Sensitivity Analysis of Elliptic Variational Inequalities of the First and the Second Kind

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Der Fakultät für Mathematik der
Technischen Universität Dortmund
vorgelegt von

Constantin Christof

im Mai 2018

Dissertation

Sensitivity Analysis of Elliptic Variational Inequalities of the First and the Second Kind

Fakultät für Mathematik
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Christian Meyer

Zweitgutachter: Prof. Dr. Martin Brokate

Tag der mündlichen Prüfung: 16. Juli 2018

Acknowledgments

I would like to thank my supervisor, Prof. Dr. Christian Meyer, for giving me the opportunity to write this thesis and for his guidance, encouragement and continuing support. Further, I would like to express my gratitude to Prof. Dr. Gerd Wachsmuth for many interesting discussions and to the German Research Foundation (DFG) for supporting this work within the priority program SPP1962 “Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization”.

Last, but certainly not least, I would like to thank my family and friends for their constant help, patience and support during my studies.

Contents

Introduction

This thesis is concerned with the differential sensitivity analysis of elliptic variational inequalities of the first and the second kind in finite and infinite dimensions. We develop a general theory that provides a sharp criterion for the Hadamard directional differentiability of the solution operator to an elliptic variational inequality and introduce several tools that facilitate the sensitivity analysis in practical applications. Our analysis is accompanied by examples from mechanics and fluid dynamics that illustrate the strengths and limitations of the obtained results. We further establish strong and Bouligand stationarity conditions for optimal control problems governed by elliptic variational inequalities in a general setting that covers, e.g., the situations where the control-to-state mapping is a metric projection or a non-smooth elliptic partial differential equation.

Due to their prominent role in the fields of elastoplasticity, contact mechanics and non-Newtonian fluid dynamics (see, e.g., [Cioranescu et al., 2016; Fuchs and Seregin, 2000; Han and Reddy, 1999; Khludnev and Sokołowski, 1991; Sofonea and Matei, 2012]), elliptic variational inequalities of the first and the second kind have been studied extensively by many mathematicians for a long time. Driven by an increasing interest in the optimal control of (fluid-) mechanical processes, the sensitivity analysis of elliptic variational inequalities in particular has received considerable attention over the course of the last forty years. The first significant contribution in this field was probably Zarantonello's famous lemma on the directional differentiability of metric projections in boundary points, see Theorem 3.2.2 and [Zarantonello, 1971, Lemma 4.6]. The latter was followed by the influential works of Mignot and Bonnans/Shapiro on the concepts of polyhedricity and second-order regularity, cf. [Bonnans and Shapiro, 2000; Haraux, 1977; Mignot, 1976; Shapiro, 1992, 1994b, 2016], and numerous other contributions on the (directional) differentiability of metric projections in Hilbert spaces. We mention exemplarily [Fitzpatrick and Phelps, 1982; Holmes, 1973; Levy, 1999; Noll, 1995; Rockafellar, 1990; Rockafellar and Wets, 1998]. The first works that explicitly addressed the sensitivity analysis of elliptic variational inequalities of the second kind in infinite dimensions were, at least to the author's best knowledge, [Do, 1992] and [Sokołowski, 1988] (cf. also with the preliminary finite-dimensional results of [Rockafellar, 1990] in this context). Since these first contributions, the study of the differentiability properties of solution operators to elliptic variational inequalities of the second kind has developed into an active field that continues to receive attention to this very day. See, e.g., [Borwein and Noll, 1994; Sokołowski and Zolésio, 1992] and the more recent [Adly and Bourdin, 2017; Christof and Meyer, 2016; Christof and Wachsmuth, 2017a,c; De los Reyes and Meyer, 2016; Hintermüller and Surowiec, 2017].

The work that offers the most insight into the mechanisms that are behind the sensitivity analysis of elliptic variational inequalities in general Hilbert spaces is probably that of [Do, 1992]. In the latter, it is proved that the directional differentiability of the solution operator to an elliptic variational inequality of the second kind involving a non-smooth functional j is equivalent to the so-called second-order epi-differentiability of j provided the problem at hand can be identified with a classical Moreau-Yosida regularization (i.e., provided the appearing strongly monotone operator is a positive scalar multiple of the Riesz isomorphism as defined in [Hackbusch, 2017, Conclusion 6.69], cf. Section 1.2). Note that variants of this differentiability result can also be found in [Borwein and Noll, 1994], [Rockafellar, 1990] and [Rockafellar and Wets, 1998].

In this thesis, we extend Do's differentiability criterion to elliptic variational inequalities that involve non-linear or asymmetric operators and, as a consequence, cannot be rewritten as minimization problems. We further provide several tools that simplify the sensitivity analysis in practical applications, study the advantages and limitations of the abstract theory, and explore in detail the relationship between the

notions of second-order epi-differentiability, (extended) polyhedricity and second-order regularity. The novel contributions and main results of this thesis may be summarized as follows:

- We generalize the differentiability criterion in [Do, 1992, Theorem 4.3] to elliptic variational inequalities that cannot be identified with a Moreau-Yosida regularization (see Theorem 1.4.1 for our main result). We obtain this generalization under minimal assumptions on the appearing operators and functionals and without ever using involved instruments from set-valued analysis and the theory of monotone operators. The latter is remarkable since the original proofs of Do (and the related literature) rely heavily on rather sophisticated tools as, e.g., Attouch's theorem on the characterization of Mosco convergence (cf. [Attouch, 1984, Theorem 3.66] and the proofs of [Do, 1992, Theorems 3.9, 4.3]).

- We analyze in detail how the classical notions of polyhedricity and second-order regularity are related to the sharp differentiability criterion that is obtained from our abstract theory. Using tangible (counter-)examples, we demonstrate further that the concept of polyhedricity typically fails as soon as dual spaces and gradient fields are involved. Our results illustrate in particular that several assertions on the polyhedricity of certain sets, that have apparently entered the mathematical folklore throughout the years, are not correct (see Section 3.4).

- We prove a sufficient criterion for the directional differentiability of the solution map to an elliptic variational inequality of the second kind that generalizes the idea behind the concept of (extended) polyhedricity to situations where curvature effects are not negligible. See Lemma 1.3.13 and the more tangible Theorem 4.3.16 for the corresponding results.

- We demonstrate that distributional curvature effects have to be taken into account when elliptic variational inequalities of the second kind in Sobolev spaces are considered (see Section 5.2). Our analysis yields in particular that the approach used in [Sokołowski, 1988; Sokołowski and Zolésio, 1988, 1992] is limited and that the structural assumptions made in [De los Reyes and Meyer, 2016] and [Hintermüller and Surowiec, 2017] are necessary and cannot be dropped without major problems (cf. also with [Christof and Wachsmuth, 2017c] in this context).

- Using a characterization result that is obtained as a byproduct of our sensitivity analysis, we prove strong stationarity conditions for optimal control problems that are governed by elliptic variational inequalities of the first and the second kind. The provided stationarity system covers, e.g., the situations considered in [Mignot and Puel, 1984], [De los Reyes and Meyer, 2016] and [Christof et al., 2017], where the optimal control of the classical obstacle problem, a variational inequality involving the L^1-norm and a non-smooth partial differential equation are studied, respectively, and is, at least to the author's best knowledge, new in its generality.

We hope that the self-containedness and the elementary nature of our approach make our analysis also accessible to those readers who are not familiar with the concepts of, e.g., protodifferentiability and graphical convergence but interested in the differentiability properties of solution operators to elliptic variational inequalities of the first and the second kind. Before we begin with our study, we give a short overview of the structure and the contents of this thesis:

Chapter 1 is devoted to the sensitivity analysis of elliptic variational inequalities in an abstract setting. Here, we first clarify the notation and introduce the problem (P) that we are concerned with in the remainder of this work (see Sections 1.1 and 1.2). In Sections 1.3 and 1.4, we then address the sensitivity analysis of the elliptic variational inequality (P) under consideration. The main result of these sections, Theorem 1.4.1, establishes that the solution operator of an elliptic variational inequality is Hadamard directionally differentiable if and only if the non-smooth functional appearing in the problem at hand is twice epi-differentiable. We point out that this theorem and its proof have already been published in a joint paper of Gerd Wachsmuth and the author (although in a slightly different variant), see [Christof

and Wachsmuth, 2017a]. We conclude the first chapter with some theoretical results that follow from Theorem 1.4.1.

In Chapter 2, we provide several tools that can be used to check the condition of second-order epi-differentiability in practical applications. Section 2.1 first addresses the second-order epi-differentiability of functions with directionally differentiable first derivatives. The main result of this section, Theorem 2.1.1, essentially goes back to [Noll, 1995]. In the subsequent Sections 2.2 to 2.4, we prove a sum rule and two chain rules for twice epi-differentiable functions. Lastly, in Section 2.5, we study the second-order epi-differentiability of functions that are obtained by superposition.

In Chapter 3, we apply the abstract theory of Chapter 1 to elliptic variational inequalities of the first kind. The first three sections of this chapter, Sections 3.1 to 3.3, demonstrate that our analysis indeed covers the classical results of [Zarantonello, 1971], [Bonnans and Shapiro, 2000] and [Mignot, 1976] on the directional differentiability of metric projections, see Theorems 3.2.2, 3.3.5 and 3.3.6. In Section 3.4, we then provide several (counter-) examples that highlight the limitations of the concepts of polyhedricity and second-order regularity in infinite dimensions (cf. also Remark 3.3.2).

Chapter 4 is devoted to elliptic variational inequalities that do not fall under the setting of Chapter 3. In Section 4.1, we first use the results of Chapter 2 to study a class of non-smooth partial differential equations. The subsequent Sections 4.2 and 4.3 are then concerned with elliptic variational inequalities of the second kind that involve seminorms. Here, we provide several criteria for second-order epi-differentiability that allow to study, e.g., the variational inequality of static elastoplasticity (as considered in [De los Reyes et al., 2016]) and partial differential equations involving singular terms (see Section 4.3.5). Sections 4.2.2 and 4.3.3 further address a convenient regularization effect that can be exploited, e.g., in the analysis of (bilevel) optimal control problems and that has, at least to the author's best knowledge, not been documented before.

In Chapter 5, we study in detail two H_0^1-elliptic variational inequalities of the second kind - the so-called Mosolov problem, which arises in non-Newtonian fluid dynamics, and an inequality that involves the L^1-norm. The results obtained in this chapter illustrate which peculiar effects and difficulties may occur in the sensitivity analysis of elliptic variational inequalities in Sobolev spaces and highlight the open questions that remain in this field (cf. Section 5.3). We remark that several of the instruments used in Chapter 5 are also interesting for their own sake. We mention exemplarily the trace criterion in Corollary 5.1.13 and the non-standard Taylor expansion in Corollary 5.2.9 that are obtained as byproducts of the analysis in Sections 5.1 and 5.2, respectively.

Chapter 6 focuses on stationarity conditions for optimal control problems that are governed by elliptic variational inequalities of the first and the second kind. Here, we use the sensitivity analysis of Chapter 1 to derive strong and Bouligand stationarity conditions in an abstract setting that covers, for instance, the situations considered in [Christof et al., 2017; De los Reyes and Meyer, 2016; Mignot and Puel, 1984]. See Theorem 6.1.7, Corollary 6.1.9 and Corollary 6.1.10 in Section 6.1 for the main results. In Section 6.1.1, we further provide several tangible examples that illustrate the applicability of our abstract theory. Section 6.2 finally contains some remarks on the relationship between the sensitivity analysis of Chapter 1 and the study of necessary and sufficient second-order optimality conditions for optimization and optimal control problems.

In the last chapter of this work, we give some concluding remarks on the implications that our results have, possible generalizations and questions that remain open regarding the differentiability properties of solution operators to elliptic variational inequalities of the first and the second kind.

The appendix of this thesis contains a bibliography and a list of symbols covering the most important notational conventions for easy reference.

1 Sensitivity Analysis in an Abstract Setting

This chapter is concerned with the sensitivity analysis of elliptic variational inequalities of the first and the second kind in general Hilbert spaces. After some remarks on the employed notation and basic concepts in Section 1.1, we introduce the problem(-class) that we consider in Section 1.2. The subsequent Section 1.3 then contains the bulk of the actual sensitivity analysis. In Section 1.4, we finally summarize our results in a main theorem, Theorem 1.4.1, and explore the consequences that our findings have, e.g., for the study of the second-order differentiability properties of convex functions.

We would like to point out that Theorem 1.4.1 and the majority of the results in Section 1.3 have already been published in [Christof and Wachsmuth, 2017a] (albeit in a slightly different setting). The author wishes to express his gratitude to Gerd Wachsmuth, the coauthor of this paper, whose remarks on the proofs of parts (ii) and (iii) of Proposition 1.3.5 and Corollary 1.4.4 helped to strengthen the obtained results significantly.

1.1 Comments on the Notation and Basic Concepts

Before we begin with our investigation, we give some preliminary remarks on the employed notation and basic concepts that are needed for our analysis.

The notation that we use in this thesis is kept as standard as possible and we hope that it is consistent enough with the available literature to make our results accessible also to those readers who are not familiar with the field under consideration. In what follows, we use the letters H, U, V to denote Hilbert spaces and the letters X, Y to denote Banach spaces. For the topological dual of a Banach space X, we write X^*. Norms, dual pairings and scalar products are denoted with the symbols $\|\cdot\|$, $\langle\cdot,\cdot\rangle$ and $(\cdot,\cdot)$, respectively (we often suppress the dependency on the space and write $\langle\cdot,\cdot\rangle_X$ etc. only if it is necessary to avoid ambiguities). The letters F, G are reserved for Banach and Hilbert space valued mappings, the letters A, B for functions that map the primal into the dual space, and the letters S, T for solution operators to variational inequalities and partial differential equations. Given two Banach spaces X and Y, we define $L(X,Y) := \{F : X \to Y \mid F \text{ is linear and bounded}\}$. With j, k, l we denote scalar functions (and the associated Nemytskii operators, respectively). The letters K and L are used for convex non-empty sets. Note that, throughout this thesis, we introduce new variables and symbols whenever necessary. Such notation is defined when it first appears in the text. We refer to the list of symbols in the appendix of this work for a fairly complete overview.

In addition to the above conventions, in what follows, we use several classical abbreviations and notations from (convex) analysis, linear algebra and measure theory. Given a convex, proper function $j : X \to (-\infty, \infty]$, we write, e.g.,

$$\begin{aligned}
\operatorname{dom}(j) &:= \{x \in X \mid j(x) < \infty\},\\
\operatorname{graph}(j) &:= \{(x, j(x)) \in X \times \mathbb{R} \mid x \in \operatorname{dom}(j)\},\\
\operatorname{epi}(j) &:= \{(x,\alpha) \in X \times \mathbb{R} \mid x \in \operatorname{dom}(j),\ \alpha \geq j(x)\},\\
\partial j(x) &:= \{x^* \in X^* \mid j(y) - j(x) \geq \langle x^*, y - x\rangle \ \forall y \in X\}
\end{aligned}$$

for the domain, the graph, the epigraph and the subdifferential at a point $x \in \operatorname{dom}(j)$ of j. Note that, in the remainder of this thesis, we work with the definition $\partial j(x) := \emptyset$ for all $x \in X \setminus \operatorname{dom}(j)$. We further use the notation $\operatorname{graph}(\cdot)$ also in the set-valued sense, i.e., we write, e.g.,

$$\operatorname{graph}(\partial j) = \{(x, x^*) \in X \times X^* \mid x \in \operatorname{dom}(j),\ x^* \in \partial j(x)\}$$

to denote the graph of the set-valued mapping $\partial j : X \rightrightarrows X^*$, $x \mapsto \partial j(x)$. For the convenience of the reader, all of the employed abbreviations ($\mathrm{span}, \mathrm{int}, \mathrm{cl}, \mathrm{tr}, \mathrm{sign}, \mathrm{conv}$ etc.) are defined upon their first appearance and also explained in the list of symbols at the end of this work.

As local approximations of convex sets are of particular importance for our analysis (cf. Section 1.3 and Chapter 3), throughout this thesis, we will make frequent use of the following classical concepts (see [Bonnans and Shapiro, 2000, Section 2.2.4], [Schirotzek, 2007, Chapter 11]):

Definition 1.1.1 (Radial, Tangent and Normal Cone)**.** *Let X be a Banach space and let $L \subset X$ be a convex, non-empty set (not necessarily closed). Then, the radial, the tangent, and the normal cone to L at a point $x \in L$ are defined, respectively, by*

$$\begin{aligned} \mathcal{T}_L^{rad}(x) &:= \mathbb{R}^+(L - x), \qquad \mathcal{T}_L(x) := \mathrm{cl}(\mathcal{T}_L^{rad}(x)), \\ \mathcal{N}_L(x) &:= \{x^* \in X^* \mid \langle x^*, z\rangle \leq 0 \quad \forall z \in \mathcal{T}_L(x)\}. \end{aligned}$$

Here, $\mathrm{cl}(\cdot)$ denotes the topological closure of a set.

Note that, from Mazur's lemma and standard arguments, we obtain that the following holds true in the situation of Definition 1.1.1 (cf. [Bonnans and Shapiro, 2000, Proposition 2.55]):

$$\begin{aligned} \mathcal{T}_L(x) &= \left\{ z \in X \;\middle|\; \exists t_n \searrow 0 \, \exists x_n \in L \text{ such that } \frac{x_n - x}{t_n} \to z \text{ in } X \right\} \\ &= \left\{ z \in X \;\middle|\; \forall t_n \searrow 0 \, \exists x_n \in L \text{ such that } \frac{x_n - x}{t_n} \to z \text{ in } X \right\} \\ &= \left\{ z \in X \;\middle|\; \exists t_n \searrow 0 \, \exists x_n \in L \text{ such that } \frac{x_n - x}{t_n} \rightharpoonup z \text{ in } X \right\} \qquad \forall x \in L. \end{aligned} \tag{1.1}$$

The above identities will be used extensively in Section 1.3. To conclude this section, we recall several notions of differentiability that are relevant for our sensitivity analysis (cf. [Bonnans and Shapiro, 2000, Section 2.2], [Schirotzek, 2007, Chapter 3] and [Shapiro, 1990]).

