff there holds $H_k \in H(w_k)$ for all $k \in \mathbb{N}$.*

Proof. Let H be mildly upper semicontinuous at $\bar{w}$. Let (w_k) be a sequence that converges to $\bar{w}$ and (H_k) be a selection of $(H(w_k))$. For $\varepsilon > 0$ there exists a neighborhood N_W of $\bar{w}$ such that $H(w) \subset N_V := B_\varepsilon(H(\bar{w}))$ for every $w \in N_W$. Since $w_k \to \bar{w}$, there exists $K \in \mathbb{N}$ such that $w_k \in N_W$ for every $k \ge K$. Hence, $\cup_{k \ge K} H(w_k) \subset N_V$, which implies $\cup_{k \ge K}\{H_k\} \subset N_V$. Thus, the definition of N_V yields that for every H_k with $k \ge K$ there is $\bar{H}_k \in H(\bar{w})$ such that $d_V(H_k, \bar{H}_k) < \varepsilon$.

Let us now assume that H is not mildly upper semicontinuous at $\bar{w}$. That is, there exists an $\varepsilon > 0$ such that for every neighborhood N_W of $\bar{w}$ there is $w \in N_W$ with $H(w) \not\subset N_V := B_\varepsilon(H(\bar{w}))$. Hence, we can inductively construct a sequence $(w_k) \subset W$ with $d_W(w_k, \bar{w}) \leq 1/k$ and $H(w_k) \not\subset N_V$ for every $k \in \mathbb{N}$. Therefore, there exists for every $k \in \mathbb{N}$ a $H_k \in H(w_k)$ that satisfies $H_k \notin N_V$. Since $N_V = B_\varepsilon(H(\bar{w}))$, this means that for every k there holds $d_V(H_k, \bar{H}) \geq \varepsilon$ for all $\bar{H} \in H(\bar{w})$. In conclusion, (w_k) converges to $\bar{w}$, but for the selection (H_k) of $(H(w_k))$ it is not possible to find $K \in \mathbb{N}$ such that for every $k \geq K$ there is $\bar{H}_k \in H(\bar{w})$ with $d_V(H_k, \bar{H}_k) < \varepsilon$. This completes the proof. □

The following result is due to Páles. It ensures the existence of an implicit function in the presence of a convex linear uniform strict prederivative.

Theorem 4 (Cf. Theorem 4 in [Pál97])**.** *Let X be a topological space and Y, Z be Banach spaces. Let $(\bar{x}, \bar{y}) \in X \times Y$ as well as a neighborhood N_Y of $\bar{y}$ be given. Suppose that the mapping $F : X \times N_Y \to Z$ satisfies*

- *$F(\bar{x}, \bar{y}) = 0$;*
- *$F(\cdot, \bar{y})$ is continuous at $\bar{x}$;*
- *there exists a convex set $\mathcal{A} \subset \mathcal{L}(Y, Z)$ of invertible operators and a constant $C > 0$ with $\sup_{A \in \mathcal{A}} \|A^{-1}\|_{\mathcal{L}(Z,Y)} \leq C$ and $\chi(\mathcal{A}) \cdot C < 1$;*
- *$\mathcal{A}$ is a uniform strict prederivative for F at $\bar{y}$ near $\bar{x}$.*

Then there exist neighborhoods $N_X \subset X$ of $\bar{x}$ and $\tilde{N}_Y \subset N_Y$ of $\bar{y}$ together with a unique mapping $g : N_X \to \tilde{N}_Y$ continuous at $\bar{x}$ such that $g(\bar{x}) = \bar{y}$ and $F(x, g(x)) = 0$ for all $x \in N_X$.

Moreover, if X is a metric space and $F(\cdot, \bar{y})$ is calm at $\bar{x}$, then g is calm at $\bar{x}$.

Proof. The main statement is [Pál97, Theorem 4] in combination with the preceding remark there. The additional part about calmness follows from the proof given there; it is a direct consequence of the estimate that is used to prove continuity of g at $\bar{x}$. □

Remark. The formulation of [Pál97, Theorem 4] does not use the neighborhood $\tilde{N}_Y$. However, the proof shows that uniqueness of g may only hold in a smaller neighborhood than N_Y. The necessity of $\tilde{N}_Y$ is, moreover, clear since the result includes the classical implicit function theorem.

From Theorem 4 we deduce a result for the special case that only X is infinite-dimensional. This case occurs in infinite-dimensional optimization problems with finite-rank constraints, e.g., optimal control problems with a constraint on the L^p-norm of the control or state (in addition to other constraints that do not have to be of finite rank).