Definition 1.1.2. *Let X and Y be Banach spaces. Assume that a convex, non-empty set $L \subset X$ (not necessarily open) and a function $F : L \to Y$ are given. Then, F is called:*

(i) directionally differentiable in an $x \in L$ in a direction $z \in \mathcal{T}_L^{rad}(x)$ if there exists an $F'(x; z) \in Y$ (the directional derivative in x in the direction z) such that for all sequences $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ and $x + t_n z \in L$ for all n, it holds

$$\lim_{n\to\infty} \frac{F(x + t_n z) - F(x)}{t_n} = F'(x; z).$$

(ii) Hadamard directionally differentiable in an $x \in L$ in a direction $z \in \mathcal{T}_L(x)$ if there exists an $F'(x; z) \in Y$ (the Hadamard directional derivative in x in the direction z) such that for all sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset X$ with $t_n \searrow 0$, $z_n \to z$ and $x + t_n z_n \in L$ for all n, it holds

$$\lim_{n\to\infty} \frac{F(x + t_n z_n) - F(x)}{t_n} = F'(x; z).$$

(iii) (Hadamard-) Gâteaux differentiable in an $x \in L$ if F is (Hadamard) directionally differentiable in all directions $z \in \mathcal{T}_L^{rad}(x)$ (or $z \in \mathcal{T}_L(x)$, respectively) and if the (Hadamard) directional derivative can be extended to a continuous linear function $F'(x; \cdot) \in L(X, Y)$.

(iv) Fréchet differentiable in a point $x \in L$ if there exists an $F'(x) \in L(X, Y)$ with

$$\lim_{z\to 0,\, x+z\in L} \frac{\|F(x + z) - F(x) - F'(x)z\|_Y}{\|z\|_X} = 0.$$

If one of the above conditions is satisfied for all $x \in L$ and all relevant directions $z \in X$, then we drop the reference to x and z and simply say that F has the respective differentiability property.

We point out that Hadamard directional differentiability is what is typically required for the derivation of chain rules and first-order optimality conditions, cf. [Bonnans and Shapiro, 2000, Proposition 2.47] and [Schirotzek, 2007, Proposition 12.1.1]. As a consequence, this notion of differentiability is of particular importance in the fields of optimal control and optimization. A special feature of the concept of Hadamard directional differentiability is that it implies the continuity of the directional derivative $F'(x;\cdot) : \mathcal{T}_L(x) \to Y$ (cf. [Bonnans and Shapiro, 2000, Proposition 2.46]):

Lemma 1.1.3. *Let X and Y be Banach spaces, and suppose that a convex, non-empty set $L \subset X$ (not necessarily open), an $x \in L$ and a function $F : L \to Y$ are given such that F is Hadamard directionally differentiable in x in all directions $z \in \mathcal{T}_L(x)$. Then, the function $F'(x;\cdot) : \mathcal{T}_L(x) \to Y$ is continuous.*

Proof. Suppose that a $z \in \mathcal{T}_L(x)$ and a sequence $\{z_n\} \subset \mathcal{T}_L(x)$ with $z_n \to z$ in X are given. From (1.1), we obtain that for every arbitrary but fixed $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$ we can find $\{z_{n,m}\} \subset X$ with $x + t_m z_{n,m} \in L$ for all n, m and $z_{n,m} \to z_n$ for all n as $m \to \infty$. Choose $\{m_n\} \subset \mathbb{N}$ such that

$$t_{m_n} + \|z_{n,m_n} - z_n\|_X + \left\| \frac{F(x + t_{m_n} z_{n,m_n}) - F(x)}{t_{m_n}} - F'(x; z_n) \right\|_Y \leq \frac{1}{n} \qquad \forall n \in \mathbb{N}.$$

Then, the definition of $\{m_n\}$ and the Hadamard directional differentiability of F in x imply

$$\left\|F'(x; z_n) - F'(x; z)\right\|_Y \leq \frac{1}{n} + \left\| \frac{F(x + t_{m_n} z_{n,m_n}) - F(x)}{t_{m_n}} - F'(x; z) \right\|_Y \to 0$$

for $n \to \infty$. This proves the claim. □

1.2 The Problem under Consideration

As already mentioned in the introduction, the aim of this thesis is to study the differentiability properties of solution operators to elliptic variational inequalities (EVIs) of the first and the second kind. The problem(-class) that we consider in the remainder of this work takes the following form:

$$w \in K, \quad \langle A(w), v - w\rangle + j(v) - j(w) \geq \langle f, v - w\rangle \quad \forall v \in K. \tag{P}$$

Our assumptions on the quantities in (P) are as follows:

Assumption 1.2.1 (Standing Assumptions and Notation for the Abstract Setting)**.**

- *V is a Hilbert space with topological dual V^* and dual pairing $\langle\cdot,\cdot\rangle$.*
- *$f \in V^*$ is a given datum (the argument of the solution map).*
- *$j : V \to (-\infty, \infty]$ is a convex, lower semicontinuous and proper function.*
- *$K := \mathrm{dom}(j)$ is the (necessarily convex and non-empty) domain of j.*
- *$A : K \to V^*$ is an operator with the following properties:*
  - *(i) A maps bounded subsets of K into bounded subsets of V^*.*
  - *(ii) A is strongly monotone on K, i.e., there exists a constant $c > 0$ such that*

$$\langle A(v_1) - A(v_2), v_1 - v_2\rangle \geq c\|v_1 - v_2\|_V^2 \quad \forall v_1, v_2 \in K. \tag{1.2}$$

  - *(iii) A is Fréchet differentiable on K in the sense of Definition 1.1.2(iv).*

Note that, in the literature, problems of the form (P) are commonly referred to as elliptic variational inequalities of the second kind while problems of the type

$$w \in K, \quad \langle A(w), v - w\rangle \geq \langle f, v - w\rangle \quad \forall v \in K \tag{1.3}$$

with a closed, convex, non-empty set $K \subset V$ and an operator A as in Assumption 1.2.1 are typically called elliptic variational inequalities of the first kind (cf., e.g., [Glowinski, 1980, Chapter 1]). This distinction, however, is largely artificial. If we start with a problem of the type (1.3) and define $j := \chi_K$, where $\chi_D : V \to \{0, \infty\}$ denotes the characteristic function of a set $D \subset V$, then we immediately arrive at an EVI of the form (P). Conversely, it is also often possible to rewrite a problem of the type (P) in the form (1.3), e.g., by dualization or by using the epigraph of the function j, cf. [Christof and Wachsmuth, 2017c], [Sokołowski, 1988], [Oden and Kikuchi, 1980, Section 1.7] and the proof of Theorem 1.2.2. Since the formulation (P) turns out to be slightly more general than (1.3), we have decided to work with elliptic variational inequalities of the second kind in this chapter.

We would like to point out that, in spite of the above observation, it often still makes sense to study EVIs of the type (1.3) separately, e.g., to exploit the special structure of the non-smoothness $j = \chi_K$. Details on this topic may be found in Chapter 3.

For later use, we mention that variational inequalities of the form (P) (and (1.3), respectively) can also often be reformulated as convex minimization problems. Indeed, if the operator $A : K \to V^*$ admits a potential $k : K \to \mathbb{R}$, i.e., if there exists a Fréchet differentiable function $k : K \to \mathbb{R}$ such that $k'(v) = A(v) \in V^*$ for all $v \in K$, then every solution $w \in K$ of the problem

$$\min_{v \in K} k(v) + j(v) - \langle f, v\rangle \tag{1.4}$$

satisfies

$$\begin{aligned} 0 &\leq \lim_{t \searrow 0} \frac{k(w + t(v - w)) + j(w + t(v - w)) - \langle f, w + t(v - w)\rangle - k(w) - j(w) + \langle f, w\rangle}{t} \\ &\leq \langle A(w), v - w\rangle + j(v) - j(w) - \langle f, v - w\rangle \qquad \forall v \in K, \end{aligned}$$

and we arrive precisely at the EVI (P). If, conversely, we are given a solution $w \in K$ to (P) in the above situation, then the function

$$\psi_v : [0, 1] \to \mathbb{R}, \qquad t \mapsto k(w + t(v - w)),$$

is Fréchet differentiable in $[0, 1]$ for all $v \in K$ with derivative

$$\psi_v'(t) = \langle A(w + t(v - w)), v - w\rangle \quad \forall t \in [0, 1],$$

and we obtain from (1.2) that

$$\psi_v'(t_1) - \psi_v'(t_2) = \langle A(w + t_1(v - w)) - A(w + t_2(v - w)), v - w\rangle \geq c(t_1 - t_2)\|v - w\|_V^2 \tag{1.5}$$

for all $0 \leq t_2 \leq t_1 \leq 1$. The estimate (1.5) implies that ψ_v' is monotonously increasing, that ψ_v is convex and that

$$\langle A(w), v - w\rangle = \psi_v'(0) = \inf_{t \in (0,1]} \frac{k(w + t(v - w)) - k(w)}{t} \leq k(v) - k(w) \qquad \forall v \in K.$$

In particular,

$$\begin{aligned} 0 &\leq \langle A(w), v - w\rangle + j(v) - j(w) - \langle f, v - w\rangle \\ &\leq k(v) + j(v) - \langle f, v\rangle - k(w) - j(w) + \langle f, w\rangle \qquad \forall v \in K. \end{aligned}$$

This shows that w is also a solution to (1.4) and that the problems (P) and (1.4) are indeed equivalent.

We remark that, using the above observation and the direct method of the calculus of variations, it is straightforward to prove that all EVIs (P) whose operators $A : K \to V^*$ possess a potential $k : K \to \mathbb{R}$ admit a unique solution $w \in K$. In the general setting of Assumption 1.2.1, this unique solvability can be established with the help of Browder's theorem as the following result shows.

Theorem 1.2.2. *Suppose that Assumption 1.2.1 holds. Then, for every right-hand side $f \in V^*$, there exists one and only one solution $w \in K$ to* (P)*. Moreover, the solution operator $S : V^* \to V$, $f \mapsto w$, associated with* (P) *satisfies*

$$\|S(f_1) - S(f_2)\|_V \leq \frac{1}{c}\|f_1 - f_2\|_{V^*} \qquad \forall f_1, f_2 \in V^*. \tag{1.6}$$

Here, $c > 0$ denotes the monotonicity constant in (1.2)*.*

Proof. The proof uses standard techniques from the theory of (pseudo-)monotone operators as found, e.g., in [Oden and Kikuchi, 1980] and [Ruzicka, 2004, Chapter 3]. We proceed in three steps:

Step 1 (Reduction to the case $j = \chi_L$): In a first step, we demonstrate that (P) can be reformulated as a variational inequality on the product space $U := V \times \mathbb{R}$ (cf. [Oden and Kikuchi, 1980, Section 1.7]). Consider the problem

$$(w, \beta) \in L, \qquad \langle B(w, \beta), (v, \alpha) - (w, \beta)\rangle_U \geq 0 \qquad \forall (v, \alpha) \in L \tag{1.7}$$

with admissible set $L := \operatorname{epi}(j)$ and operator

$$B : L \to U^* \cong V^* \times \mathbb{R}, \qquad (v, \alpha) \mapsto (A(v) - f, 1),$$

and suppose that a solution (w, β) to (1.7) is given. Then, it necessarily holds $j(w) \in \mathbb{R}$, and we may test (1.7) with $(w, j(w)) \in L$ to obtain

$$\langle B(w, \beta), (w, j(w)) - (w, \beta)\rangle_U = j(w) - \beta \geq 0. \tag{1.8}$$

Since $\beta \geq j(w)$ by the definition of L, (1.8) implies $\beta = j(w)$. Choosing test functions of the form $(v, j(v)) \in L$ in (1.7) now yields

$$\langle B(w, j(w)), (v, j(v)) - (w, j(w))\rangle_U = \langle A(w) - f, v - w\rangle_V + j(v) - j(w) \geq 0 \quad \forall v \in K.$$

The above shows that a tuple $(w, \beta) \in L$ can only be a solution to (1.7) if $\beta = j(w)$ and if the first component w is a solution to (P). If, conversely, we start with a solution $w \in K$ to (P), then it trivially holds $j(w) \in \mathbb{R}$ and

$$\begin{aligned}\langle B(w, j(w)), (v, \alpha) - (w, j(w))\rangle_U &= \langle A(w) - f, v - w\rangle_V + \alpha - j(w) \\ &\geq \langle A(w) - f, v - w\rangle_V + j(v) - j(w) \geq 0 \qquad \forall (v, \alpha) \in L.\end{aligned}$$

The last estimate implies that the tuple $(w, j(w)) \in L$ solves (1.7) and that there is indeed a one-to-one correspondence between the solutions of the problems (P) and (1.7).

Step 2 (Solvability of (P)): From Step 1, it follows that it suffices to prove the solvability of (1.7) to obtain the existence of a solution to the problem (P). Note that the set L appearing in (1.7) is trivially convex, closed and non-empty (since L is the epigraph of a convex, lower semicontinuous and proper function). Further, the strong monotonicity of A implies

$$\langle B(v_1, \alpha_1) - B(v_2, \alpha_2), (v_1, \alpha_1) - (v_2, \alpha_2)\rangle_U = \langle A(v_1) - A(v_2), v_1 - v_2\rangle_V \geq c\|v_1 - v_2\|_V^2$$

for all $(v_1, \alpha_1), (v_2, \alpha_2) \in L$, and from the Fréchet differentiability and the boundedness of the operator $A : K \to V^*$, it follows immediately that $B : L \to U^*$ is continuous and bounded on bounded sets. In particular, B is pseudomonotone by [Ruzicka, 2004, Section 3.2, Lemma 2.6], and it holds

$$\begin{aligned}&\langle B(v, \alpha), (v, \alpha) - (u, j(u))\rangle_U \\ &\quad \geq c\|v - u\|_V^2 + \langle B(u, j(u)), (v, \alpha) - (u, j(u))\rangle_U \\ &\quad \geq c\|v - u\|_V^2 + \alpha - j(u) - \|A(u) - f\|_{V^*}\|v - u\|_V \\ &\quad \geq c\|v - u\|_V^2 + |\alpha| + 2\min(0, j(v)) - j(u) - \|A(u) - f\|_{V^*}\|v - u\|_V\end{aligned} \tag{1.9}$$

for all $u \in \mathrm{dom}(j)$ and all $(v, \alpha) \in L$. Recall that the convexity and the lower semicontinuity of j yield that there exist an $l \in V^*$ and a $\tilde{c} \in \mathbb{R}$ with

$$j(v) \geq \langle l, v \rangle_V + \tilde{c} \qquad \forall v \in V.$$

If we use this estimate and Young's inequality in (1.9), then we obtain that there are constants $c_1, c_2 > 0$ independent of (v, α) with

$$\langle B(v,\alpha), (v,\alpha) - (u, j(u)) \rangle_U \geq c_1 \left(\|v\|_V^2 + |\alpha| \right) - c_2 \qquad \forall (v,\alpha) \in L.$$

The above implies that for every $u \in \mathrm{dom}(j)$ we can find an $r = r(u) > 0$ with

$$\langle B(v,\alpha), (v,\alpha) - (u, j(u)) \rangle_U > 0 \qquad \forall (v,\alpha) \in L, \quad \|(v,\alpha)\|_U > r,$$

and that the operator $B : L \to U^*$ is coercive in the sense of [Ruzicka, 2004, Theorem 3.43] and [Oden and Kikuchi, 1980, Condition (6.9)]. Using Browder's theorem as found, e.g., in [Ruzicka, 2004, Section 3.3.3] and [Oden and Kikuchi, 1980, Theorem 1-6.2], it now follows straightforwardly that (1.7) admits at least one solution (w, β). This proves that (P) is solvable.

Step 3 (Uniqueness of the solution and Lipschitz estimate): Suppose that $w_1, w_2 \in K$ solve the problem (P) with right-hand side f_1 and f_2, respectively. Then, it holds $j(w_1), j(w_2) \in \mathbb{R}$ and it follows from the EVIs for w_1 and w_2 that

$$\begin{aligned} \langle A(w_1), w_2 - w_1 \rangle + j(w_2) - j(w_1) &\geq \langle f_1, w_2 - w_1 \rangle, \\ \langle A(w_2), w_1 - w_2 \rangle + j(w_1) - j(w_2) &\geq \langle f_2, w_1 - w_2 \rangle. \end{aligned} \tag{1.10}$$

If we add the inequalities in (1.10) and use the strong monotonicity of A, then we obtain

$$c\|w_1 - w_2\|_V^2 \leq \langle f_1 - f_2, w_1 - w_2 \rangle \leq \|f_1 - f_2\|_{V^*} \|w_1 - w_2\|_V. \tag{1.11}$$

The above proves, on the one hand, that for each $f \in V^*$ there can only be one solution to (P) and, on the other hand, that the solution operator $S : V^* \to V$ associated with (P) satisfies (1.6). □

The reader who is familiar with the theory of monotone operators may have noticed that the above proof does not rely on the Hilbert space structure of V and that the unique solvability of (P) can also be obtained under monotonicity assumptions that are weaker than our condition (1.2). We do not work with a more general setting than that in Assumption 1.2.1 here because this is not sensible when the (directional) differentiability of the solution operator $f \mapsto w$ is the property of interest. Consider, for example, the situation

$$V := \mathbb{R}, \qquad j := 0, \qquad K := \mathbb{R}, \qquad A(v) := |v|^{q-2} v \in V^* \cong \mathbb{R}, \tag{1.12}$$

where $q > 2$ is arbitrary but fixed. Then, the quantities V, j, K and A clearly satisfy all conditions in Assumption 1.2.1 except for (1.2), and it holds (cf. [Lindqvist, 2017, Section 12])

$$\left\langle |v_1|^{q-2} v_1 - |v_2|^{q-2} v_2, v_1 - v_2 \right\rangle \geq 2^{2-q} |v_1 - v_2|^q \qquad \forall v_1, v_2 \in K.$$

It is easy to check that the above monotonicity property is still enough to prove the unique solvability of the problem

$$w \in K, \quad \langle A(w), v - w \rangle + j(v) - j(w) \geq \langle f, v - w \rangle \quad \forall v \in K \tag{1.13}$$

with the technique that we have used for the derivation of Theorem 1.2.2. However, by direct calculation, we may also compute that the solution operator to (1.13) with V, j, K and A as in (1.12) is given by $S(f) = \mathrm{sgn}(f)|f|^{1/(q-1)}$, where $\mathrm{sgn}(\cdot)$ denotes the signum function. The non-differentiability that appears here demonstrates that the solution map $f \mapsto w$ to a variational inequality of the form (P) cannot be expected to be (directionally) differentiable if the strong monotonicity condition (1.2) is relaxed. We

point out that the effect that is responsible for the non-smoothness in the above is also present (and even more severe) when variational inequalities in function spaces are considered whose operators A behave, e.g., like a q-Laplacian with $q > 2$. This shows that it makes sense to restrict the attention to the case (1.2) in the subsequent analysis.

To answer the question of why we work with a Hilbert space in Assumption 1.2.1 and not, e.g., with a reflexive Banach space, we note the following:

Lemma 1.2.3. *Assume that X is a Banach space with topological dual X^* and dual pairing $\langle \cdot, \cdot \rangle$. Suppose further that $j : X \to (-\infty, \infty]$ is a proper function with a convex domain $\mathrm{dom}(j) =: K$, and that $A : K \to X^*$ is an operator with the following properties:*

- *A is strongly monotone in K, i.e., there exists a constant $c > 0$ such that*

$$\langle A(x_1) - A(x_2), x_1 - x_2 \rangle \geq c\|x_1 - x_2\|_X^2 \quad \forall x_1, x_2 \in K. \tag{1.14}$$

- *A is Fréchet differentiable on K in the sense of Definition 1.1.2(iv).*

Then, the following holds true:

(i) For every $x \in K$, we have

$$\big\langle A'(x)(z_1 - z_2), z_1 - z_2 \big\rangle \geq c\|z_1 - z_2\|_X^2 \quad \forall z_1, z_2 \in \mathcal{T}_K(x). \tag{1.15}$$

Here, c is the monotonicity constant of A in (1.14) *and $\mathcal{T}_K(x)$ is the tangent cone to K in x.*

(ii) If $x \in K$ is arbitrary but fixed and if $H := \mathrm{cl}(\mathrm{span}(K - x))$, where $\mathrm{span}(\cdot)$ denotes the linear span of a set, i.e.,

$$\mathrm{span}(D) := \left\{ \sum_{n=1}^{N} \alpha_n x_n \;\middle|\; N \in \mathbb{N},\ x_n \in D,\ \alpha_n \in \mathbb{R} \right\} \quad \forall D \subset X,$$

then there exists a norm $\|\cdot\|_H : H \to [0, \infty)$ such that $(H, \|\cdot\|_H)$ is a Hilbert space and such that the norms $\|\cdot\|_H$ and $\|\cdot\|_X$ are equivalent on H.

(iii) Let $x_0 \in K$ and $f \in X^*$ be arbitrary but fixed. Then, $y \in X$ is a solution to the problem

$$y \in K, \quad \langle A(y), x - y \rangle + j(x) - j(y) \geq \langle f, x - y \rangle \quad \forall x \in K \tag{1.16}$$

if and only if $y - x_0 \in \mathrm{cl}(\mathrm{span}(K - x_0))$ is a solution to the variational inequality

$$\tilde{y} \in K - x_0, \quad \langle A(\tilde{y} + x_0), \tilde{x} - \tilde{y} \rangle + j(\tilde{x} + x_0) - j(\tilde{y} + x_0) \geq \langle f, \tilde{x} - \tilde{y} \rangle \quad \forall \tilde{x} \in K - x_0. \tag{1.17}$$

Proof. Ad (i): Let $x \in K$ and $z_1, z_2 \in \mathcal{T}_K^{rad}(x) = \mathbb{R}^+(K - x)$ be arbitrary but fixed. Then, the convexity of K implies that there exists an $\varepsilon > 0$ with $x + tz_1, x + tz_2 \in K$ for all $0 < t < \varepsilon$, and we obtain from the strong monotonicity and the Fréchet differentiability of the operator $A : K \to X^*$ that

$$c\|z_1 - z_2\|_X^2 \leq \frac{1}{t} \langle A(x + tz_1) - A(x + tz_2), z_1 - z_2 \rangle = \big\langle A'(x)(z_1 - z_2), z_1 - z_2 \big\rangle + o(1)$$

holds for all $0 < t < \varepsilon$. Passing to the limit $t \searrow 0$ in the above and using that $A'(x) \in L(X, X^*)$ and $\mathcal{T}_K(x) = \mathrm{cl}(\mathcal{T}_K^{rad}(x))$, we arrive at (1.15) as claimed.

Ad (ii): Suppose that an $x \in K$ is given. From the convexity and the cone property of the set $\mathcal{T}_K^{rad}(x)$, it follows that every

$$h := \sum_{n=1}^{N} \alpha_n (x_n - x) \in \mathrm{span}(K - x), \quad N \in \mathbb{N}, \quad x_n \in K, \quad \alpha_n \in \mathbb{R},$$

satisfies

$$h = \sum_{\alpha_n > 0} |\alpha_n|(x_n - x) - \sum_{\alpha_n < 0} |\alpha_n|(x_n - x) \in \mathcal{T}_K^{rad}(x) - \mathcal{T}_K^{rad}(x).$$

The above implies in combination with part (i) of the lemma that

$$\big\langle A'(x)h, h \big\rangle \geq c\|h\|_X^2 \quad \forall h \in H := \mathrm{cl}(\mathrm{span}(K - x)).$$

Consider now the bilinear form

$$(h_1, h_2)_H := \frac{1}{2}\Big(\big\langle A'(x)h_1, h_2 \big\rangle + \big\langle A'(x)h_2, h_1 \big\rangle \Big), \quad h_1, h_2 \in H.$$

Then, $(\cdot, \cdot)_H$ clearly defines a scalar product on H and it holds

$$c\|h\|_X^2 \leq (h, h)_H \leq \|A'(x)\|_{L(X,X^*)}\|h\|_X^2 \quad \forall h \in H.$$

This proves the claim.

Ad (iii): The equivalence of (1.16) and (1.17) follows from a trivial translation argument. □

Lemma 1.2.3 demonstrates that, if we are given a variational inequality of the form (1.16) in some Banach space X with a convex, proper and lower semicontinuous function $j : X \to (-\infty, \infty]$ and a strongly monotone and Fréchet differentiable operator $A : \mathrm{dom}(j) \to X^*$, then we can always rewrite this problem as an EVI in an appropriately defined Hilbert space (namely, the closure of the space of directions of the affine hull $\mathrm{aff}(\mathrm{dom}(j))$). This illustrates that working with Banach spaces is not useful in the presence of the strong monotonicity condition (1.2) (which, as we have seen, is a reasonable assumption when the directional differentiability of the solution operator is studied) and shows that Assumption 1.2.1 is, in a sense, optimal.

We would like to point out that the setting that we have introduced in this section is very flexible and covers a multitude of different situations. Compare, e.g., with [Borwein and Noll, 1994; Christof et al., 2017; De los Reyes et al., 2016; De los Reyes and Meyer, 2016; Do, 1992; Han and Reddy, 1999; Haraux, 1977; Hintermüller and Surowiec, 2017; Mignot, 1976; Noll, 1995; Shapiro, 1994b; Sokołowski, 1988] in this context. For tangible examples of problems that fall under the scope of our analysis, we refer to Chapters 3 to 5.

1.3 Differential Sensitivity Analysis

In what follows, we always tacitly assume that a problem of the form (P) is given and that the conditions in Assumption 1.2.1 are satisfied.

To study the (directional) differentiability of the solution operator $S : V^* \to V$ associated with (P), we have to analyze how the difference quotients

$$\delta_t := \frac{S(f + tg) - S(f)}{t} \in V, \quad t > 0, \quad f, g \in V^*, \tag{1.18}$$

behave as t tends to zero (cf. Definition 1.1.2). So let us consider an arbitrary but fixed right-hand side $f \in V^*$ and some direction $g \in V^*$, and let δ_t be defined as in (1.18). Then, the global Lipschitz continuity of S implies $\|\delta_t\|_V \leq \|g\|_{V^*}/c$ for all $t > 0$ (see Theorem 1.2.2), and it follows from the definition of δ_t that

$$\langle A(w + t\delta_t), v - w - t\delta_t \rangle + j(v) - j(w + t\delta_t) \geq \langle f + tg, v - w - t\delta_t \rangle \quad \forall v \in K, \tag{1.19}$$

where $w := S(f)$ denotes the solution of the unperturbed problem. If we use test functions of the form $v = w + tz \in K$ in the above and exploit the Fréchet differentiability of $A : K \to V^*$, then (1.19) can be rewritten as

$$\begin{aligned}
&\delta_t \in \frac{1}{t}\,(K - w)\,,\\
&\big\langle A'(w)\delta_t, z - \delta_t\big\rangle + \frac{1}{t}\left(\frac{j(w + tz) - j(w)}{t} - \langle f - A(w), z\rangle\right)\\
&\qquad - \frac{1}{t}\left(\frac{j(w + t\delta_t) - j(w)}{t} - \langle f - A(w), \delta_t\rangle\right) \geq \langle g, z - \delta_t\rangle - r_t(g)\|z - \delta_t\|_V\\
&\qquad\qquad \forall z \in \frac{1}{t}\,(K - w)
\end{aligned} \tag{1.20}$$

with a remainder $r_t(g) \in \mathbb{R}$ satisfying

$$\begin{aligned}
0 \leq r_t(g) &:= \frac{\|A(w + t\delta_t) - A(w) - tA'(w)\delta_t\|_{V^*}}{t}\\
&\leq \frac{\|A(w + t\delta_t) - A(w) - tA'(w)\delta_t\|_{V^*}}{\|t\delta_t\|_V}\,\frac{\|g\|_{V^*}}{c} \to 0
\end{aligned}$$

for $t \searrow 0$ independently of z. Note that the functions

$$j_t : V \to (-\infty, \infty], \qquad z \mapsto \frac{2}{t}\left(\frac{j(w + tz) - j(w)}{t} - \langle f - A(w), z\rangle\right), \quad t > 0,$$

appearing in (1.20) are convex, lower semicontinuous and proper for all $t > 0$, and that Lemma 1.2.3 yields

$$\big\langle A'(v)(z_1 - z_2), z_1 - z_2\big\rangle \geq c\|z_1 - z_2\|_V^2 \quad \forall z_1, z_2 \in \mathcal{T}_K(v) \quad \forall v \in K. \tag{1.21}$$

The above implies that the bilinear form $(z_1, z_2) \mapsto \langle A'(w)z_1, z_2\rangle$ in (1.20) is coercive on the domain $\mathrm{dom}(j_t) = \frac{1}{t}\,(K - w) \subset \mathcal{T}_K^{rad}(w) \subset \mathcal{T}_K(w)$ and that the difference quotients δ_t are themselves solutions to elliptic variational inequalities of the second kind that are perturbed by the term $r_t(g)\|z - \delta_t\|_V$.

In what follows, our aim will be to pass to the limit $t \searrow 0$ in (1.20) and to obtain information about the convergence properties of the difference quotients δ_t from the limiting behavior of the associated variational inequalities. Unfortunately, the limit transition in (1.20) is far from straightforward. Since the function j is not assumed to possess first or second derivatives, it is perfectly possible that the term $j_t(z)$ in (1.20) diverges, and since the behavior of the difference quotients δ_t is presently completely unknown, the quantity $j_t(\delta_t)$ is hard to control. To overcome these problems, we introduce the following concept:

Definition 1.3.1 (Weak Second Subderivative)**.** *Given a tuple $(v, \varphi) \in \mathrm{graph}(\partial j)$, we define the (weak) second subderivative $Q_j^{v,\varphi} : V \to [0, \infty]$ of j at v for φ by*

$$Q_j^{v,\varphi}(z) := \inf\left\{\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n\rangle\right) \;\middle|\; \begin{array}{l}\{t_n\} \subset \mathbb{R}^+, \{z_n\} \subset V,\\ t_n \searrow 0,\ z_n \rightharpoonup z\end{array}\right\}. \tag{1.22}$$

Recall that the definition of the convex subdifferential yields

$$\frac{2}{t}\left(\frac{j(v + tz) - j(v)}{t} - \langle \varphi, z\rangle\right) \geq 0 \quad \forall z \in V,\ \ \forall t \in \mathbb{R}^+,\ \ \forall (v, \varphi) \in \mathrm{graph}(\partial j). \tag{1.23}$$

The functional $Q_j^{v,\varphi}$ is thus indeed non-negative for all $(v, \varphi) \in \mathrm{graph}(\partial j)$. From the variational inequality (P), we further obtain that $f - A(w) \in \partial j(w)$. The latter implies that the j_t-terms in (1.20) have exactly the same structure as the expressions in the limes inferior on the right-hand side of (1.22)

and demonstrates that it makes sense to study the quantity $Q_j^{v,\varphi}$ when we try to pass to the limit $t \searrow 0$ in the variational inequality for the difference quotients δ_t.

It should be noted that the fractions in (1.22) (and (1.23), respectively) can be interpreted as second-order difference quotients in which the (possibly non-existent) first derivative $j'(v) \in V^*$ is replaced with a subgradient $\varphi \in \partial j(v)$. This shows that Definition 1.3.1 generalizes the classical notion of (directional) second derivative to arbitrary convex functions and explains why we use the term "second subderivative" for the functional $Q_j^{v,\varphi} : V \to [0, \infty]$. Compare also with the results of Section 2.1 in this context, and in particular with Corollary 2.1.2 which yields that the second subderivative of a convex C^2-function $j : V \to \mathbb{R}$ is given by

$$Q_j^{v,\varphi}(z) = j''(v)z^2 \in \mathbb{R} \qquad \forall z \in V,$$

where $j''(v) : V \times V \to \mathbb{R}$ denotes the second Fréchet derivative at a point $v \in V$, where $\varphi := j'(v)$ and where $j''(v)z^2$ is short for $j''(v)(z, z)$ for all $z \in V$.

We would like to point out that the construction in Definition 1.3.1 is not the only way to define a notion of second (sub)derivative for an arbitrary convex function. In fact, there are various competing approaches to generalized second derivatives, each with its own advantages and disadvantages. See, e.g., [Rockafellar, 1990], [Rockafellar and Wets, 1998, Section 13] and the references therein. The notion of second subderivative that we employ in this thesis has, at least to the author's best knowledge, first been introduced in [Rockafellar, 1985] for the study of convex functions on the Euclidean space. It was later extended to general Hilbert spaces in [Do, 1992] and has since appeared in numerous works on the (sensitivity) analysis of optimization problems and metric projections (although under different names and with slightly varying definitions, see, e.g., [Adly and Bourdin, 2017; Borwein and Noll, 1994; Christof and Wachsmuth, 2017b; Levy, 1999; Noll, 1995]). We remark that it is possible to identify the epigraph of the functional $Q_j^{v,\varphi}$ with an appropriately defined Kuratowski limit of the sets $\operatorname{epi}(j_t)$. Details on this topic can be found in [Do, 1992, Section 1].

Before we continue with the analysis of the variational inequality (1.20), we collect some elementary facts about the second subderivative in Definition 1.3.1. Our first result concerns the homogeneity properties of the functional $Q_j^{v,\varphi}$.

Lemma 1.3.2. *The second subderivative $Q_j^{v,\varphi} : V \to [0, \infty]$ is positively homogeneous of degree two for all $(v, \varphi) \in \operatorname{graph}(\partial j)$, i.e.,*

$$Q_j^{v,\varphi}(\alpha z) = \alpha^2 Q_j^{v,\varphi}(z) \quad \forall z \in V, \ \forall \alpha > 0, \ \forall (v, \varphi) \in \operatorname{graph}(\partial j).$$

Moreover, it holds $Q_j^{v,\varphi}(0) = 0$ for all $(v, \varphi) \in \operatorname{graph}(\partial j)$.

Proof. Consider an arbitrary but fixed tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ and some $z \in V$, $\alpha > 0$. Then, for all sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset V$ with $t_n \searrow 0$ and $z_n \rightharpoonup z$, we may define $\tilde{t}_n := t_n/\alpha$, $\tilde{z}_n := \alpha z_n$ to obtain sequences $\{\tilde{t}_n\} \subset \mathbb{R}^+$, $\{\tilde{z}_n\} \subset V$ with $\tilde{t}_n \searrow 0$, $\tilde{z}_n \rightharpoonup \alpha z$ and

$$\liminf_{n\to\infty} \frac{2}{\tilde{t}_n} \left(\frac{j(v + \tilde{t}_n \tilde{z}_n) - j(v)}{\tilde{t}_n} - \langle \varphi, \tilde{z}_n \rangle \right) = \alpha^2 \liminf_{n\to\infty} \frac{2}{t_n} \left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle \right).$$

Taking the infimum over all sequences $\{t_n\}$, $\{z_n\}$, we obtain $Q_j^{v,\varphi}(\alpha z) \leq \alpha^2 Q_j^{v,\varphi}(z)$. Since z and α were arbitrary, the positive homogeneity of degree two follows immediately. The identity $Q_j^{v,\varphi}(0) = 0$ for all $(v, \varphi) \in \operatorname{graph}(\partial j)$ is trivial. □

Lemma 1.3.2 implies in particular that the domain of the second subderivative is always a pointed cone (where pointed means that the cone contains the origin). In what follows, we will call this cone the reduced critical cone $\mathcal{K}_j^{red}(v, \varphi)$.

Definition 1.3.3 (Reduced Critical Cone)**.** *Let $(v, \varphi) \in \operatorname{graph}(\partial j)$ be arbitrary but fixed. The set*

$$\mathcal{K}_j^{red}(v, \varphi) := \operatorname{dom}\left(Q_j^{v,\varphi}\right) = \left\{ z \in V \;\middle|\; Q_j^{v,\varphi}(z) < \infty \right\} \tag{1.24}$$

is called the reduced critical cone associated with the tuple (v, φ).