Corollary 2. *Let X be a topological space and $Y = Z = \mathbb{R}^n$. Let $(\bar{x}, \bar{y}) \in X \times Y$ as well as a neighborhood N_Y of $\bar{y}$ be given. Suppose that the mapping $F : X \times N_Y \to Z$ satisfies*

- *$F(\bar{x}, \bar{y}) = 0$;*
- *$F(\cdot, \bar{y})$ is continuous at $\bar{x}$;*
- *$F(x, \cdot)$ is Lipschitz in N_Y for every $x \in X$;*

- *the mapping $H : X \times N_Y \rightrightarrows \mathbb{R}^{n\times n}$, $H(x,y) := \partial_y^{Cl} F(x,y)$ is mildly upper semicontinuous at $(\bar{x},\bar{y})$. Here, $\partial_y^{Cl} F(x,y)$ denotes Clarke's generalized Jacobian of $F(x,\cdot)$ at $y \in N_Y$;*
- *every $A \in \partial_y^{Cl} F(\bar{x},\bar{y})$ is invertible.*

Then there exist neighborhoods $N_X \subset X$ of $\bar{x}$ and $\tilde{N}_Y \subset N_Y$ of $\bar{y}$ together with a unique mapping $g : N_X \to \tilde{N}_Y$ continuous at $\bar{x}$ such that $g(\bar{x}) = \bar{y}$ and $F(x, g(x)) = 0$ for all $x \in N_X$.

Moreover, if X is a metric space and $F(\cdot,\bar{y})$ is calm at $\bar{x}$, then g is calm at $\bar{x}$.

Proof. We set $\bar{w} := (\bar{x},\bar{y})$ and check that the requirements of Theorem 4 are satisfied for $\mathcal{A} := \partial_y^{\mathrm{Cl}} F(\bar{w})$. Since the Clarke differential $\partial_y^{\mathrm{Cl}} F(\bar{w})$ is convex and compact, it follows that $\mathcal{A}$ is convex and that $\{A^{-1} \in \mathbb{R}^{n\times n} : A \in \mathcal{A}\}$ is compact (as image of a compact set under a continuous mapping). Thus, there exists $C > 0$ such that $\sup_{A\in\mathcal{A}} \|A^{-1}\|_{\mathcal{L}(Z,Y)} \leq C$. Moreover, the compactness of $\mathcal{A}$ implies $\chi(\mathcal{A}) = 0$. It remains to establish that $\mathcal{A}$ is a uniform strict prederivative for F at $\bar{y}$ near $\bar{x}$. To do so, let $\varepsilon > 0$ and consider a neighborhood $N_W = N_X \times B_\delta(\bar{y})$ of $\bar{w}$ such that for all $(x,y) \in N_W$ there holds $\partial_y^{\mathrm{Cl}} F(x,y) \subset \partial_y^{\mathrm{Cl}} F(\bar{w}) + B_\varepsilon(0)$. Such a neighborhood exists due to the mild upper semicontinuity of $\partial_y^{\mathrm{Cl}} F$ at $\bar{w}$. Let $x \in N_X$ and $y_1, y_2 \in B_\delta(\bar{y})$. Denoting by $\mathrm{co}(S)$ the convex hull of the set $S \subset \mathbb{R}^n$ and by $y(t)$ for $t \in [0,1]$ the point $y(t) := y_1 + t(y_2 - y_1)$, the mean-value theorem for Clarke's generalized Jacobian, cf. [Cla90, Proposition 2.6.5], yields

$$F(x,y_2) - F(x,y_1) \in \mathrm{co}\Big(\Big\{A(y_2 - y_1) \in \mathbb{R}^n : A \in \partial_y^{\mathrm{Cl}} F(x, y(t)) \text{ for a } t \in [0,1]\Big\}\Big).$$

Hence, there are $t_1, t_2, \ldots, t_m \in [0,1]$ and $A_1, \ldots, A_m$ with $A_i \in \partial_y^{\mathrm{Cl}} F(x, y(t_i))$ for $1 \leq i \leq m$, as well as $\lambda_1, \ldots, \lambda_m \in [0,1]$ with $\sum_{i=1}^m \lambda_i = 1$ such that

$$F(x,y_2) - F(x,y_1) = \sum_{i=1}^m \lambda_i A_i (y_2 - y_1)$$

is satisfied. Since $\partial_y^{\mathrm{Cl}} F(x,y) \subset \partial_y^{\mathrm{Cl}} F(\bar{w}) + B_\varepsilon(0)$ holds for all $(x,y) \in N_W$, there exist $\bar{A}_1, \ldots, \bar{A}_m \in \mathcal{A}$ with $\|A_i - \bar{A}_i\|_{\mathcal{L}(Y,Z)} < \varepsilon$ for $1 \leq i \leq m$. The convexity of $\mathcal{A}$ implies that $A := \sum_{i=1}^m \lambda_i \bar{A}_i \in \mathcal{A}$. Thus, we can conclude