Using standard results from convex analysis, we can prove:

Lemma 1.3.4. *For every $(v,\varphi) \in \operatorname{graph}(\partial j)$, it holds $\mathcal{K}_j^{red}(v,\varphi) \subset \mathcal{T}_K(v)$. If, moreover, the tuple $(v,\varphi) \in \operatorname{graph}(\partial j)$ is such that $j|_K : K \to \mathbb{R}$ is Hadamard directionally differentiable in v in all directions $z \in \mathcal{T}_K(v)$, then it is true that*

$$\mathcal{K}_j^{red}(v,\varphi) \subset \big\{ z \in \mathcal{T}_K(v) \;\big|\; j'(v;z) = \langle \varphi, z \rangle \big\}. \tag{1.25}$$

Proof. Let $(v,\varphi) \in \operatorname{graph}(\partial j)$ be arbitrary but fixed and assume that a $z \in \mathcal{K}_j^{red}(v,\varphi)$ is given. From the definition of $\mathcal{K}_j^{red}(v,\varphi)$, we obtain that there exist sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset V$ with

$$t_n \searrow 0, \quad z_n \rightharpoonup z, \quad 0 \leq \liminf_{n\to\infty} \frac{2}{t_n} \left(\frac{j(v+t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle \right) \leq Q_j^{v,\varphi}(z) + 1 < \infty. \tag{1.26}$$

By passing over to a subsequence, we may assume w.l.o.g. that $v + t_n z_n \in K$ holds for all $n \in \mathbb{N}$. The latter implies in combination with (1.1) that $z \in \mathcal{T}_K(v)$. Consequently, $\mathcal{K}_j^{red}(v,\varphi) \subset \mathcal{T}_K(v)$ and the first claim is proved. Suppose now that $j|_K : K \to \mathbb{R}$ is Hadamard directionally differentiable in v in all directions $z \in \mathcal{T}_K(v)$. Then, Lemma 1.1.3 yields that the function $j'(v;\cdot) : \mathcal{T}_K(v) \to \mathbb{R}$ is continuous, and we obtain from (1.1) that for all $z_1, z_2 \in \mathcal{T}_K(v)$ and all $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$, we can find sequences $\{z_{1,n}\} \subset V$, $\{z_{2,n}\} \subset V$ with $v + t_n z_{1,n} \in K$, $v + t_n z_{2,n} \in K$, $z_{1,n} \to z_1$ and $z_{2,n} \to z_2$. From the latter and the convexity of j, it follows that

$$\begin{aligned} j'(v; \alpha z_1 + (1-\alpha) z_2) &= \lim_{n\to\infty} \frac{j(v + t_n \alpha z_{1,n} + t_n (1-\alpha) z_{2,n}) - j(v)}{t_n} \\ &\leq \alpha j'(v; z_1) + (1-\alpha) j'(v; z_2) \end{aligned}$$

holds for all $z_1, z_2 \in \mathcal{T}_K(v)$ and all $\alpha \in [0,1]$. This shows that the function $j'(v;\cdot) : \mathcal{T}_K(v) \to \mathbb{R}$ is also convex. Consider now again an arbitrary but fixed $z \in \mathcal{K}_j^{red}(v,\varphi)$ and assume that $\{t_n\} \subset \mathbb{R}^+$ and $\{z_n\} \subset V$ are sequences satisfying (1.26) and $v + t_n z_n \in K$ for all $n \in \mathbb{N}$. Then, the properties of j and $j'(v;\cdot)$ and the weak lower semicontinuity of convex and continuous functions (see the lemma of Mazur) imply

$$\begin{aligned} 0 &\geq \liminf_{n\to\infty} \left(\frac{j(v+t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle \right) \\ &\geq \liminf_{n\to\infty} \big(j'(v; z_n) - \langle \varphi, z_n \rangle \big) \\ &\geq j'(v;z) - \langle \varphi, z \rangle \\ &\geq 0. \end{aligned}$$

The above yields (1.25) as claimed. □

Recall that, in optimization, the term "critical cone" is typically used for the set of all directions that satisfy a given first-order optimality condition with equality (see, e.g., [Bonnans and Shapiro, 2000, Section 3.1.1] and [Wachsmuth, 2016, Section 2]). Consider, for example, the model problem

$$\min_{u\in K} j(u) - \langle \varphi, u \rangle, \tag{1.27}$$

where φ satisfies $\varphi \in \partial j(v)$ for some arbitrary but fixed $v \in \operatorname{dom}(j)$, and suppose for the moment that the function $j|_K : K \to \mathbb{R}$ is Hadamard directionally differentiable. In this situation, v is trivially first-order stationary for (1.27) in the sense that

$$v \in K, \quad j'(v;z) - \langle \varphi, z \rangle \geq 0 \quad \forall z \in \mathcal{T}_K(v),$$

and the critical cone associated with v takes the form

$$\mathcal{T}_K^{crit}(v) := \big\{ z \in \mathcal{T}_K(v) \;\big|\; j'(v;z) = \langle \varphi, z \rangle \big\}. \tag{1.28}$$

Note that the set in (1.28) is precisely the right-hand side of (1.25). This shows that, for a Hadamard directionally differentiable j, the domain $\mathrm{dom}(Q_j^{v,\varphi})$ is always contained in the critical cone $\mathcal{T}_K^{crit}(v)$ of the prototypical minimization problem (1.27) and that it is sensible to refer to the set in (1.24) as the "reduced critical cone". Compare also with (1.4) and the comments after Definition 1.3.1 in this context.

We would like to point out that the set $\mathcal{K}_j^{red}(v,\varphi)$ is typically not closed w.r.t. the norm $\|\cdot\|_V$ (in contrast to the critical, the normal and the tangent cone), and that the inclusion in (1.25) is in general strict. See, e.g., Chapters 4 and 5 for examples.

Having introduced and studied the notion of second subderivative, we are now in the position to prove a first result on the differentiability properties of the solution operator $S : V^* \to V$ associated with the variational inequality (P):

Proposition 1.3.5 (Necessary Condition for Directional Differentiability)**.** *Suppose that $f, g \in V^*$ are given such that the solution operator $S : V^* \to V$ associated with* (P) *is directionally differentiable in f in the direction g with directional derivative $\delta := S'(f;g)$. Define $w := S(f)$ and $\varphi := f - A(w)$. Then, the following holds true:*

(i) The derivative δ is uniquely characterized by the variational inequality

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w,\varphi), \\ &\left\langle A'(w)\delta, z-\delta\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi). \end{aligned} \tag{1.29}$$

(ii) It holds

$$Q_j^{w,\varphi}(\delta) = \langle g, \delta\rangle - \left\langle A'(w)\delta, \delta\right\rangle. \tag{1.30}$$

(iii) For every sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$, the difference quotients $\delta_n := \delta_{t_n}$ satisfy

$$\delta_n \to \delta, \qquad Q_j^{w,\varphi}(\delta) = \lim_{n\to\infty} \frac{2}{t_n}\left(\frac{j(w+t_n\delta_n)-j(w)}{t_n} - \langle \varphi, \delta_n\rangle\right). \tag{1.31}$$

Proof. Suppose that S is directionally differentiable in $f \in V^*$ in the direction $g \in V^*$, i.e., that the difference quotients

$$\delta_t = \frac{S(f+tg)-S(f)}{t}, \quad t > 0,$$

converge strongly to some $\delta = S'(f;g)$ as $t \searrow 0$, and assume that a $z \in \mathcal{K}_j^{red}(w,\varphi)$ is given. From the definitions of the quantities $\mathcal{K}_j^{red}(w,\varphi)$ and $Q_j^{w,\varphi}(z)$, it follows that for each arbitrary but fixed $\varepsilon > 0$ we can find sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset V$ with $t_n \searrow 0$, $z_n \rightharpoonup z$, $w + t_n z_n \in K$ and

$$\frac{1}{2}Q_j^{w,\varphi}(z) \leq \liminf_{n\to\infty} \frac{1}{t_n}\left(\frac{j(w+t_nz_n)-j(w)}{t_n} - \langle \varphi, z_n\rangle\right) \leq \frac{1}{2}Q_j^{w,\varphi}(z) + \varepsilon \in \mathbb{R}.$$

If we use the above $\{z_n\}$, $\{t_n\}$ in (1.20), then we obtain

$$\begin{aligned} \left\langle A'(w)\delta_n, z_n-\delta_n\right\rangle &+ \frac{1}{t_n}\left(\frac{j(w+t_nz_n)-j(w)}{t_n} - \langle \varphi, z_n\rangle\right) \\ &- \frac{1}{t_n}\left(\frac{j(w+t_n\delta_n)-j(w)}{t_n} - \langle \varphi, \delta_n\rangle\right) \geq \langle g, z_n-\delta_n\rangle - r_{t_n}(g)\|z_n-\delta_n\|_V, \end{aligned} \tag{1.32}$$

where $\delta_n := \delta_{t_n}$ denotes the sequence of difference quotients associated with $\{t_n\}$. Taking the limes inferior for $n \to \infty$ in (1.32) yields

$$\begin{aligned} \left\langle A'(w)\delta, z-\delta\right\rangle &+ \frac{1}{2}Q_j^{w,\varphi}(z) + \varepsilon \\ &- \frac{1}{2}\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(w+t_n\delta_n)-j(w)}{t_n} - \langle \varphi, \delta_n\rangle\right) \geq \langle g, z-\delta\rangle. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary and because of Definition 1.3.1, we may deduce

$$\left\langle A'(w)\delta, z-\delta\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi).$$

If we test the above variational inequality with $z = 0 \in \mathcal{K}_j^{red}(w,\varphi)$, then it follows from Lemma 1.3.2 and the non-negativity of $Q_j^{w,\varphi}$ that

$$0 \leq \frac{1}{2}Q_j^{w,\varphi}(\delta) \leq \langle g, \delta\rangle - \left\langle A'(w)\delta, \delta\right\rangle < \infty.$$

This shows that $\delta \in \mathcal{K}_j^{red}(w,\varphi)$ and that δ is indeed a solution to (1.29). To complete the proof of part (i) of the proposition, we note that (1.21) and Lemma 1.3.4 yield

$$\left\langle A'(w)(z_1-z_2), z_1-z_2\right\rangle \geq c\|z_1-z_2\|_V^2 \quad \forall z_1, z_2 \in \mathcal{K}_j^{red}(w,\varphi). \tag{1.33}$$

The above implies that (1.29) can have at most one solution (cf. the argumentation with (1.10) and (1.11) in Section 1.2) and that δ is indeed uniquely characterized by the variational inequality (1.29). This proves the first claim.

To obtain (ii), we use the test function $z = s\delta \in \mathcal{K}_j^{red}(w,\varphi)$, $s > 0$, in (1.29) (recall that $\mathcal{K}_j^{red}(w,\varphi)$ is a cone). This yields

$$(s-1)\left\langle A'(w)\delta, \delta\right\rangle + \frac{1}{2}(s^2-1)Q_j^{w,\varphi}(\delta) \geq (s-1)\langle g, \delta\rangle$$

and, consequently,

$$\begin{aligned}\left\langle A'(w)\delta, \delta\right\rangle + \frac{1}{2}(s+1)Q_j^{w,\varphi}(\delta) &\geq \langle g, \delta\rangle \quad \forall s > 1,\\ \left\langle A'(w)\delta, \delta\right\rangle + \frac{1}{2}(s+1)Q_j^{w,\varphi}(\delta) &\leq \langle g, \delta\rangle \quad \forall s < 1.\end{aligned}$$

Letting $s \searrow 1$ and $s \nearrow 1$ in the above, (1.30) follows immediately.

It remains to prove part (iii). To this end, we consider an arbitrary but fixed $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$. Denote the difference quotients associated with the sequence $\{st_n\}$ with $\delta_n(s)$ for all $s > 0$, and define

$$Q := Q_j^{w,\varphi}(\delta), \qquad \Theta(s) := \limsup_{n\to\infty} \frac{2}{st_n}\left(\frac{j(w+st_n\delta_n(s)) - j(w)}{st_n} - \langle \varphi, \delta_n(s)\rangle\right), \quad s > 0.$$

Then, we may test the variational inequality (1.20) for $\delta_n(s_1)$ with $s_2\delta_n(s_2)/s_1$ to obtain

$$\begin{aligned}&\left\langle A'(w)\delta_n(s_1), \frac{s_2}{s_1}\delta_n(s_2) - \delta_n(s_1)\right\rangle - \left\langle g, \frac{s_2}{s_1}\delta_n(s_2) - \delta_n(s_1)\right\rangle\\ &\qquad + \left(\frac{s_2}{s_1}\right)^2 \frac{1}{s_2t_n}\left(\frac{j(w+t_ns_2\delta_n(s_2)) - j(w)}{s_2t_n} - \langle \varphi, \delta_n(s_2)\rangle\right)\\ &\qquad\qquad \geq \frac{1}{s_1t_n}\left(\frac{j(w+s_1t_n\delta_n(s_1)) - j(w)}{s_1t_n} - \langle \varphi, \delta_n(s_1)\rangle\right) + o(1) \qquad \forall s_1, s_2 > 0.\end{aligned} \tag{1.34}$$

Taking the limes superior for $n \to \infty$ in (1.34) yields (with $\delta_n(s) \to \delta$ for all s as $n \to \infty$ and (1.30))

$$0 \leq \Theta(s_1) \leq 2\left(1 - \frac{s_2}{s_1}\right)Q + \left(\frac{s_2}{s_1}\right)^2 \Theta(s_2) \quad \forall s_1, s_2 > 0. \tag{1.35}$$

Using the above functional inequality and induction, we will prove that

$$\Theta(s) \leq \left(1 + \frac{1}{m}\right)Q \qquad \forall m \in \mathbb{N}, \quad \forall s > 0. \tag{1.36}$$

Note that by choosing the test function $z = 0$ in the variational inequality (1.20) for $\delta(s)$ and by passing to the limit $n \to \infty$, we obtain that $\Theta(s) \leq 2(\langle g, \delta \rangle - \langle A'(w)\delta, \delta \rangle) = 2Q$ holds for all $s > 0$. This is precisely (1.36) for $m = 1$. For the transition $m \mapsto m + 1$, we use that (1.35) with $s_2 = s_1 m/(m + 1)$ and the induction hypothesis yield

$$\begin{aligned}\Theta(s_1) &\leq 2\left(1 - \frac{m}{m+1}\right)Q + \left(\frac{m}{m+1}\right)^2 \Theta(s_2) \\ &\leq \frac{2}{m+1}Q + \left(\frac{m}{m+1}\right)^2 \left(\frac{m+1}{m}\right) Q \\ &\leq \left(1 + \frac{1}{m+1}\right) Q \quad \forall s_1 > 0.\end{aligned}$$

This completes the induction step and proves that we indeed have $\Theta(s) \leq (1 + 1/m)Q$ for all $m \in \mathbb{N}$ and all $s > 0$. By letting $m \to \infty$ in (1.36), we obtain in particular that $\Theta(s) \leq Q$ holds for all $s > 0$. If we combine the latter with the definitions of Q and $\Theta(s)$, then we arrive at the chain of inequalities

$$Q_j^{w,\varphi}(\delta) \leq \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(w + t_n\delta_n) - j(w)}{t_n} - \langle \varphi, \delta_n \rangle\right) \leq \Theta(1) \leq Q = Q_j^{w,\varphi}(\delta).$$

The assertion in (iii) now follows immediately. □

It should be noted that it is a priori completely unclear whether the functional $Q_j^{w,\varphi} : V \to [0, \infty]$ is convex or lower semicontinuous. This implies in particular that we can, in general, not say anything about the existence of a solution to (1.29) when we consider this variational inequality as a stand-alone problem, i.e., independently of the background of our differential sensitivity analysis. We always know, however, that (1.29) can have at most one solution, cf. the coercivity condition (1.33) and the usual contradiction argument.

Proposition 1.3.5 demonstrates that solutions to variational inequalities of the form (1.29) are the only candidates for the directional derivatives $S'(f; g)$ of S, and that the solution operator to (P) can only be expected to be directionally differentiable if the sequence $j_t(\delta_t)$ in (1.20) converges for $t \searrow 0$. To the author's best knowledge, these necessary conditions for directional differentiability (that are also applicable when only certain directions $g \in V^*$ are considered) have not been documented before.

In practical applications, we are, of course, typically interested in criteria that are sufficient for the directional differentiability of the solution operator S associated with (P), i.e., we would like to have the converse of Proposition 1.3.5. To obtain such a result, we introduce the following concept that is motivated by the recovery condition (1.31).

Definition 1.3.6 (Second-Order Epi-Differentiability)**.** *Suppose that a tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ is given. Then, j is said to be twice epi-differentiable in v for φ in a direction $z \in V$ if for every sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ there exists a sequence $\{z_n\} \subset V$ such that*

$$z_n \to z \quad \textit{and} \quad Q_j^{v,\varphi}(z) = \lim_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle\right). \tag{1.37}$$

If the above holds for all $z \in V$, then we say that j is twice epi-differentiable in v for φ.

Note that part (iii) of Proposition 1.3.5 yields that, if the solution operator $S : V^* \to V$ to (P) is directionally differentiable in a point $f \in V^*$ in a direction $g \in V^*$, then j is necessarily twice epi-differentiable in $w := S(f)$ for $\varphi := f - A(w) \in \partial j(w)$ in the direction $\delta := S'(f; g)$ (and the recovery sequence $\{z_n\}$ for a given $\{t_n\}$ may be chosen as the sequence of difference quotients $\{\delta_{t_n}\}$).

For later use, we mention the following:

Lemma 1.3.7. *Let $(v, \varphi) \in \operatorname{graph}(\partial j)$ be given. Then, j is twice epi-differentiable in v for φ if and only if j is twice epi-differentiable in v for φ in all directions $z \in \mathcal{K}_j^{red}(v, \varphi)$.*

Proof. For every $z \in V$ with $Q_j^{v,\varphi}(z) = \infty$ and every $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$, it holds

$$\begin{aligned}\infty &\geq \limsup_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z)-j(v)}{t_n} - \langle \varphi, z\rangle\right)\\ &\geq \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z)-j(v)}{t_n} - \langle \varphi, z\rangle\right)\\ &\geq Q_j^{v,\varphi}(z)\\ &= \infty.\end{aligned}$$

The above shows that recovery sequences with the properties in (1.37) can trivially be found for all $z \in V$ with $Q_j^{v,\varphi}(z) = \infty$ and that we can indeed ignore all $z \in V \setminus \mathcal{K}_j^{red}(v,\varphi)$ when we check the condition of second-order epi-differentiability in v for φ. □

As the notion of second subderivative, the concept of second-order epi-differentiability essentially goes back to [Rockafellar, 1985] in finite dimensions and [Do, 1992] in infinite dimensions. Compare also with [Adly and Bourdin, 2017; Borwein and Noll, 1994; Hintermüller and Surowiec, 2017; Levy, 1999; Noll, 1995] and the comments after Theorem 1.4.1 in this context.

We remark that recovery conditions similar to that in Definition 1.3.6 appear very naturally in many different branches of optimization. See, e.g., [Christof and Wachsmuth, 2017b] and Section 6.2 for applications in fields other than the sensitivity analysis of elliptic variational inequalities of the first and the second kind. It should be noted that the property of second-order epi-differentiability is closely related to the (probably more prominent) concept of Mosco epi-convergence. Recall that this notion of convergence is defined as follows (cf. [Attouch, 1984, Definition 3.17]):

Definition 1.3.8. *A family of extended real-valued functions $F_t : V \to [0,\infty]$, $t > 0$, is said to be Mosco epi-convergent to some $F : V \to [0,\infty]$ for $t \searrow 0$ if for every sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ and every $z \in V$ it holds*

$$\forall \{z_n\} \subset V \text{ with } z_n \rightharpoonup z: \ F(z) \leq \liminf_{n\to\infty} F_{t_n}(z_n), \tag{1.38}$$

$$\exists \{z_n\} \subset V \text{ with } z_n \to z: \ F(z) \geq \limsup_{n\to\infty} F_{t_n}(z_n). \tag{1.39}$$

From the above, we immediately obtain:

Proposition 1.3.9. *Suppose that a tuple $(v,\varphi) \in \operatorname{graph}(\partial j)$ is given. Then, j is twice epi-differentiable in v for φ if and only if the second-order difference quotient functions*

$$j_t : V \to [0,\infty], \qquad j_t(z) := \frac{2}{t}\left(\frac{j(v+tz)-j(v)}{t} - \langle \varphi, z\rangle\right),$$

are Mosco epi-convergent to the second subderivative $Q_j^{v,\varphi}$ for $t \searrow 0$.