$$\|F(x,y_2) - F(x,y_1) - A(y_2 - y_1)\|_Z < \varepsilon \, \|y_2 - y_1\|_Y \,,$$

from which the assertion follows. ☐

Remark. In finite-dimensional implicit function theorems where F is Lipschitz continuous in *both* variables and Clarke's generalized Jacobian is used, it is often required that every $A \in \pi_Y \partial^{\mathrm{Cl}} F(\bar{w})$ be invertible. Since the definitions imply $\partial_y^{\mathrm{Cl}} F(\bar{w}) \subset \pi_Y \partial^{\mathrm{Cl}} F(\bar{w})$, the assumption in Corollary 2 is weaker. However, it should be noted that in the finite-dimensional setting with Lipschitz continuous F a more general implicit function theorem can be derived based on Kummer's implicit function theorem; cf. [Kum91, Theorem 1] for Kummer's theorem and [MSZ05, Corollary 2] for its application to semismoothness.

The following lemma establishes mild upper semicontinuity for a class of superposition operators that frequently occur in optimal control with partial differential equations. In particular, there are situations in which the mild upper semicontinuity required in Corollary 2 can be shown by use of this lemma, cf. Example 5.

Lemma 4. *Let $m \in \mathbb{N}$ and $1 \le q \le p \le \infty$. Let $\Omega \subset \mathbb{R}^n$ be a Lebesgue measurable set with $0 < |\Omega| < \infty$ and define*

$$P := \prod_{i=1}^{m} L^p(\Omega), \qquad Q := \prod_{i=1}^{m} L^q(\Omega), \qquad \text{and} \qquad \hat{Q} := L^q(\Omega).$$

Endow P, Q and $L^1(\Omega)^m$ with the norms $\|v\|_P := \sum_{i=1}^m \|v_i\|_{L^p(\Omega)}$, $\|v\|_Q := \sum_{i=1}^m \|v_i\|_{L^q(\Omega)}$, and $\|v\|_1 := \sum_{i=1}^m \|v_i\|_{L^1(\Omega)}$, respectively. Let X be a metric space and Y be a real Banach space. Let $(\bar{x}, \bar{y}) \in X \times Y$ as well as an open neighborhood N_Y of $\bar{y}$ be given. Suppose that the mappings $G : X \times N_Y \to Q$ and $\psi : \mathbb{R}^m \to \mathbb{R}$ satisfy

- *$G(x, \cdot) : N_Y \to Q$ is Gâteaux differentiable for every fixed $x \in X$;*
- *$G_y : X \times N_Y \to \mathcal{L}(Y, Q)$ is continuous at $(\bar{x}, \bar{y})$;*
- *ψ is Lipschitz continuous with modulus $L_\psi > 0$.*

Furthermore, for $v = (v_i)_{i=1}^m \in L^1(\Omega)^m$ and $s \in \Omega$ let us denote

$$v(s) := \big(v_1(s), v_2(s), \dots, v_m(s)\big)^T.$$

We consider the superposition operator

$$\Psi : X \times N_Y \to \hat{Q}, \qquad \Psi(x, y)(s) := \psi(G(x, y)(s))$$

together with Ulbrich's generalized derivative $\partial_y^{Ul}\Psi : X \times N_Y \rightrightarrows \mathcal{L}(Y, \hat{Q})$ of Ψ with respect to y. That is,

$$\partial_y^{Ul}\Psi(x, y) := \Big\{ M : \ Mh(s) = g(s)^T \big(G_y(x, y)h(s)\big), \ g \in L^\infty(\Omega)^m, \\ g(s) \in \partial^{Cl}\psi(G(x, y)(s)) \text{ for a.a. } s \in \Omega \Big\}.$$

Then $(x, y) \mapsto \partial_y^{Ul}\Psi(x, y)$ is mildly upper semicontinuous at $(\bar{x}, \bar{y})$ provided that either

- *$p = \infty$, G maps to $P = L^\infty(\Omega)^m$, and $G : X \times N_Y \to P$ is continuous at $(\bar{x}, \bar{y})$*

  or

- *$p > q$, $G_y(\bar{x}, \bar{y}) \in \mathcal{L}(Y, P)$, and $G : X \times N_Y \to L^1(\Omega)^m$ is continuous at $(\bar{x}, \bar{y})$.*

Proof. To establish the assertion we use the characterization of mild upper semicontinuity based on sequences, cf. Lemma 3. To this end, let $(w_k) \subset X \times N_Y$ be a sequence with $w_k \to \bar{w} := (\bar{x}, \bar{y})$. For every $k \in \mathbb{N}$ select an $M_k \in \partial_y^{\mathrm{Ul}}\Psi(w_k)$. Let $\varepsilon > 0$. We have to show that there exists $K \in \mathbb{N}$ and a sequence $(\bar{M}_k) \subset \partial_y^{\mathrm{Ul}}\Psi(\bar{w})$ such that $\|M_k - \bar{M}_k\|_{\mathcal{L}(Y,\hat{Q})} < \varepsilon$ for all $k \ge K$. The proof simplifies in the case $p = \infty$, so let us deal with $p < \infty$ first. In this case we have $p - q > 0$ by assumption. As ψ is Lipschitz with rank $L_\psi > 0$ and $G_y(\bar{x}, \bar{y}) \in \mathcal{L}(Y, P)$ holds, we can introduce the positive numbers $\alpha := \max\{L_\psi, \|G_y(\bar{w})\|_{\mathcal{L}(Y,P)}\}$ and $\beta := pq/(p - q)$. Now we define $\hat{\varepsilon} := [\varepsilon/(8\alpha^2)]^\beta$, $\bar{v} := G(\bar{w})$, and $v_k := G(w_k)$ for all $k \in \mathbb{N}$. Since $\bar{v} \in Q$, there are $C > 0$ and $\Omega_{\hat{\varepsilon}} \subset \Omega$ with $|\Omega_{\hat{\varepsilon}}| < \hat{\varepsilon}/2$ such that $\bar{v}(s) \in [-C, C]^m$ for a.e. $s \in \Omega \setminus \Omega_{\hat{\varepsilon}}$. This follows by considering for fixed $1 \le i \le m$ and $j \to \infty$ the sets $\Omega_j^i := \{s \in \Omega : |\bar{v}_i(s)| \le j\}$. As $\partial^{\mathrm{Cl}}\psi$ is upper semicontinuous, in particular mildly upper semicontinuous, we infer that for every $a \in [-C, C]^m$ there is $\delta_a > 0$ such that $\partial^{\mathrm{Cl}}\psi(\tilde{a}) \subset \partial^{\mathrm{Cl}}\psi(a) + B_{\tilde{\varepsilon}}^{\mathbb{R}^m}(0)$ is satisfied for all

$\tilde a \in B_{\delta_a}(a)$, where $\tilde\varepsilon := \frac{\varepsilon}{1+4\|G_y(\bar w)\|_{\mathcal{L}(Y,Q)}}$. Since $[-C,C]^m$ is compact, we obtain the existence of $\delta > 0$ such that for all $a \in [-C,C]^m$ there holds $\partial^{\mathrm{Cl}}\psi(\tilde a) \subset \partial^{\mathrm{Cl}}\psi(a) + B_{\tilde\varepsilon}^{\mathbb{R}^m}(0)$ for all $\tilde a \in B_\delta(a)$. The continuity of G at $\bar w$ yields $\|v_k - \bar v\|_1 \to 0$ for $k \to \infty$. Hence, there exists $K_1 \in \mathbb{N}$ and for every $k \geq K_1$ a set $\omega_k \subset \Omega$ with $|\omega_k| < \hat\varepsilon/2$ such that $\|v_k - \bar v\|_{L^\infty(\Omega\setminus\omega_k)^m} < \delta$ is satisfied for every $k \geq K_1$, where we used $\|v\|_{L^\infty(\omega)^m} := \sum_{i=1}^m \|v_i\|_{L^\infty(\omega)}$ as norm on $L^\infty(\omega)^m$ for every $\omega \subset \Omega$. Setting $\Omega_k := \Omega_{\hat\varepsilon} \cup \omega_k$ we have established that there hold

$$|\Omega_k|^{\frac{1}{\beta}} < \hat\varepsilon^{\frac{1}{\beta}} = \frac{\varepsilon}{8\alpha^2} \qquad \text{and} \qquad \partial^{\mathrm{Cl}}\psi(v_k(s)) \subset \partial^{\mathrm{Cl}}\psi(\bar v(s)) + B_{\tilde\varepsilon}^{\mathbb{R}^m}(0) \tag{3}$$