Proof. Assume that j is twice epi-differentiable in v for φ. Then, Definitions 1.3.1 and 1.3.6 immediately yield that j_t and $Q_j^{v,\varphi}$ satisfy (1.38) and (1.39). If, conversely, we know that the functions j_t are Mosco epi-convergent to the second subderivative $Q_j^{v,\varphi}$ for $t \searrow 0$, then the second-order epi-differentiability of j follows straightforwardly from (1.38) and (1.39). □

We are now in the position to prove:

Proposition 1.3.10 (Sufficient Condition for Directional Differentiability)**.** *Let $f, g \in V^*$ be given and set $w := S(f)$, $\varphi := f - A(w)$. Let $\{\delta_t\}$ be the family of difference quotients defined in* (1.18)*, and suppose that there exists a $D \subset V$ such that D contains all weak accumulation points of $\{\delta_t\}$ for $t \searrow 0$ and such that j is twice epi-differentiable in w for φ in all directions $z \in D$. Then, the solution operator $S : V^* \to V$ associated with* (P) *is Hadamard directionally differentiable in f in the direction g.*

Proof. Recall that the difference quotients $\{\delta_t\}$ satisfy $\|\delta_t\|_V \leq \|g\|_{V^*}/c$ for all $t > 0$ and

$$\begin{aligned}\left\langle A'(w)\delta_t, z - \delta_t\right\rangle + \frac{1}{t}\left(\frac{j(w+tz)-j(w)}{t} - \langle \varphi, z\rangle\right)& \\ - \frac{1}{t}\left(\frac{j(w+t\delta_t)-j(w)}{t} - \langle \varphi, \delta_t\rangle\right) \geq \langle g, z-\delta_t\rangle - r_t(g)\|z-\delta_t\|_V& \\ \forall z \in \frac{1}{t}(K-w).&\end{aligned} \tag{1.40}$$

Consider now an arbitrary but fixed sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$. Then, the reflexivity of V and the uniform boundedness of the difference quotients $\{\delta_t\}$ yield that we may pass over to a subsequence of $\{t_n\}$ (unrelabeled) such that $\delta_n := \delta_{t_n} \rightharpoonup \delta$ holds for some $\delta \in D$. Note that (1.21), (1.23) (with $v = w$, $\varphi = f - A(w)$ and $z = \delta_t$) and (1.40) (with $z = 0$) imply

$$\langle g, \delta_n\rangle \geq \langle g, \delta_n\rangle - \left\langle A'(w)\delta_n, \delta_n\right\rangle \geq \frac{1}{t_n}\left(\frac{j(w+t_n\delta_n)-j(w)}{t_n} - \langle \varphi, \delta_n\rangle\right) + o(1) \geq o(1).$$

Taking the $\liminf$ on the left- and the right-hand side of the above inequality yields $Q_j^{w,\varphi}(\delta) < \infty$, i.e., the weak limit δ of our sequence δ_n is an element of the reduced critical cone $\mathcal{K}_j^{red}(w,\varphi)$. Suppose now that a $z \in \mathcal{K}_j^{red}(w,\varphi) \cap D$ is given. Then, the second-order epi-differentiability of j in w for φ in the direction z implies that we can find a sequence $\{z_n\} \subset V$ such that

$$z_n \to z, \quad Q_j^{w,\varphi}(z) = \lim_{n\to\infty} \frac{2}{t_n}\left(\frac{j(w+t_nz_n)-j(w)}{t_n} - \langle \varphi, z_n\rangle\right).$$

Note that we may assume w.l.o.g. that $z_n \in \frac{1}{t_n}(K-w)$ holds for all n (since there can only be finitely many n with $w + t_nz_n \notin K = \operatorname{dom}(j)$ due to the convergence to $Q_j^{w,\varphi}(z) \in \mathbb{R}$, and since these finitely many elements can be changed to zero without losing the convergence properties). If we use the sequence $\{z_n\}$ in (1.40) (with $t = t_n$), then we arrive at the inequality

$$\begin{aligned}\langle g, \delta_n - z_n\rangle + \left\langle A'(w)\delta_n, z_n\right\rangle + \frac{1}{t_n}\left(\frac{j(w+t_nz_n)-j(w)}{t_n} - \langle \varphi, z_n\rangle\right)& \\ \geq \frac{1}{t_n}\left(\frac{j(w+t_n\delta_n)-j(w)}{t_n} - \langle \varphi, \delta_n\rangle\right) + o(1) + \left\langle A'(w)\delta_n, \delta_n\right\rangle.&\end{aligned}$$

Using the properties of the sequence z_n, the weak convergence $\delta_n \rightharpoonup \delta$, Definition 1.3.1, and the weak lower semicontinuity of the map $\mathcal{T}_K(w) \ni z \mapsto \langle A'(w)z, z\rangle \in \mathbb{R}$ (which follows from the continuity and the convexity), we may pass to the limit in the above to obtain

$$\begin{aligned}\langle g, \delta - z\rangle + \left\langle A'(w)\delta, z\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) &\geq \frac{1}{2}Q_j^{w,\varphi}(\delta) + \limsup_{n\to\infty}\left\langle A'(w)\delta_n, \delta_n\right\rangle \\ &\geq \frac{1}{2}Q_j^{w,\varphi}(\delta) + \liminf_{n\to\infty}\left\langle A'(w)\delta_n, \delta_n\right\rangle \\ &\geq \frac{1}{2}Q_j^{w,\varphi}(\delta) + \left\langle A'(w)\delta, \delta\right\rangle \qquad \forall z \in \mathcal{K}_j^{red}(w,\varphi)\cap D.\end{aligned} \tag{1.41}$$

The above chain of inequalities has several consequences: First, choosing $z = \delta$ in (1.41) yields

$$\left\langle A'(w)\delta, \delta\right\rangle \geq \limsup_{n\to\infty}\left\langle A'(w)\delta_n, \delta_n\right\rangle \geq \liminf_{n\to\infty}\left\langle A'(w)\delta_n, \delta_n\right\rangle \geq \left\langle A'(w)\delta, \delta\right\rangle$$

and, consequently,

$$\begin{aligned}c\|\delta - \delta_n\|_V^2 &\leq \left\langle A'(w)(\delta-\delta_n), \delta-\delta_n\right\rangle \\ &= \left\langle A'(w)\delta, \delta\right\rangle - \left\langle A'(w)\delta_n, \delta\right\rangle - \left\langle A'(w)\delta, \delta_n\right\rangle + \left\langle A'(w)\delta_n, \delta_n\right\rangle \\ &\to 0.\end{aligned}$$

This proves that the sequence δ_n converges even strongly to δ. Second, (1.41) implies that the limit δ of the difference quotients δ_n is a solution of the problem

$$\begin{aligned}&\delta \in \mathcal{K}_j^{red}(w,\varphi) \cap D,\\ &\big\langle A'(w)\delta, z-\delta\big\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi)\cap D.\end{aligned}$$

Since the above variational inequality can have at most one solution (by the usual contradiction argument with (1.21)), it follows that the weak limit δ does not depend on the (sub)sequence $\{t_n\}$ that we started with. We have thus proven the following: Every sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ contains a subsequence $\{t_{n_m}\}$ such that the associated difference quotients $\delta_{t_{n_m}}$ converge strongly in V to a uniquely determined $\delta \in V$. Using contradiction, it now follows straightforwardly that the difference quotients δ_t converge to δ as $t \searrow 0$. This proves the directional differentiability of S in f in the direction g. The Hadamard directional differentiability follows immediately from the Lipschitz estimate (1.6) and the triangle inequality (cf. [Bonnans and Shapiro, 2000, Proposition 2.49]). This completes the proof. □

In practical applications, the set D in Proposition 1.3.10 can be used to exploit additional information about the properties of the difference quotients $\{\delta_t\}$. Note that, the smaller the set D, i.e., the more a priori knowledge about the accumulation points of the family $\{\delta_t\}$ for $t \searrow 0$ is available, the less elements the condition of second-order epi-differentiability has to be checked for. If we do not know anything about the limiting behavior of $\{\delta_t\}$ beforehand, then we can, of course, always choose D to be the whole space V. This choice leads to:

Corollary 1.3.11. *Let $f \in V^*$ be given and set $w := S(f)$, $\varphi := f - A(w)$. Assume that j is twice epi-differentiable in w for φ. Then, the solution operator $S : V^* \to V$ associated with* (P) *is Hadamard directionally differentiable in f in all directions $g \in V^*$ and the directional derivative $\delta := S'(f;g)$ in f in a direction g is uniquely characterized by the variational inequality*

$$\begin{aligned}&\delta \in \mathcal{K}_j^{red}(w,\varphi),\\ &\big\langle A'(w)\delta, z-\delta\big\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi).\end{aligned} \tag{1.42}$$

Proof. The claim follows immediately from Propositions 1.3.5 and 1.3.10. □

Note that, as a byproduct of our sensitivity analysis, we obtain that (1.42) admits a unique solution for all $g \in V^*$ in the situation of Corollary 1.3.11. This makes sense as the following lemma illustrates:

Lemma 1.3.12. *Assume that a tuple $(v,\varphi) \in \operatorname{graph}(\partial j)$ is given and that j is twice epi-differentiable in v for φ. Then, the second subderivative $Q_j^{v,\varphi} : V \to [0,\infty]$ is a proper, convex and lower semicontinuous functional and $\mathcal{K}_j^{red}(v,\varphi)$ is a convex, non-empty and pointed cone.*

Proof. Let $z_1, z_2 \in \mathcal{K}_j^{red}(v,\varphi)$ be given, and let $\{t_n\} \subset \mathbb{R}^+$ be an arbitrary but fixed sequence with $t_n \searrow 0$. Let $\{z_{1,n}\} \subset V, \{z_{2,n}\} \subset V$ be recovery sequences for z_1 and z_2 w.r.t. $\{t_n\}$ as in the definition of the second-order epi-differentiability, and let $\alpha \in [0,1]$ be arbitrary but fixed. Then, we may use the convexity of j to compute:

$$\begin{aligned}&Q_j^{v,\varphi}(\alpha z_1 + (1-\alpha)z_2)\\ &\quad\leq \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n(\alpha z_{1,n}+(1-\alpha)z_{2,n})) - j(v)}{t_n} - \langle \varphi, \alpha z_{1,n} + (1-\alpha)z_{2,n}\rangle\right)\\ &\quad\leq \liminf_{n\to\infty} \left(\frac{2\alpha}{t_n}\left(\frac{j(v+t_n z_{1,n}) - j(v)}{t_n} - \langle \varphi, z_{1,n}\rangle\right)\right.\\ &\qquad\qquad \left. + \frac{2(1-\alpha)}{t_n}\left(\frac{j(v+t_n z_{2,n}) - j(v)}{t_n} - \langle \varphi, z_{2,n}\rangle\right)\right)\\ &\quad= \alpha Q_j^{v,\varphi}(z_1) + (1-\alpha)Q_j^{v,\varphi}(z_2) < \infty.\end{aligned}$$

This proves that the second subderivative $Q_j^{v,\varphi}$ and its domain, the reduced critical cone $\mathcal{K}_j^{red}(v,\varphi)$, are convex. Consider now a sequence $\{z_n\} \subset \mathcal{K}_j^{red}(v,\varphi)$ with $z_n \to z$ for some $z \in V$ and let $\{t_m\} \subset \mathbb{R}^+$ be a sequence with $t_m \searrow 0$. Then, for each n we can find a recovery sequence $\{z_{n,m}\} \subset V$ as in the definition of the second-order epi-differentiability. Choose a strictly increasing sequence $\{m_n\}$ such that

$$\|z_n - z_{n,m_n}\|_V + \left| \frac{2}{t_{m_n}} \left(\frac{j(v + t_{m_n} z_{n,m_n}) - j(v)}{t_{m_n}} - \langle \varphi, z_{n,m_n} \rangle \right) - Q_j^{v,\varphi}(z_n) \right| \leq \frac{1}{n} \quad \forall n \in \mathbb{N}.$$

Then, the sequences $\tilde{t}_n := t_{m_n}$ and $\tilde{z}_n := z_{n,m_n}$ satisfy $\tilde{t}_n \searrow 0$, $\tilde{z}_n \to z$ and

$$Q_j^{v,\varphi}(z) \leq \liminf_{n\to\infty} \frac{2}{\tilde{t}_n} \left(\frac{j(v + \tilde{t}_n \tilde{z}_n) - j(v)}{\tilde{t}_n} - \langle \varphi, \tilde{z}_n \rangle \right) = \liminf_{n\to\infty} Q_j^{v,\varphi}(z_n).$$

This proves the lower semicontinuity. The remaining claims are trivial (cf. Lemma 1.3.2). □

As Lemma 1.3.12 shows, the second subderivative $Q_j^{v,\varphi} : V \to [0,\infty]$ is a proper, convex and lower semicontinuous function if j is twice epi-differentiable in v for φ. This implies that, under the assumption of second-order epi-differentiability of j in $w := S(f)$ for $\varphi := f - A(w)$, the solvability of the variational inequality

$$\delta \in \mathcal{K}_j^{red}(w,\varphi), \quad \big\langle A'(w)\delta, z - \delta \big\rangle + \frac{1}{2} Q_j^{w,\varphi}(z) - \frac{1}{2} Q_j^{w,\varphi}(\delta) \geq \langle g, z - \delta \rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi)$$

can be proved with Theorem 1.2.2 and that the existence of a solution to (1.42) is, in fact, granted in the situation of Corollary 1.3.11.

We point out that, in practice, it is typically hard to check whether a given function j is twice epi-differentiable in a point v for some $\varphi \in \partial j(v)$ or not. The following lemma is helpful in this context not only for practical applications but also for theoretical considerations:

Lemma 1.3.13 (Sufficient Criterion for Second-Order Epi-Differentiability)**.** *Let $(v,\varphi) \in \operatorname{graph}(\partial j)$ be arbitrary but fixed. Suppose that sets $\mathcal{Z}$ and $\mathcal{K}$ and a function $Q : \mathcal{K} \to [0,\infty)$ are given such that:*

(i) $\mathcal{Z} \subset \mathcal{K}_j^{red}(v,\varphi) \subset \mathcal{K}$,

(ii) for each $z \in \mathcal{K}$, we have $Q_j^{v,\varphi}(z) \geq Q(z)$,

(iii) for each $z \in \mathcal{Z}$ and each $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ there exists a sequence $\{z_n\} \subset V$ satisfying

$$z_n \to z \quad \text{and} \quad \frac{2}{t_n} \left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle \right) \to Q(z) \tag{1.43}$$

for $n \to \infty$,

(iv) for each $z \in \mathcal{K}$ there exists a sequence $\{z_n\} \subset \mathcal{Z}$ with $z_n \to z$ and

$$Q(z) \geq \liminf_{n\to\infty} Q(z_n).$$

Then, it holds $\mathcal{K} = \mathcal{K}_j^{red}(v,\varphi)$, $Q = Q_j^{v,\varphi}$ and j is twice epi-differentiable in v for φ.

Proof. From Definition 1.3.1 and the properties of Q, we immediately obtain that $Q(z) = Q_j^{v,\varphi}(z)$ holds for all $z \in \mathcal{Z}$ and that j is twice epi-differentiable in v for φ in all directions $z \in \mathcal{Z}$. Consider now an arbitrary but fixed $z \in \mathcal{K}_j^{red}(v,\varphi) \setminus \mathcal{Z} \subset \mathcal{K}$ and some $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$. Then, our assumptions imply that there exist $z_n \in \mathcal{Z}$ with $z_n \to z$ for $n \to \infty$ and

$$Q(z) \geq \liminf_{n\to\infty} Q(z_n).$$

Choose sequences $\{z_{n,m}\} \subset V$ such that $v + t_m z_{n,m} \in K$ holds for all n, m and such that

$$z_{n,m} \to z_n \quad \text{and} \quad \frac{2}{t_m}\left(\frac{j(v+t_m z_{n,m}) - j(v)}{t_m} - \langle \varphi, z_{n,m}\rangle\right) \to Q(z_n)$$

holds for all n as $m \to \infty$ (this is possible due to the properties of $\mathcal{Z}$). Select further a strictly increasing sequence $\{m_n\}$ such that

$$\|z_n - z_{n,m_n}\|_V + \left|\frac{2}{t_{m_n}}\left(\frac{j(v+t_{m_n} z_{n,m_n}) - j(v)}{t_{m_n}} - \langle \varphi, z_{n,m_n}\rangle\right) - Q(z_n)\right| \leq \frac{1}{n} \quad \forall n \in \mathbb{N}.$$

Then, it trivially holds $t_{m_n} \searrow 0$ and $z_{m_n} \to z$ for $n \to \infty$, and we obtain

$$\begin{aligned} Q_j^{v,\varphi}(z) &\leq \liminf_{n\to\infty} \frac{2}{t_{m_n}}\left(\frac{j(v+t_{m_n} z_{n,m_n}) - j(v)}{t_{m_n}} - \langle \varphi, z_{n,m_n}\rangle\right) \\ &= \liminf_{n\to\infty} Q(z_n) \\ &\leq Q(z) \\ &\leq Q_j^{v,\varphi}(z). \end{aligned}$$