for all $k \geq K_1$ and almost all $s \in \Omega \setminus \Omega_k$. The definition of Ulbrich's generalized derivative implies that every M_k has the form $M_k = g_k^T G_y(w_k)$ with a $g_k \in L^\infty(\Omega)^m$ that fulfills $g_k(s) \in \partial^{\mathrm{Cl}}\psi(v_k(s))$ for a.e. $s \in \Omega$. The statement on the right in (3) therefore yields that for every $k \geq K_1$ there is $\bar g_k \in L^\infty(\Omega)^m$ with $\bar g_k(s) \in \partial^{\mathrm{Cl}}\psi(\bar v(s))$ for a.e. $s \in \Omega$ such that $\|g_k - \bar g_k\|_{L^\infty(\Omega\setminus\Omega_k)^m} < \tilde\varepsilon$. For every $k \geq K_1$ let us define $\bar M_k := \bar g_k^T G_y(\bar w)$. Evidently, we have $\bar M_k \in \partial_y^{\mathrm{Ul}}\Psi(\bar w)$ for all $k \geq K_1$. Moreover, we have for these k

$$\begin{aligned} \left\|M_k - \bar M_k\right\|_{\mathcal{L}(Y,\hat Q)} &= \left\|g_k^T G_y(w_k) - \bar g_k^T G_y(\bar w)\right\|_{\mathcal{L}(Y,\hat Q)} \\ &\leq \left\|g_k^T(G_y(w_k) - G_y(\bar w))\right\|_{\mathcal{L}(Y,\hat Q)} + \left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,\hat Q)}. \end{aligned} \tag{4}$$

Since $g_k(s) \in \partial^{\mathrm{Cl}}\psi(v_k(s))$ for a.e. $s \in \Omega$, it follows that $g_k(s) \in [-L_\psi, L_\psi]^m$ for a.e. $s \in \Omega$ and all $k \in \mathbb{N}$. Hence, the first term on the right-hand side in (4) is for all $k \geq K_2$ smaller than $\varepsilon/2$, provided $K_2 \in \mathbb{N}$ is chosen such that $\|G_y(w_k) - G_y(\bar w)\|_{\mathcal{L}(Y,Q)} < \varepsilon/(2L_\psi)$ for all $k \geq K_2$. Such a K_2 exists due to the continuity of G_y at $\bar w$. In view of (4) it suffices to demonstrate that $\|(g_k - \bar g_k)^T G_y(\bar w)\|_{\mathcal{L}(Y,\hat Q)} \leq \varepsilon/2$ is satisfied for all $k \geq K_1$ to complete the proof. Obviously, there holds

$$\left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,\hat Q)} \leq \left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,L^q(\Omega\setminus\Omega_k))} + \left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,L^q(\Omega_k))}$$

for all $k \geq K_1$. To estimate the left summand we recall that $\|g_k - \bar g_k\|_{L^\infty(\Omega\setminus\Omega_k)^m} < \tilde\varepsilon$ is true for all $k \geq K_1$. Hence, the definition of $\tilde\varepsilon$ leads to

$$\left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,L^q(\Omega\setminus\Omega_k))} \leq \|g_k - \bar g_k\|_{L^\infty(\Omega\setminus\Omega_k)^m} \|G_y(\bar w)\|_{\mathcal{L}(Y,L^q(\Omega\setminus\Omega_k)^m)} < \frac{\varepsilon}{4}$$

for these k. For the right summand we employ $\partial^{\mathrm{Cl}}\psi \subset [-L_\psi, L_\psi]^m$, Hölder's inequality, and (3) to arrive at

$$\begin{aligned} \left\|(g_k - \bar g_k)^T G_y(\bar w)\right\|_{\mathcal{L}(Y,L^q(\Omega_k))} &\leq 2L_\psi \|G_y(\bar w)\|_{\mathcal{L}(Y,L^q(\Omega_k)^m)} \\ &\leq 2L_\psi |\Omega_k|^{\frac{p-q}{pq}} \|G_y(\bar w)\|_{\mathcal{L}(Y,L^p(\Omega_k)^m)} \leq \frac{\varepsilon L_\psi \|G_y(\bar w)\|_{\mathcal{L}(Y,P)}}{4\alpha^2} \leq \frac{\varepsilon}{4} \end{aligned}$$

for all $k \geq K_1$, where we have used $(p-q)/pq = 1/\beta$. This finishes the proof for $p < \infty$.

The case $p = \infty$ can be handled analogously. In fact, the above proof can be copied almost verbatim if the following changes are made: $\alpha := L_\psi$, $\beta := 1$, $\Omega_{\hat\varepsilon} := \emptyset$, $\Omega_k := \omega_k := \emptyset$ for all $k \in \mathbb{N}$. □

Remark. 1) An inspection of the proof of Lemma 4 reveals that in the case $p = \infty$ it is enough to require that ψ be *locally* Lipschitz (since this implies that ψ is Lipschitz on every compact set).

2) The assumptions on G imply that $y \mapsto G(\bar{x}, y) \in Q$ is Fréchet differentiable at $y = \bar{y}$.

3) Ulbrich's generalized derivative for superposition operators is treated in [Ulb11, Section 3]. In particular, its semismoothness is analyzed there. A different approach to semismoothness of superposition operators can be found in [Sch08]. The latter reference shows that the assumption that ψ is Lipschitz continuous may be weakened.

4) Our requirements on Ω, $G(x, \cdot)$, G_y and ψ are weaker than the ones used by Ulbrich in [Ulb11, Section 3.3.1]. However, Ulbrich's generalized derivative for Ψ is still well-defined under the weaker assumptions that we use. This follows by inspection of the results in [Ulb11, Section 3.3.2] (note that our p corresponds to r_i there and our q corresponds to r, while we do not require the q_i that are used there).

5) It is easy to see that if $G : W \rightrightarrows V$ is mildly upper semicontinuous and $f : V \to Z$ is uniformly continuous (e.g. Lipschitz continuous), then $f \circ G : W \rightrightarrows Z$, $w \mapsto \cup_{v \in G(w)} f(v)$ is mildly upper semicontinuous. This observation can be helpful to establish mild upper semicontinuity of composed mappings containing a superposition operator; cf. Example 5.

The following example presents a scenario in which Lemma 4 can be used to establish the mild upper semicontinuity required in Corollary 2.

Example 5. In the notation of Lemma 4 we take $m = 1$, $q \in [1, \infty]$, $p > q$ if $q < \infty$ and $p = \infty$ if $q = \infty$, $X = L^p(\Omega)$ and $Y = \mathbb{R}$ with arbitrary $(\bar{x}, \bar{y})$ in what follows. Let $F : L^p(\Omega) \times \mathbb{R} \to \mathbb{R}$, $F(x, y) := \int_\Omega |x(s) - y| \, \mathrm{d}s = \int_\Omega \Psi(x, y) \, \mathrm{d}s$, where $\Psi(x, y)(s) = \psi(x(s) - \iota y)$ with $\psi(t) = |t|$ and $\iota : \mathbb{R} \to L^p(\Omega)$ denoting the embedding of $\mathbb{R}$ into $L^p(\Omega)$ given by $\iota(t)(s) := t$ for every $s \in \Omega$. Then $G : L^p(\Omega) \times \mathbb{R} \to L^p(\Omega)$, $G(x, y) = x - \iota y$ is continuous, continuously differentiable with respect to y for every fixed x, and its derivative $G_y(x, y) = -\iota \in \mathcal{L}(\mathbb{R}, L^p(\Omega))$ is continuous with respect to (x, y). Thus, the prerequisites of Lemma 4 are satisfied. (In fact, the assumptions of Lemma 4 are weaker since in that lemma it is, for instance, sufficient to consider G as a mapping to $L^q(\Omega)$ instead of $L^p(\Omega)$ if $p < \infty$.) In conclusion, $\partial_y^{\mathrm{Ul}} \Psi$ is mildly upper semicontinuous as a mapping to $\mathcal{L}(\mathbb{R}, L^q(\Omega))$. Hence, $f : L^p(\Omega) \times \mathbb{R} \rightrightarrows \mathcal{L}(\mathbb{R}, \mathbb{R})$, $f(x, y)h := \int_\Omega (\partial_y^{\mathrm{Ul}} \Psi(x, y)h)(s) \, \mathrm{d}s$ is mildly upper semicontinuous. Note that $\mathcal{L}(\mathbb{R}, \mathbb{R})$ can be identified with $\mathbb{R}$. A computation reveals that $f(x, y) \subset \mathbb{R}$ is given by $f(x, y)1 = [V_{<0} - V_{>0} - V_{=0}, V_{<0} - V_{>0} + V_{=0}]$, where $V_{<0} := V_{<0}(x, y) := |\{s \in \Omega : x(s) - y < 0\}|$, $V_{>0} := V_{>0}(x, y) := |\{s \in \Omega : x(s) - y > 0\}|$, and $V_{=0} := V_{=0}(x, y) := |\{s \in \Omega : x(s) - y = 0\}|$. This shows that $f(x, y) \subset \mathbb{R}$ coincides with $\partial_y^{\mathrm{Cl}} F(x, y)$. In effect, we have proven the mild upper semicontinuity of $(x, y) \mapsto \partial_y^{\mathrm{Cl}} F(x, y)$ that is needed in Corollary 2.