The above proves, on the one hand, that $Q = Q_j^{v,\varphi}$ holds on $\mathcal{K}_j^{red}(v,\varphi)$ and, on the other hand, that, given an arbitrary but fixed $z \in \mathcal{K}_j^{red}(v,\varphi)$ and a sequence $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$, we can always find a subsequence $\{t_{m_n}\}$ such that there exist $z_{m_n} \in V$ with

$$z_{m_n} \to z \quad \text{and} \quad \frac{2}{t_{m_n}}\left(\frac{j(v+t_{m_n} z_{m_n}) - j(v)}{t_{m_n}} - \langle \varphi, z_{m_n}\rangle\right) \to Q_j^{v,\varphi}(z)$$

as $n \to \infty$. Assume now that the function j is not twice epi-differentiable in v for φ. Then, there exists a $z \in \mathcal{K}_j^{red}(v,\varphi)$ and a sequence $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$ such that there is no sequence $\{z_m\} \subset V$ with

$$z_m \to z \quad \text{and} \quad \frac{2}{t_m}\left(\frac{j(v+t_m z_m) - j(v)}{t_m} - \langle \varphi, z_m\rangle\right) \to Q_j^{v,\varphi}(z)$$

for $m \to \infty$. The above implies that

$$\inf_{u\in V} \|z-u\|_V + \left|Q_j^{v,\varphi}(z) - \frac{2}{t_m}\left(\frac{j(v+t_m u) - j(v)}{t_m} - \langle \varphi, u\rangle\right)\right| \not\to 0$$

as $m \to \infty$. Passing over to a subsequence (unrelabeled), we now obtain a sequence $\{t_m\}$ with $t_m \searrow 0$ and

$$\inf_{u\in V} \|z-u\|_V + \left|Q_j^{v,\varphi}(z) - \frac{2}{t_m}\left(\frac{j(v+t_m u) - j(v)}{t_m} - \langle \varphi, u\rangle\right)\right| > \varepsilon$$

for some $\varepsilon > 0$. From the first part of the proof, however, we know that for this sequence $\{t_m\}$, there exists a subsequence $\{t_{m_n}\}$ such that there are $z_{m_n} \in V$ with

$$z_{m_n} \to z \quad \text{and} \quad \frac{2}{t_{m_n}}\left(\frac{j(v+t_{m_n} z_{m_n}) - j(v)}{t_{m_n}} - \langle \varphi, z_{m_n}\rangle\right) \to Q_j^{v,\varphi}(z)$$

as $n \to \infty$. This is a contradiction. Accordingly, j is twice epi-differentiable in v for φ with $Q = Q_j^{v,\varphi}$ on $\mathcal{K}_j^{red}(v,\varphi)$. It remains to show that $\mathcal{K} = \mathcal{K}_j^{red}(v,\varphi)$. This, however, follows immediately from the assumptions of the lemma, the second-order epi-differentiability and the lower semicontinuity property in Lemma 1.3.12. □

Remark 1.3.14. *In spite of its simplicity, Lemma 1.3.13 is a surprisingly useful and powerful tool in the sensitivity analysis of elliptic variational inequalities of the first and the second kind. It allows to check if a lower estimate $Q : \mathcal{K} \to [0, \infty)$ of the expression on the right-hand side of* (1.22) *is identical to the second subderivative $Q_j^{v,\varphi}$ and demonstrates that it suffices to check the recovery condition in* (1.37) *for a dense subset of the reduced critical cone $\mathcal{K}_j^{red}(v, \varphi)$ to obtain the second-order epi-differentiability of j in a given $v \in V$ for some $\varphi \in \partial j(v)$ (provided $Q_j^{v,\varphi}$ has suitable semicontinuity properties). Compare, e.g., with Theorem 4.3.16 and its applications in Chapters 4 and 5 in this context. We point out that Lemma 1.3.13 covers in particular those situations where j is the characteristic function of a polyhedric or extended polyhedric set (see Section 3.3). Lemma 1.3.13 can further be used, e.g., if a classical second-order Taylor expansion of j is only available in certain directions z but not in others (cf. the concept of two-norms discrepancy in [Tröltzsch, 2010]).*

Using Proposition 1.3.5, Lemma 1.3.12 and Lemma 1.3.13, we can prove that the second-order epi-differentiability of j in $w := S(f)$ for $\varphi := f - A(w)$ is not only sufficient but also necessary for the directional differentiability of the solution operator $S : V^* \to V$ of (P) in f in all directions g.

Proposition 1.3.15. *Let $f \in V^*$ be arbitrary but fixed and set $w := S(f)$, $\varphi := f - A(w)$. Assume that the solution operator $S : V^* \to V$ is directionally differentiable in f in all directions $g \in V^*$. Then, j is twice epi-differentiable in w for φ and for every $z \in \mathcal{K}_j^{red}(w, \varphi)$ there exists a sequence of directional derivatives $\zeta_n \in S'(f; V^*)$ such that $\zeta_n \to z$ holds in V and such that $Q_j^{w,\varphi}(\zeta_n) \nearrow Q_j^{w,\varphi}(z)$ holds for $n \to \infty$. In particular, $S'(f; V^*)$ is a dense subset of $\mathcal{K}_j^{red}(w, \varphi)$.*

Proof. If the solution map S is directionally differentiable in $f \in V^*$ in all directions $g \in V^*$, then it follows from Proposition 1.3.5 that the problem

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w, \varphi), \\ &\big\langle A'(w)\delta, z - \delta \big\rangle + \frac{1}{2} Q_j^{w,\varphi}(z) - \frac{1}{2} Q_j^{w,\varphi}(\delta) \geq \langle g, z - \delta \rangle \quad \forall z \in \mathcal{K}_j^{red}(w, \varphi) \end{aligned} \tag{1.44}$$

admits a unique solution for every $g \in V^*$ and that, if we are given a solution δ to a variational inequality of the type (1.44), then it necessarily holds $\delta = S'(f; g)$. Define $\varepsilon := c/(2\|A'(w)\|_{L(V,V^*)}) > 0$, where c is the monotonicity constant in (1.2). We claim that the variational inequality

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w, \varphi), \\ &(1 + n\varepsilon) \big\langle A'(w)\delta, z - \delta \big\rangle + \frac{1}{2} Q_j^{w,\varphi}(z) - \frac{1}{2} Q_j^{w,\varphi}(\delta) \geq \langle g, z - \delta \rangle \quad \forall z \in \mathcal{K}_j^{red}(w, \varphi) \end{aligned} \tag{1.45}$$

admits a unique solution for all $n \in \mathbb{N}_0$ and all $g \in V^*$. Note that this unique solvability is indeed non-trivial since we do not know anything about the second subderivative $Q_j^{w,\varphi}$ at this point. We prove the claim by induction: For $n = 0$, (1.45) is identical to (1.44) and there is nothing to prove. For the induction step $n \mapsto n + 1$, we fix a $g \in V^*$, assume that a $u \in V$ is given and consider the variational inequality

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w, \varphi), \\ &(1 + n\varepsilon) \big\langle A'(w)\delta, z - \delta \big\rangle + \frac{1}{2} Q_j^{w,\varphi}(z) - \frac{1}{2} Q_j^{w,\varphi}(\delta) \geq \big\langle g - \varepsilon A'(w)u, z - \delta \big\rangle \quad \forall z \in \mathcal{K}_j^{red}(w, \varphi). \end{aligned} \tag{1.46}$$

Note that (1.46) is uniquely solvable for all u and all g by the induction hypothesis. This implies that the solution mapping $T : V \to V$, $u \mapsto \delta$, associated with (1.46) is well-defined. Note further that for all $u_1, u_2 \in V$ with associated solutions $\delta_1 := T(u_1)$, $\delta_2 := T(u_2)$, it holds

$$(1 + n\varepsilon) \big\langle A'(w)\delta_1, \delta_2 - \delta_1 \big\rangle + \frac{1}{2} Q_j^{w,\varphi}(\delta_2) - \frac{1}{2} Q_j^{w,\varphi}(\delta_1) \geq \big\langle g - \varepsilon A'(w)u_1, \delta_2 - \delta_1 \big\rangle$$

and

$$(1+n\varepsilon)\left\langle A'(w)\delta_2, \delta_1-\delta_2\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(\delta_1) - \frac{1}{2}Q_j^{w,\varphi}(\delta_2) \geq \left\langle g-\varepsilon A'(w)u_2, \delta_1-\delta_2\right\rangle.$$

Adding the last two inequalities yields

$$(1+n\varepsilon)\left\langle A'(w)(\delta_1-\delta_2), \delta_1-\delta_2\right\rangle \leq \left\langle \varepsilon A'(w)u_1 - \varepsilon A'(w)u_2, \delta_2-\delta_1\right\rangle$$

and (by the monotonicity and the definition of ε)

$$\|T(u_1)-T(u_2)\|_V = \|\delta_1-\delta_2\|_V \leq \frac{1}{2\,(1+n\varepsilon)}\|u_1-u_2\|_V \leq \frac{1}{2}\|u_1-u_2\|_V.$$

The above shows that the map $T: V \to V$ is a contraction. From Banach's fixpoint theorem, we now obtain that there exists exactly one $\tilde{\delta} \in V$ with $T(\tilde{\delta}) = \tilde{\delta} \in \mathcal{K}_j^{red}(w,\varphi)$, i.e., with

$$\begin{aligned}
&\tilde{\delta} \in \mathcal{K}_j^{red}(w,\varphi),\\
&(1+n\varepsilon)\left\langle A'(w)\tilde{\delta}, z-\tilde{\delta}\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\tilde{\delta}) \geq \left\langle g-\varepsilon A'(w)\tilde{\delta}, z-\tilde{\delta}\right\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi).
\end{aligned}$$

If we rewrite the above, we obtain (1.45) for $n+1$ and the induction step is complete. Consider now an arbitrary but fixed $\tilde{z} \in \mathcal{K}_j^{red}(w,\varphi)$. Then, it follows from the above considerations that for every $n \in \mathbb{N}$ there exists a unique $\zeta_n \in \mathcal{K}_j^{red}(w,\varphi)$ with

$$\begin{aligned}
&\zeta_n \in \mathcal{K}_j^{red}(w,\varphi),\\
&(1+n\varepsilon)\left\langle A'(w)\zeta_n, z-\zeta_n\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\zeta_n) \geq (1+n\varepsilon)\left\langle A'(w)\tilde{z}, z-\zeta_n\right\rangle \qquad (1.47)\\
&\qquad\qquad \forall z \in \mathcal{K}_j^{red}(w,\varphi).
\end{aligned}$$

Note that $\zeta_n = S'(f;(1+n\varepsilon)A'(w)\tilde{z} - n\varepsilon A'(w)\zeta_n) \in S'(f;V^*)$ for all n. Testing (1.47) with $\tilde{z}$ yields

$$\begin{aligned}
\frac{1}{2}Q_j^{w,\varphi}(\tilde{z}) &\geq \frac{1}{2}Q_j^{w,\varphi}(\zeta_n) + (1+n\varepsilon)\left\langle A'(w)\tilde{z} - A'(w)\zeta_n, \tilde{z}-\zeta_n\right\rangle\\
&\geq \frac{1}{2}Q_j^{w,\varphi}(\zeta_n) + c(1+n\varepsilon)\|\tilde{z}-\zeta_n\|_V^2.
\end{aligned}$$

The above entails $S'(f;V^*) \ni \zeta_n \to \tilde{z}$ in V and

$$Q_j^{w,\varphi}(\tilde{z}) \geq \liminf_{n\to\infty} Q_j^{w,\varphi}(\zeta_n). \qquad (1.48)$$

In particular, $S'(f;V^*)$ is dense in $\mathcal{K}_j^{red}(w,\varphi)$. Using part (iii) of Proposition 1.3.5 and Lemma 1.3.13 with $Q := Q_j^{w,\varphi}$, $\mathcal{K} := \mathcal{K}_j^{red}(w,\varphi)$ and $\mathcal{Z} := S'(f;V^*)$, it now follows immediately that j is twice epi-differentiable in w for φ. This, in turn, implies that Lemma 1.3.12 is applicable and that the second subderivative is lower semicontinuous, so that the estimate (1.48) can be continued with

$$Q_j^{w,\varphi}(\tilde{z}) \geq \liminf_{n\to\infty} Q_j^{w,\varphi}(\zeta_n) \geq Q_j^{w,\varphi}(\tilde{z}).$$

This proves the convergence $Q_j^{w,\varphi}(\zeta_n) \nearrow Q_j^{w,\varphi}(\tilde{z})$ and completes the proof. □

It should be noted that the second-order epi-differentiability of j (in all directions z) is not necessary for the directional differentiability of the solution operator $S: V^* \to V$ of (P) in single directions g. Indeed, we can make the following trivial observation:

Proposition 1.3.16. *Let $f \in V^*$ be arbitrary but fixed and let w denote the solution $S(f)$. Then, S is directionally differentiable in all directions $g \in \mathbb{R}^+(\partial j(w) + A(w) - f)$ with $S'(f;g) = 0$.*

Proof. If a $g \in \mathbb{R}^+(\partial j(w) + A(w) - f)$ is given, then the convexity of the subdifferential $\partial j(w)$ implies that there exists an $\varepsilon > 0$ with $f - A(w) + tg \in \partial j(w)$ for all $0 < t < \varepsilon$. The latter yields that

$$\langle A(w), v - w\rangle + j(v) - j(w) \geq \langle f + tg, v - w\rangle \quad \forall v \in V \quad \forall 0 < t < \varepsilon.$$

This implies $S(f + tg) = w$ for all $0 < t < \varepsilon$ and completes the proof. □

The above result shows that, given an $f \in V^*$ with $\partial j(w) \neq \{f - A(w)\}$, we can always find a non-trivial direction $g \in V^*$ such that S is directionally differentiable in f in the direction g - no matter how exotic the function j might be and regardless of whether there are directions for which the condition of second-order epi-differentiability fails. Compare, e.g., with the counterexample in Section 3.2 in this context and also with the comments in Section 3.4.

1.4 Main Theorem and Consequences

We summarize the main results of the last section in:

Theorem 1.4.1 (Main Theorem of the Sensitivity Analysis)**.** *Suppose that a variational inequality of the form* (P) *is given and that Assumption 1.2.1 is satisfied. Denote the solution operator to* (P) *with $S : V^* \to V$, let $f \in V^*$ be arbitrary but fixed, and define $w := S(f)$, $\varphi := f - A(w)$. Then, the following statements are equivalent:*

(I) The solution map S is Hadamard directionally differentiable in f in all directions $g \in V^*$.

(II) The function j is twice epi-differentiable in w for φ.

If one of the above holds, then the following is true:

(i) The second subderivative $Q_j^{w,\varphi} : V \to [0, \infty]$ is proper, convex, lower semicontinuous and positive homogeneous of degree two, and $\mathcal{K}_j^{red}(w, \varphi)$ is a convex, non-empty and pointed cone.

(ii) The directional derivative $\delta := S'(f; g)$ in f in a direction $g \in V^*$ is uniquely characterized by the variational inequality

$$\delta \in \mathcal{K}_j^{red}(w, \varphi), \quad \left\langle A'(w)\delta, z - \delta\right\rangle + \frac{1}{2}Q_j^{w,\varphi}(z) - \frac{1}{2}Q_j^{w,\varphi}(\delta) \geq \langle g, z - \delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w, \varphi) \tag{1.49}$$

and it holds

$$Q_j^{w,\varphi}(\delta) = \langle g, \delta\rangle - \left\langle A'(w)\delta, \delta\right\rangle.$$

(iii) If $g \in V^*$ is fixed and if $\{\delta_t\}$ is the family of difference quotients defined in (1.18)*, then $\{\delta_t\}$ is a recovery sequence for $\delta := S'(f; g)$ in the sense that*

$$\delta_t \to \delta \qquad \text{and} \qquad \frac{2}{t}\left(\frac{j(w + t\delta_t) - j(w)}{t} - \langle \varphi, \delta_t\rangle\right) \to Q_j^{w,\varphi}(\delta)$$

holds for $t \searrow 0$.

(iv) For every $z \in \mathcal{K}_j^{red}(w, \varphi)$ there exists a sequence of directional derivatives $\zeta_n \in S'(f; V^*)$ with

$$\zeta_n \to z \quad \text{and} \quad Q_j^{w,\varphi}(\zeta_n) \nearrow Q_j^{w,\varphi}(z)$$

as $n \to \infty$. In particular, $S'(f; V^*)$ is a dense subset of $\mathcal{K}_j^{red}(w, \varphi)$.

Proof. Combine Propositions 1.3.5 and 1.3.15, Corollary 1.3.11, and Lemma 1.3.12. □

Some remarks are in order regarding Theorem 1.4.1 and its proof:

Remark 1.4.2.

(i) Theorem 1.4.1 significantly generalizes [Do, 1992, Theorem 4.3] where the equivalence between (I) and (II) is proved under the assumption that the operator A is a positive scalar multiple of the Riesz isomorphism, i.e., $\langle A(v_1), v_2\rangle = \alpha(v_1, v_2)_V$ for some $\alpha > 0$. Compare also with [Borwein and Noll, 1994, Proposition 6.3], [Noll, 1995, Theorem 3.3] and [Rockafellar, 1990, Corollary 3.9] in this context. Note that the proofs in the latter references rely heavily on results on the second-order epi-differentiability of infimal convolutions and are, as a consequence, only applicable if the variational inequality at hand can be identified with a classical Moreau-Yosida regularization. The argumentation that we have used in the last section for the derivation of Theorem 1.4.1 does not suffer from this drawback and also covers those cases where the operator A is non-linear and where the EVI (P) *cannot be identified with a minimization problem.*

(ii) In contrast to the approaches in [Do, 1992], [Rockafellar, 1990] and [Adly and Bourdin, 2017], our method of proof is completely elementary. We do not have to invoke, e.g., Attouch's theorem on the characterization of Mosco convergence (see [Attouch, 1984, Theorem 3.66]) to obtain the equivalence of (I) and (II) in Theorem 1.4.1 and can completely avoid working with the concept of protodifferentiability (cf. [Do, 1992, Section 3] and the proof of [Do, 1992, Theorem 4.3]). We remark that, if the concept of protodifferentiability is what one is interested in, then one can employ, e.g., [Do, 1992, Theorem 3.9] to deduce that the conditions (I) and (II) are also equivalent to

(III) The subdifferential $\partial j : V \rightrightarrows V^*$ is protodifferentiable in w relative to φ.

We refer the interested reader to [Do, 1992] for details on this topic.

(iii) It can be shown that the assumption of convexity on j in Theorem 1.4.1 can be dropped if the well-definedness of the solution operator $S : V^* \to V$ to (P) *can be established with means other than Theorem 1.2.2, cf. [Christof and Wachsmuth, 2017a, Section 4]. It is further possible to extend the analysis of Section 1.3 to problems that are not only perturbed in the right-hand side f but also in the operator A and the functional j. See [Christof and Wachsmuth, 2017a, Theorem 4.1] and [Christof and Meyer, 2016] for details on this topic and also [Adly and Bourdin, 2017] for a competing approach.*

To highlight the implications of Theorem 1.4.1, we note the following:

Corollary 1.4.3. *Suppose that a problem of the form* (P) *is given and that Assumption 1.2.1 is satisfied. Denote the solution operator to* (P) *with $S : V^* \to V$. Then, the following conditions are equivalent:*

(I) The map S is Hadamard directionally differentiable in all $f \in V^*$ in all directions $g \in V^*$.

(II) The function j is twice epi-differentiable in all $v \in \mathrm{dom}(\partial j)$ for all $\varphi \in \partial j(v)$.