Remark. In Example 5 we have observed for a concrete mapping that the composition of Ulbrich's differential with an integration operator is identical to Clarke's generalized gradient; this also follows from [Cla90, Theorem 2.7.2]. We conjecture that this identity holds in a rather general setting, but leave this as a topic for future research.

The next theorem might also be helpful to establish the existence of an implicit function in the presence of semismoothness. It is a special case of a result taken from the book [DR14].

Theorem 5 (Cf. Theorem 5F.4 in [DR14])**.** *Let X be a metric space, Y be a Banach space, and Z be a normed linear space. Let $(\bar{x}, \bar{y}) \in X \times Y$, a neighborhood N_Y of $\bar{y}$, and $F : X \times N_Y \to Z$ with $F(\bar{x}, \bar{y}) = 0$ be given. Suppose that*

- *$F(\cdot, \bar{y})$ is continuous at $\bar{x}$;*

- *there exist a constant* $c > 0$*, a neighborhood* $N_W \subset X \times N_Y$ *of* $(\bar{x}, \bar{y})$*, and a function* $f : N_Y \to Z$ *with* $f(\bar{y}) = 0$ *such that*

$$\|[F(x, y_2) - f(y_2)] - [F(x, y_1) - f(y_1)]\|_Z \leq c\,\|y_2 - y_1\|_Y$$

holds for all $(x, y_1), (x, y_2) \in N_W$*;*

- f *is injective and its inverse mapping is Lipschitz continuous with a constant* $L_{f^{-1}} > 0$ *that satisfies* $c \cdot L_{f^{-1}} < 1$*.*

Then there exist neighborhoods $N_X \subset X$ *of* $\bar{x}$ *and* $\tilde{N}_Y \subset N_Y$ *of* $\bar{y}$ *together with a unique mapping* $g : N_X \to \tilde{N}_Y$ *continuous at* $\bar{x}$ *such that* $g(\bar{x}) = \bar{y}$ *and* $F(x, g(x)) = 0$ *for all* $x \in N_X$*.*

Moreover, if $F(\cdot, y)$ *is continuous in* N_X *for every fixed* $y \in N_Y$*, then* g *is continuous.*

If $F(\cdot, \bar{y})$ *is calm at* $\bar{x}$*, then* g *is calm at* $\bar{x}$*.*

If $F|_{N_X \times \tilde{N}_Y}$ *is Lipschitz continuous with respect to* x *uniformly in* y*, then* g *is Lipschitz continuous.*

Proof. See the proof in [DR14]. Only the claim on the continuity of g in N_X is not treated there. However, it follows directly from the estimate (1) there. □

Remark. 1) If the approximation f of F has the form $f(y) := A(y - \bar{y})$ with an invertible linear operator $A : Y \to Z$, then $L_{f^{-1}}$ is apparently given by $L_{f^{-1}} = \|A^{-1}\|_{\mathcal{L}(Z,Y)}$.

2) It is possible to extend Theorem 5 to set-valued approximations f, cf. the discussion after Theorem 3 in the paper [CD15] by Cibulka and Dontchev.

4 Conclusion

This paper provides an investigation of the semismoothness of implicit functions in infinite-dimensional settings. Certainly, there are interesting topics left. For instance, it could be beneficial to have an existence result similar to Corollary 2, but with Clarke's generalized Jacobian replaced by Ulbrich's generalized derivative for superposition operators. Regarding the existence of implicit functions for superposition operators it may be noteworthy that Gâteaux differentiability can be sufficient to derive an inverse function theorem (from which an implicit function theorem can be deduced in the usual manner), cf. [Eke11, Theorem 2]. Finally, let us remark that besides the implicit function theorems already mentioned in this paper the implicit function theorems from [Tar98] might be helpful to further extend the theory.

References

[ADT14] S. Adly, A.L. Dontchev, and M. Théra. On one-sided Lipschitz stability of set-valued contractions. *Numer. Funct. Anal. Optim.*, 35(7-9):837–850, 2014.

[AF09] J.-P. Aubin and H. Frankowska. *Set-valued analysis.* Boston, MA, Birkhäuser, reprint of the 1990 original edition, 2009.

[AK93] W. Alt and I. Kolumbán. An implicit-function theorem for a class of monotone generalized equations. *Kybernetika*, 29(3):210–221, 1993.

[CD15] R. Cibulka and A.L. Dontchev. A nonsmooth Robinson's inverse function theorem in Banach spaces. *Math. Program. A*, published electronically, April 2015.

[Cia13] P.G. Ciarlet. *Linear and nonlinear functional analysis with applications.* Philadelphia, PA, SIAM, 2013.

[Cla90] F.H. Clarke. *Optimization and nonsmooth analysis.* Philadelphia, PA, SIAM, reprint edition, 1990.

[CNQ00] X. Chen, Z. Nashed, and L. Qi. Smoothing methods and semismooth methods for nondifferentiable operator equations. *SIAM J. Numer. Anal.*, 38(4):1200–1216, 2000.

[De 15] J.C. De los Reyes. *Numerical PDE-constrained optimization.* Cham, Springer, 2015.

[DR14] A.L. Dontchev and R.T. Rockafellar. *Implicit functions and solution mappings. A view from variational analysis.* New York, NY, Springer, 2nd updated edition, 2014.

[Eke11] I. Ekeland. An inverse function theorem in Fréchet spaces. *Ann. Inst. Henri Poincaré, Anal. Non Linéaire*, 28(1):91–105, 2011.

[Gow04] M.S. Gowda. Inverse and implicit function theorems for H-differentiable and semismooth functions. *Optim. Methods Softw.*, 19(5):443–461, 2004.

[HPUU09] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints.* Mathematical Modelling: Theory and Applications 23. Berlin, Springer, 2009.

[IK08] K. Ito and K. Kunisch. *Lagrange multiplier approach to variational problems and applications.* Philadelphia, PA, SIAM, 2008.

[Iof81] A.D. Ioffe. Nonsmooth analysis: differential calculus of nondifferentiable mappings. *Trans. Am. Math. Soc.*, 266:1–56, 1981.

[Jit78] K. Jittorntrum. An implicit function theorem. *J. Optim. Theory Appl.*, 25:575–577, 1978.

[KK16] F. Kruse and K. Kunisch. Optimal control with L^1-constraints. *To appear*, 2016.

[Kum80] S. Kumagai. An implicit function theorem: Comment. *J. Optim. Theory Appl.*, 31:285–288, 1980.

[Kum91] B. Kummer. An implicit-function theorem for $C^{0,1}$-equations and parametric $C^{1,1}$-optimization. *J. Math. Anal. Appl.*, 158(1):35–46, 1991.

[Mif77] R. Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optim.*, 15:959–972, 1977.

[MSZ05] F. Meng, D. Sun, and G. Zhao. Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. *Math. Program.*, 104(2-3 (B)):561–581, 2005.

[Pál97] Z. Páles. Inverse and implicit function theorems for nonsmooth maps in Banach spaces. *J. Math. Anal. Appl.*, 209(1):202–220, 1997.

[PSS03] J.-S. Pang, D. Sun, and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Math. Oper. Res.*, 28(1):39–63, 2003.

[PZ07] Z. Páles and V. Zeidan. Generalized Jacobian for functions with infinite dimensional range and domain. *Set-Valued Anal.*, 15(4):331–375, 2007.

[Qi93] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.*, 18(1):227–244, 1993.

[QS93] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Math. Program.*, 58(3 (A)):353–367, 1993.

[Ral90] D. Ralph. *Rank-1 Support Functionals and the Rank-1 Generalized Jacobian, Piecewise Linear Homeomorphisms.* PhD thesis, University of Wisconsin, Madison, WI, 1990.

[Rob79] S.M. Robinson. Generalized equations and their solutions. I: Basic theory. *Math. Program. Study*, 10:128–141, 1979.

[RW98] R.T. Rockafellar and R.J.-B. Wets. *Variational analysis.* Berlin, Springer, 1998.

[Sch08] A. Schiela. A simplified approach to semismooth Newton methods in function space. *SIAM J. Optim.*, 19(3):1417–1432, 2008.

[Sun01] D. Sun. A further result on an implicit function theorem for locally Lipschitz functions. *Oper. Res. Lett.*, 28(4):193–198, 2001.

[Tar98] A. Tarsia. Differential equations and implicit function: A generalization of the near operators theorem. *Topol. Methods Nonlinear Anal.*, 11(1):115–133, 1998.

[Thi82] L. Thibault. On generalized differentials and subdifferentials of Lipschitz vector-valued functions. *Nonlinear Anal., Theory Methods Appl.*, 6:1037–1053, 1982.

[Ulb11] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces.* Philadelphia, PA, SIAM, 2011.

[vK08] A. von Heusinger and C. Kanzow. SC^1 optimization reformulations of the generalized Nash equilibrium problem. *Optim. Methods Softw.*, 23(6):953–973, 2008.

[War78] J. Warga. An implicit function theorem without differentiability. *Proc. Am. Math. Soc.*, 69:65–69, 1978.

[Zei93] E. Zeidler. *Nonlinear functional analysis and its applications. Volume I: Fixed-point theorems.* New York, NY, Springer, 2nd corrected printing edition, 1993.