Proof. The implication (II) $\Rightarrow$ (I) is trivial. To prove that (I) yields (II), it suffices to note that every $(v, \varphi) \in \mathrm{graph}(\partial j)$ satisfies $v = S(\varphi + A(v)) \in K$ (see the definition of the convex subdifferential). The claim then follows immediately from Theorem 1.4.1. □

The important point about Corollary 1.4.3 is that (II) is a condition only on the function j. This implies that the operator A is completely irrelevant when the directional differentiability of the solution mapping $S : V^* \to V$ in all points f in all directions g is what we aim for - the only property that matters in this context is the second-order epi-differentiability of the involved convex functional. Note that Corollary 1.4.3 yields in particular that, if we have proved the directional differentiability of the solution operator S for a problem of the form (P), then we have effectively proved the directional differentiability

of S for all problems of the form (P) that share the same convex functional j. This observation will be exploited frequently in Chapter 2.

We point out that, as a consequence of Theorem 1.4.1, we also obtain the following characterization result for Gâteaux points of the solution operator S:

Corollary 1.4.4 (Gâteaux Points)**.** *Suppose that a variational inequality of the form* (P) *is given and that Assumption 1.2.1 is satisfied. Denote the solution operator to* (P) *with $S : V^* \to V$, let $f \in V^*$ be arbitrary but fixed, and define $w := S(f)$, $\varphi := f - A(w)$. Then, the following statements are equivalent:*

(I) The solution map S is Hadamard-Gâteaux differentiable in f.

(II) The function j is twice epi-differentiable in w for φ, the set $\mathcal{K}_j^{red}(w,\varphi)$ is a subspace, and there exists a symmetric, positive semidefinite, bilinear form $q_j^{w,\varphi} : \mathcal{K}_j^{red}(w,\varphi) \times \mathcal{K}_j^{red}(w,\varphi) \to \mathbb{R}$ such that

$$Q_j^{w,\varphi}(z) = q_j^{w,\varphi}(z,z) \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi).$$

If one of the above holds, then the directional derivative $\delta := S'(f;g)$ in f in a direction $g \in V^*$ is uniquely characterized by the variational equality

$$\delta \in \mathcal{K}_j^{red}(w,\varphi), \quad \big\langle A'(w)\delta, z\big\rangle + q_j^{w,\varphi}(\delta,z) = \langle g, z\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi)$$

and the set $\mathcal{K}_j^{red}(w,\varphi)$ is a Hilbert space with the product $(\cdot,\cdot)_{\mathcal{K}} := (\cdot,\cdot)_V + q_j^{w,\varphi}(\cdot,\cdot)$.

Proof. We first prove (II) $\Rightarrow$ (I): Since j is twice epi-differentiable in w for φ, we know that S is Hadamard directionally differentiable in f in all directions g and that the derivative $\delta := S'(f;g)$ is uniquely characterized by the variational inequality

$$\delta \in \mathcal{K}_j^{red}(w,\varphi), \quad \big\langle A'(w)\delta, z-\delta\big\rangle + \frac{1}{2}q_j^{w,\varphi}(z,z) - \frac{1}{2}q_j^{w,\varphi}(\delta,\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi).$$

If we use test functions of the form $z = \delta \pm tv$ in the above with $v \in \mathcal{K}_j^{red}(w,\varphi)$ and $t > 0$ (recall that the reduced critical cone is a subspace), then we obtain

$$\delta \in \mathcal{K}_j^{red}(w,\varphi), \quad \pm\big\langle A'(w)\delta, v\big\rangle \pm q_j^{w,\varphi}(\delta,v) + \frac{t}{2}q_j^{w,\varphi}(v,v) \geq \pm\langle g, v\rangle \quad \forall v \in \mathcal{K}_j^{red}(w,\varphi). \quad (1.50)$$

Letting $t \searrow 0$ in (1.50) yields

$$\delta \in \mathcal{K}_j^{red}(w,\varphi), \quad \big\langle A'(w)\delta, v\big\rangle + q_j^{w,\varphi}(\delta,v) = \langle g, v\rangle \quad \forall v \in \mathcal{K}_j^{red}(w,\varphi).$$

The above implies that the map $V^* \ni g \mapsto S'(f;g) \in V$ is linear and that S is indeed Hadamard-Gâteaux differentiable in f (cf. Lemma 1.1.3).

It remains to prove (I) $\Rightarrow$ (II). To this end, we note that the Hadamard-Gâteaux differentiability of S implies that j is twice epi-differentiable in w for φ, that $Q_j^{w,\varphi}$ is convex and lower semicontinuous, and that $S'(f;\alpha g_1 + \beta g_2) = \alpha S'(f;g_1) + \beta S'(f;g_2)$ holds for all $g_1, g_2 \in V^*$ and all $\alpha, \beta \in \mathbb{R}$. In particular, $S'(f;V^*)$ is a subspace. Consider now some $z \in \mathcal{K}_j^{red}(w,\varphi)$. Then, Theorem 1.4.1 yields that there exist $\zeta_n \in S'(f;V^*)$ with $\zeta_n \to z$ in V and $Q_j^{w,\varphi}(\zeta_n) \nearrow Q_j^{w,\varphi}(z)$. This, the lower semicontinuity of the second subderivative and (1.30) imply that the sequence $-\zeta_n \in S'(f;V^*)$ satisfies

$$-\zeta_n \to -z, \qquad Q_j^{w,\varphi}(-z) \leq \liminf_{n\to\infty} Q_j^{w,\varphi}(-\zeta_n) = \lim_{n\to\infty} Q_j^{w,\varphi}(\zeta_n) = Q_j^{w,\varphi}(z).$$

Accordingly, $-z \in \mathcal{K}_j^{red}(w,\varphi)$ and $Q_j^{w,\varphi}(-z) = Q_j^{w,\varphi}(z)$. Since $Q_j^{w,\varphi}$ is also convex and positively homogeneous of degree two, we obtain that

$$\begin{aligned}
0 &\leq Q_j^{w,\varphi}(\alpha z_1 + \beta z_2)\\
&= (|\alpha| + |\beta|)^2 Q_j^{w,\varphi}\left(\frac{|\alpha|}{|\alpha|+|\beta|}\operatorname{sgn}(\alpha) z_1 + \frac{|\beta|}{|\alpha|+|\beta|}\operatorname{sgn}(\beta) z_2\right)\\
&\leq (|\alpha| + |\beta|)|\alpha| Q_j^{w,\varphi}(z_1) + (|\alpha| + |\beta|)|\beta| Q_j^{w,\varphi}(z_2)
\end{aligned}$$

holds for all $z_1, z_2 \in \mathcal{K}_j^{red}(w,\varphi)$ and all $\alpha, \beta \in \mathbb{R}$. This proves that $\mathcal{K}_j^{red}(w,\varphi)$ is a subspace. Recall that every directional derivative $\delta = S'(f;g)$ satisfies

$$Q_j^{w,\varphi}(\delta) = \langle g, \delta \rangle - \left\langle A'(w)\delta, \delta \right\rangle .$$

The above and the Gâteaux differentiability of S yield that for all $\delta_1 := S'(f;g_1)$ and all $\delta_2 := S'(f;g_2)$, we have the polarization identity

$$Q_j^{w,\varphi}(\delta_1) + Q_j^{w,\varphi}(\delta_2) = \frac{Q_j^{w,\varphi}(\delta_1+\delta_2) + Q_j^{w,\varphi}(\delta_1-\delta_2)}{2}. \tag{1.51}$$

Using part (iv) of Theorem 1.4.1 and the lower semicontinuity of the second subderivative, we obtain that (1.51) also holds for all $z_1, z_2 \in \mathcal{K}_j^{red}(w,\varphi)$. Consider now the function

$$q_j^{w,\varphi} : \mathcal{K}_j^{red}(w,\varphi) \times \mathcal{K}_j^{red}(w,\varphi) \to \mathbb{R}, \qquad q_j^{w,\varphi}(z_1,z_2) := \frac{1}{4}\left(Q_j^{w,\varphi}(z_1+z_2) - Q_j^{w,\varphi}(z_1-z_2)\right).$$

Then, $q_j^{w,\varphi}$ is trivially symmetric. We claim that $q_j^{w,\varphi}$ is also additive in the first argument, i.e.,

$$q_j^{w,\varphi}(z_1+z_2,z_3) = q_j^{w,\varphi}(z_1,z_3) + q_j^{w,\varphi}(z_2,z_3) \qquad \forall z_1,z_2,z_3 \in \mathcal{K}_j^{red}(w,\varphi). \tag{1.52}$$

To see this, we note that the polarization identity (1.51) implies

$$Q_j^{w,\varphi}(z_1+z_3) + Q_j^{w,\varphi}(z_2) = \frac{Q_j^{w,\varphi}(z_1+z_2+z_3) + Q_j^{w,\varphi}(z_1-z_2+z_3)}{2}.$$

The above yields (with an interchange of z_1 and z_2),

$$\begin{aligned} Q_j^{w,\varphi}(z_1+z_2+z_3) &= 2Q_j^{w,\varphi}(z_1+z_3) + 2Q_j^{w,\varphi}(z_2) - Q_j^{w,\varphi}(z_1-z_2+z_3) \\ &= 2Q_j^{w,\varphi}(z_2+z_3) + 2Q_j^{w,\varphi}(z_1) - Q_j^{w,\varphi}(-z_1+z_2+z_3). \end{aligned}$$

By addition, we now obtain

$$\begin{aligned} Q_j^{w,\varphi}(z_1+z_2+z_3) = Q_j^{w,\varphi}(z_1) + Q_j^{w,\varphi}(z_2) + Q_j^{w,\varphi}(z_1+z_3) + Q_j^{w,\varphi}(z_2+z_3) \\ - \frac{1}{2}Q_j^{w,\varphi}(z_1-z_2+z_3) - \frac{1}{2}Q_j^{w,\varphi}(-z_1+z_2+z_3) \end{aligned}$$

and (exchanging z_3 with $-z_3$ and using $Q_j^{w,\varphi}(-z) = Q_j^{w,\varphi}(z)$)

$$\begin{aligned} Q_j^{w,\varphi}(z_1+z_2-z_3) = Q_j^{w,\varphi}(z_1) + Q_j^{w,\varphi}(z_2) + Q_j^{w,\varphi}(z_1-z_3) + Q_j^{w,\varphi}(z_2-z_3) \\ - \frac{1}{2}Q_j^{w,\varphi}(-z_1+z_2+z_3) - \frac{1}{2}Q_j^{w,\varphi}(z_1-z_2+z_3). \end{aligned}$$

Combining all of the above, we arrive at

$$\begin{aligned} q_j^{w,\varphi}(z_1+z_2,z_3) &= \frac{1}{4}\left(Q_j^{w,\varphi}(z_1+z_2+z_3) - Q_j^{w,\varphi}(z_1+z_2-z_3)\right) \\ &= \frac{1}{4}\left(Q_j^{w,\varphi}(z_1+z_3) + Q_j^{w,\varphi}(z_2+z_3) - Q_j^{w,\varphi}(z_1-z_3) - Q_j^{w,\varphi}(z_2-z_3)\right) \\ &= q_j^{w,\varphi}(z_1,z_3) + q_j^{w,\varphi}(z_2,z_3). \end{aligned}$$

This is precisely (1.52). The property of symmetry now yields that $q_j^{w,\varphi}(\cdot,\cdot)$ is biadditive and, as a consequence, $\mathbb{Q}$-bilinear. From the convexity of $Q_j^{w,\varphi}$, we obtain further that the map

$$\mathbb{R} \ni \alpha \mapsto q_j^{w,\varphi}(\alpha z_1, z_2) = \frac{1}{4}\left(Q_j^{w,\varphi}(\alpha z_1 + z_2) - Q_j^{w,\varphi}(\alpha z_1 - z_2)\right) \in \mathbb{R}$$

is continuous. This yields that $q_j^{w,\varphi}(\cdot,\cdot)$ is also $\mathbb{R}$-bilinear. Since $q_j^{w,\varphi}(\cdot,\cdot)$ is positive semidefinite by construction, the claim now follows immediately.

It remains to prove that $\mathcal{K}_j^{red}(w,\varphi)$ is a Hilbert space with the product $(\cdot,\cdot)_{\mathcal{K}} = (\cdot,\cdot)_V + q_j^{w,\varphi}(\cdot,\cdot)$ when one of the conditions (I) or (II) is satisfied. To this end, we first note that $(\cdot,\cdot)_{\mathcal{K}}$ is trivially a scalar product. This implies that it suffices to prove completeness to obtain the claim. So let us assume that a $\|\cdot\|_{\mathcal{K}}$-Cauchy sequence $\{z_n\} \subset \mathcal{K}_j^{red}(w,\varphi)$ is given. Then, $\{z_n\}$ is also Cauchy in V and it holds $z_n \to z$ for some $z \in V$. Further, the Cauchy-Schwarz inequality for $(\cdot,\cdot)_{\mathcal{K}}$, the Cauchy property of $\{z_n\}$ and the lower semicontinuity of the second subderivative $Q_j^{w,\varphi}$ yield that $Q_j^{w,\varphi}(z_n)$ is bounded, that z is an element of $\mathcal{K}_j^{red}(w,\varphi)$ and that, given an arbitrary but fixed $\varepsilon > 0$, we can find an $N > 0$ with

$$0 \le Q_j^{w,\varphi}(z_n - z_m) \le \varepsilon \quad \forall m,n \ge N.$$

The above implies in combination with the lower semicontinuity of $Q_j^{w,\varphi}$ that

$$0 \le Q_j^{w,\varphi}(z - z_n) \le \liminf_{m\to\infty} Q_j^{w,\varphi}(z_m - z_n) \le \varepsilon \quad \forall n \ge N.$$

This proves that $\|z - z_n\|_{\mathcal{K}} \to 0$ and that $(\mathcal{K}_j^{red}(w,\varphi), \|\cdot\|_{\mathcal{K}})$ is indeed a Hilbert space. □

We point out that Corollary 1.4.4 is a useful tool in the study of Bouligand subdifferentials of solution maps to elliptic variational inequalities of the first and the second kind. Compare, e.g., with the analysis in [Christof et al., 2017] in this context and in particular with [Christof et al., 2017, Theorem 3.18].

The relationship between the Hadamard-Gâteaux differentiability of the solution operator S and the second-order epi-differentiability of the functional j established in Corollary 1.4.4 can further be used to derive general results on the second-order (epi-)differentiability properties of convex functions. The following version of Alexandrov's theorem that has been first proved (by different means) in [Kato, 1989], for example, follows straightforwardly from the equivalence between the conditions (I) and (II) in the last result (see also [Borwein and Noll, 1994, Proposition 6.5]):

Corollary 1.4.5 (An Epi-Differentiability Version of Alexandrov's Theorem in Infinite Dimensions)**.** *Let V be a separable Hilbert space and let $j : V \to (-\infty,\infty]$ be a convex, lower semicontinuous and proper function. Let $\Gamma \subset V \times V^*$ denote the graph of the subdifferential $\partial j : V \rightrightarrows V^*$ of j. Then, there exists a dense subset D of Γ such that for all $(v,\varphi) \in D$ there exist a subspace H of V and a symmetric, positive semidefinite, bilinear form $q : H \times H \to \mathbb{R}$ (both depending on (v,φ)) with the following properties:*

(i) The space H is Hilbert when equipped with the product $(\cdot,\cdot)_H := (\cdot,\cdot)_V + q(\cdot,\cdot)$.

(ii) The second-order difference quotients

$$\frac{2}{t}\left(\frac{j(v+tz)-j(v)}{t} - \langle\varphi, z\rangle\right), \quad z \in V, \quad t > 0, \tag{1.53}$$

are Mosco epi-convergent to the function

$$Q(z) := \begin{cases} q(z,z) & \text{if } z \in H \\ +\infty & \text{else} \end{cases} \tag{1.54}$$

for all $(v,\varphi) \in D$.

Proof. Consider for $f \in V^*$ the variational inequality

$$w \in V, \quad (w, v-w)_V + j(v) - j(w) \ge \langle f, v-w\rangle \quad \forall v \in V \tag{1.55}$$

and denote the solution operator to (1.55) with $S : V^* \to V$, $f \mapsto w$. Then, we know from Theorem 1.2.2 that S is well-defined and globally Lipschitz and it follows from the separability of the dual space V^* and

Mignot's extension of Rademacher's theorem, see [Mignot, 1976, Théorème 1.1] and [Aronszajn, 1976, Theorem 1], that there exists a set $\tilde{D} \subset V^*$ that is dense in V^* such that S is Gâteaux in every $f \in \tilde{D}$. According to Proposition 1.3.9, Corollary 1.4.4 and [Bonnans and Shapiro, 2000, Proposition 2.49], the latter implies that the second-order difference quotients

$$\frac{2}{t}\left(\frac{j(w+tz)-j(w)}{t} - \langle f - A(w), z\rangle\right), \qquad z \in V, \quad t > 0,$$

are Mosco epi-convergent to

$$Q_j^{w,\varphi}(z) = \begin{cases} q_j^{w,\varphi}(z,z) & \text{if } z \in \mathcal{K}_j^{red}(w,\varphi) \\ +\infty & \text{else} \end{cases}$$

for all $f \in \tilde{D}$ as $t \searrow 0$, where $w := S(f)$ and $\varphi := f - A(w)$. Here, $q_j^{w,\varphi}(\cdot,\cdot)$ is the symmetric, positive semidefinite, bilinear form in Corollary 1.4.4, $\mathcal{K}_j^{red}(w,\varphi)$ is the reduced critical cone (which is a Hilbert space when equipped with the product $(\cdot,\cdot)_{\mathcal{K}} := (\cdot,\cdot)_V + q_j^{w,\varphi}(\cdot,\cdot)$) and $A : V \to V^*$ is the map $v \mapsto (v,\cdot)_V$. We claim that the set

$$D := \left\{(w, f - A(w)) \;\middle|\; f \in \tilde{D}, w = S(f)\right\}$$

is dense in $\operatorname{graph}(\partial j) \subset V \times V^*$. To see this, we first note that we trivially have $D \subset \operatorname{graph}(\partial j)$. Suppose now that a tuple $(v,\varphi) \in \operatorname{graph}(\partial j)$ is given. Then, it holds

$$j(u) - j(v) \geq \langle \varphi, u - v\rangle \quad \forall u \in V$$

and, consequently, $v = S(\varphi + A(v))$. Set $f := \varphi + A(v)$. Then, we know that there exists a sequence $f_n \in \tilde{D}$ with $f_n \to f$ in V^* and we obtain from the Lipschitz continuity of the solution operator S that $w_n := S(f_n) \to S(f) = v$ holds for $n \to \infty$, i.e., we have $(w_n, f_n) \to (v, f)$ in $V \times V^*$. The latter yields that $D \ni (w_n, f_n - A(w_n)) \to (v, f - A(v)) = (v,\varphi)$ in $V \times V^*$. This proves the density. Combining all of the above, the claim follows immediately. □

Note that, in contrast to the classical theorem of Alexandrov, see [Alexandrov, 1939], [Niculescu and Persson, 2006, Theorem 3.11.2] and [Borwein and Vanderwerff, 2010, Theorem 2.6.4], Corollary 1.4.5 makes a statement about the second-order epi-differentiability of j on a dense subset of the graph $\operatorname{graph}(\partial j) \subset V \times V^*$ and not about the existence of a second-order Taylor expansion of j on a dense subset of the domain $\operatorname{dom}(j) \subset V$. The latter implies that Corollary 1.4.5 provides more information about the convex function under consideration than its classical counterpart, at the expense that the obtained information is of lesser quality (since the notion of second-order epi-differentiability is more general than that of second-order expansibility). Consider, for example, the absolute value function $j : \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$. For this map, the classical version of Alexandrov's theorem does not yield any information at the origin $x = 0$. This is different for Corollary 1.4.5 which implies that there exists a dense set $D \subset \partial j(0) = [-1, 1]$ such that the difference quotients in (1.53) are Mosco epi-convergent in $x = 0$ to some functional of the form (1.54) for all $\varphi \in D \subset \partial j(0)$. We remark that the notion of second-order epi-differentiability is also the natural concept for the study of, e.g., necessary and sufficient second-order optimality conditions, cf. [Christof and Wachsmuth, 2017b], [Rockafellar and Wets, 1998] and Theorem 6.2.1 in Section 6.2. Because of this, Corollary 1.4.5 is preferable to classical variants of Alexandrov's theorem in many situations.

2 Calculus Rules for Second-Order Epi-Derivatives

As we have seen in the last chapter, the notion of second-order epi-differentiability plays a central role in the sensitivity analysis of elliptic variational inequalities of the first and the second kind. To simplify working with this type of differentiability, in what follows, we collect and prove several calculus rules for twice epi-differentiable functions. In Section 2.1, we begin our analysis by studying the relationship between the property of second-order epi-differentiability and the classical concepts in Definition 1.1.2. The main result of this section, Theorem 2.1.1, yields that a convex function $j : V \to \mathbb{R}$ with a continuous and directionally differentiable first derivative $j' : V \to V^*$ is always twice epi-differentiable with $Q_j^{v,\varphi}(z) = \langle (j')'(v;z), z\rangle$ for all $z \in V$ and all $(v,\varphi) \in \operatorname{graph}(\partial j)$. In the subsequent Sections 2.2 to 2.4, we prove a sum rule and two chain rules that are used extensively in Chapters 3 to 5, see Theorems 2.2.1, 2.3.1 and 2.4.8. Lastly, Section 2.5 is concerned with the second-order epi-differentiability of functions that involve superposition operators.

2.1 Second-Order Epi-Differentiability of C^1-Functions

In what follows, we study the second-order epi-differentiability of functions that possess a continuous and directionally differentiable first derivative. The main result of this section is:

Theorem 2.1.1. *Let V be a Hilbert space, and let $j : V \to (-\infty,\infty]$ be a convex, lower semicontinuous and proper function. Suppose that a $v \in \operatorname{dom}(j)$ and a $z \in V$ are given such that:*

(i) j is continuously Gâteaux differentiable in an open neighborhood $D \subset \operatorname{dom}(j)$ of v,

(ii) the map $D \ni u \mapsto j'(u) \in V^*$ is directionally differentiable in v in the direction z.

Then, it holds $\partial j(v) = \{j'(v)\}$ and j is twice epi-differentiable in v for $\varphi := j'(v)$ in the direction z with

$$Q_j^{v,\varphi}(z) = \big\langle (j')'(v;z), z\big\rangle\,.$$

Here and in the remainder of this work, with "continuously Fréchet/Gâteaux differentiable in D" we mean that the function under consideration is Fréchet/Gâteaux differentiable everywhere in D and that the map which assigns to each point in D its Fréchet/Gâteaux derivative is continuous. Recall that functions with the latter properties are also called C^1-functions and that the notions of continuous Gâteaux differentiability and continuous Fréchet differentiability coincide on open sets. We refer to [Drábek and Milota, 2007, Section 3.2] and [Bonnans and Shapiro, 2000, Section 2.2.1] for details.

Proof of Theorem 2.1.1. We follow the lines of [Noll, 1995, proof of Proposition 3.2]: Suppose that sequences $\{t_n\} \subset \mathbb{R}^+$ and $\{z_n\} \subset V$ are given such that $t_n \searrow 0$ and $z_n \rightharpoonup z$ in V. Assume w.l.o.g. that $v + t_n z \in D$ holds for all $n \in \mathbb{N}$ and define

$$j_{t_n} : V \to [0,\infty], \qquad j_{t_n}(u) := \frac{2}{t_n}\left(\frac{j(v+t_n u) - j(v)}{t_n} - \big\langle j'(v), u\big\rangle\right),$$

and

$$\varphi_n := \frac{2}{t_n}\big(j'(v+t_n z) - j'(v)\big)\,.$$

Then, for every $u \in V$ it is true that

$$j_{t_n}(u) - j_{t_n}(z) - \langle \varphi_n, u - z \rangle = \frac{2}{t_n} \left(\frac{j(v + t_n u) - j(v + t_n z)}{t_n} - \langle j'(v + t_n z), u - z \rangle \right) \geq 0,$$

where the last estimate follows from $\partial j(v + t_n z) = \{j'(v + t_n z)\}$ for all $n \in \mathbb{N}$, cf. [Ekeland and Temam, 1976, Proposition 5.3]. In particular, we have

$$j_{t_n}(z_n) \geq j_{t_n}(z) + \langle \varphi_n, z_n - z \rangle .$$

Note that the directional differentiability of the map $D \ni u \mapsto j'(u) \in V^*$ in v in the direction z implies

$$\langle \varphi_n, z_n - z \rangle = 2 \left\langle \frac{j'(v + t_n z) - j'(v)}{t_n}, z_n - z \right\rangle \to 0.$$

We may thus deduce that

$$\liminf_{n \to \infty} j_{t_n}(z_n) \geq \liminf_{n \to \infty} j_{t_n}(z). \tag{2.1}$$

Recall further that the definition of directional differentiability yields

$$\exists \varepsilon > 0 \, : \, \left\| \frac{j'(v + tz) - j'(v)}{t} - (j')'(v; z) \right\|_{V^*} < 1 \quad \forall t \in (0, \varepsilon).$$

The latter implies in tandem with the dominated convergence theorem, the mean value theorem [Drábek and Milota, 2007, Theorem 3.2.6] and the directional differentiability of the map $D \ni u \mapsto j'(u) \in V^*$ in v in the direction z that

$$\begin{aligned}
\liminf_{n \to \infty} j_{t_n}(z) &= \liminf_{n \to \infty} \frac{2}{t_n} \left(\frac{j(v + t_n z) - j(v)}{t_n} - \langle j'(v), z \rangle \right) \\
&= \liminf_{n \to \infty} \frac{2}{t_n} \left(\int_0^1 \langle j'(v + s t_n z), z \rangle \, \mathrm{d}s - \langle j'(v), z \rangle \right) \\
&= \liminf_{n \to \infty} 2 \left(\int_0^1 \left\langle \frac{j'(v + s t_n z) - j'(v)}{t_n}, z \right\rangle \mathrm{d}s \right) \\
&= 2 \left(\int_0^1 s \, \mathrm{d}s \right) \langle (j')'(v; z), z \rangle \\
&= \langle (j')'(v; z), z \rangle .
\end{aligned} \tag{2.2}$$

If we combine (2.2) with Definition 1.3.1 and (2.1), then we obtain

$$\langle (j')'(v; z), z \rangle \leq Q_j^{v,\varphi}(z) \leq \liminf_{n \to \infty} j_{t_n}(z) = \lim_{n \to \infty} j_{t_n}(z) = \langle (j')'(v; z), z \rangle .$$

This proves the claim. □

We remark that, despite its rather restrictive regularity assumptions on the function j, Theorem 2.1.1 is a surprisingly useful tool in, e.g., the sensitivity analysis of non-smooth partial differential equations. Details on this topic may be found in Section 4.1.

Note that Theorem 2.1.1 implies that the second subderivative $Q_j^{v,\varphi}$ is indeed a generalization of the classical (directional) second derivative. In particular, we have (cf. [Do, 1992, Proposition 4.1]):

Corollary 2.1.2. *Let V be a Hilbert space, and let $j : V \to (-\infty, \infty]$ be a convex, lower semicontinuous and proper function. Suppose that a $v \in \mathrm{dom}(j)$ is given such that j is continuously Gâteaux differentiable in an open neighborhood of v and twice Gâteaux differentiable in v with second derivative $j''(v) : V \times V \to \mathbb{R}$ (cf. [Drábek and Milota, 2007, Section 3.2]). Then, j is twice epi-differentiable in v for $\varphi := j'(v)$ with*

$$\mathcal{K}_j^{red}(v, \varphi) = V \qquad \text{and} \qquad Q_j^{v,\varphi}(z) = j''(v) z^2 \quad \forall z \in V.$$

2.2 A Sum Rule

Having established that the functional $Q_j^{v,\varphi}$ coincides with the classical (directional) second derivative for sufficiently regular j, we now turn our attention to the question of how twice epi-differentiable functions behave under addition and composition. We begin our investigation by proving:

Theorem 2.2.1 (Sum Rule)**.** *Let V be a Hilbert space, and let $j_1, j_2 : V \to (-\infty, \infty]$ be convex, lower semicontinuous and proper functions. Suppose that a $v \in \operatorname{dom}(j_1) \cap \operatorname{dom}(j_2)$, a $\varphi_2 \in \partial j_2(v)$, a $z \in V$ and an $\alpha > 0$ are given such that the following is true:*

(i) j_1 is continuously Gâteaux differentiable in an open neighborhood $D \subset \operatorname{dom}(j_1)$ of v,

(ii) the map $D \ni u \mapsto j_1'(u) \in V^*$ is Hadamard directionally differentiable in v in the direction z,

(iii) j_2 is twice epi-differentiable in v for φ_2 in the direction z.

Define $\varphi_1 := j_1'(v)$. Then, it holds $\varphi_1 + \alpha\varphi_2 \in \partial(j_1 + \alpha j_2)(v)$, and the function $j_1 + \alpha j_2$ is twice epi-differentiable in v for $\varphi_1 + \alpha\varphi_2$ in the direction z with

$$Q_{j_1+\alpha j_2}^{v,\varphi_1+\alpha\varphi_2}(z) = \left\langle (j_1')'(v;z), z \right\rangle + \alpha Q_{j_2}^{v,\varphi_2}(z).$$

Proof. From Definition 1.3.1, Theorem 2.1.1 and [Ekeland and Temam, 1976, Section 5.3], it follows straightforwardly that $\varphi_1 + \alpha\varphi_2 \in \partial(j_1 + \alpha j_2)(v)$ and

$$Q_{j_1+\alpha j_2}^{v,\varphi_1+\alpha\varphi_2}(z) \geq Q_{j_1}^{v,\varphi_1}(z) + \alpha Q_{j_2}^{v,\varphi_2}(z) = \left\langle (j_1')'(v;z), z \right\rangle + \alpha Q_{j_2}^{v,\varphi_2}(z). \tag{2.3}$$

Consider now an arbitrary but fixed $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$. Then, the second-order epi-differentiability of j_2 in v for φ_2 in the direction z yields that there exists a sequence $\{z_n\} \subset V$ with

$$z_n \to z \quad \text{and} \quad Q_{j_2}^{v,\varphi_2}(z) = \lim_{n\to\infty} \frac{2}{t_n} \left(\frac{j_2(v + t_n z_n) - j_2(v)}{t_n} - \langle \varphi_2, z_n \rangle \right), \tag{2.4}$$

and we may compute analogously to (2.2) that

$$\lim_{n\to\infty} \frac{2}{t_n} \left(\frac{j_1(v + t_n z_n) - j_1(v)}{t_n} - \langle \varphi_1, z_n \rangle \right) = \left\langle (j_1')'(v;z), z \right\rangle. \tag{2.5}$$

If we combine (2.4) and (2.5), then we obtain

$$\lim_{n\to\infty} \frac{2}{t_n} \left(\frac{(j_1 + \alpha j_2)(v + t_n z_n) - (j_1 + \alpha j_2)(v)}{t_n} - \langle \varphi_1 + \alpha\varphi_2, z_n \rangle \right) = \left\langle (j_1')'(v;z), z \right\rangle + \alpha Q_{j_2}^{v,\varphi_2}(z).$$

The claim now follows immediately from (2.3) and Definition 1.3.6. □

If the function $D \ni u \mapsto j_1'(u) \in V^*$ in Theorem 2.2.1 is not only directionally differentiable but even Fréchet differentiable in v, then it is also possible to recover the second-order epi-differentiability of j_2 in v for φ_2 from the second-order epi-differentiability of the sum $j_1 + \alpha j_2$ in v for $\varphi_1 + \alpha\varphi_2$, cf. [Do, 1992, Proposition 4.2]. In the case that j_1 is twice Fréchet differentiable everywhere in D, this can be proved quite elegantly with Theorem 1.4.1:

Theorem 2.2.2 (Sum Rule in the Presence of Second Fréchet Derivatives)**.** *Let V be a Hilbert space, and let $j_1, j_2 : V \to (-\infty, \infty]$ be convex, lower semicontinuous and proper functions. Suppose that a $v \in \operatorname{dom}(j_1) \cap \operatorname{dom}(j_2)$, a $\varphi_2 \in \partial j_2(v)$ and an $\alpha > 0$ are given such that j_1 is twice Fréchet differentiable in an open neighborhood of v. Then, the following statements are equivalent:*

(i) The function j_2 is twice epi-differentiable in v for φ_2.

(ii) The function $j_1 + \alpha j_2$ is twice epi-differentiable in v for $\varphi_1 + \alpha\varphi_2$, where $\varphi_1 := j_1'(v)$.

If one of the above is satisfied, then it holds:

$$\mathcal{K}^{red}_{j_1+\alpha j_2}(v,\varphi_1+\alpha\varphi_2) = \mathcal{K}^{red}_{j_2}(v,\varphi_2) \;\; \textit{and} \;\; Q^{v,\varphi_1+\alpha\varphi_2}_{j_1+\alpha j_2}(z) = j_1''(v)z^2 + \alpha Q^{v,\varphi_2}_{j_2}(z) \quad \forall z \in V. \tag{2.6}$$

Proof. Choose an $r > 0$ such that j_1 is twice Fréchet differentiable in an open neighborhood of the closed ball $B_r(v) := \{u \in V \mid \|u - v\|_V \leq r\}$ and such that $\|j_1'(u)\|_{L(V,V^*)} \leq M$ holds for all $u \in B_r(v)$ and some $M > 0$ (possible due to the continuity of the first derivative). Consider further the minimization problem

$$\min_{u \in B_r(v)} \frac{1}{2}\|u\|_V^2 + j_1(u) + \alpha j_2(u) - \langle f, u\rangle\,, \tag{2.7}$$

where $f \in V^*$ is assumed to be a given datum. Then, it follows from the convexity of j_1 and [Ekeland and Temam, 1976, Proposition 5.5] that the map $B_r(v) \ni u \mapsto \langle j_1'(u), \cdot\rangle + (u,\cdot)_V \in V^*$ satisfies the conditions (i), (ii) and (iii) in Assumption 1.2.1, and we obtain from our considerations in Section 1.2 that (2.7) is equivalent to both the elliptic variational inequalities

$$w \in B_r(v), \quad (w, u-w)_V + j_1(u) + \alpha j_2(u) - j_1(w) - \alpha j_2(w) \geq \langle f, u - w\rangle \quad \forall u \in B_r(v) \tag{2.8}$$

and

$$w \in B_r(v), \quad (w, u-w)_V + \big\langle j_1'(w), u - w\big\rangle + \alpha j_2(u) - \alpha j_2(w) \geq \langle f, u - w\rangle \quad \forall u \in B_r(v) \tag{2.9}$$

for all $f \in V^*$. Note that the solution to (2.7), (2.8) and (2.9) with right-hand side $f = v + \varphi_1 + \alpha\varphi_2$ is precisely v. This implies in combination with Theorem 1.4.1 that the following statements are equivalent:

(i) The function $\chi_{B_r(v)} + j_1 + \alpha j_2$ is twice epi-differentiable in v for $\varphi_1 + \alpha\varphi_2$.

(ii) The solution operator $S : V^* \to V$, $f \mapsto w$, associated with (2.7), (2.8) and (2.9) is Hadamard directionally differentiable in $v + \varphi_1 + \alpha\varphi_2$ in all directions $g \in V^*$.

(iii) The function $\chi_{B_r(v)} + \alpha j_2$ is twice epi-differentiable in v for $\alpha\varphi_2$.

Here, $\chi_{B_r(v)} : V \to \{0, \infty\}$ denotes the characteristic function of the set $B_r(v)$. Since v is contained in the interior $\mathrm{int}(B_r(v))$ of the ball $B_r(v)$, the above equivalencies remain valid when we replace the expressions $\chi_{B_r(v)} + j_1 + \alpha j_2$ and $\chi_{B_r(v)} + \alpha j_2$ in (i) and (iii) with $j_1 + \alpha j_2$ and αj_2, respectively, cf. Definitions 1.3.1 and 1.3.6. Further, it is straightforward to check that αj_2 is twice epi-differentiable in v for $\alpha\varphi_2$ if and only if j_2 is twice epi-differentiable in v for φ_2. We may thus deduce that the second-order epi-differentiability of $j_1 + \alpha j_2$ in v for $\varphi_1 + \alpha\varphi_2$ is indeed equivalent to the second-order epi-differentiability of j_2 in v for φ_2. To complete the proof, it remains to show (2.6). This, however, follows immediately from Theorem 2.2.1. □

We conclude this section with some comments on Theorems 2.2.1 and 2.2.2:

Remark 2.2.3.

(i) By invoking a slightly more general variant of Theorem 1.4.1 in the proof of Theorem 2.2.2, namely [Christof and Wachsmuth, 2017a, Theorem 4.1], it is possible to relax the regularity assumptions on j_1 in the last theorem such that they coincide with those of [Do, 1992, Proposition 4.2].

(ii) The proof of [Do, 1992, Proposition 4.2] relies heavily on Attouch's theorem (cf. [Attouch, 1984, Theorem 3.66]) and other results from set-valued analysis. By arguing with the auxiliary problem (2.7) *and the equivalency in Theorem 1.4.1, we can completely avoid working with these instruments. Compare also with the argumentation used in Section 2.5 in this context.*

(iii) If $j_1, j_2 : V \to (-\infty, \infty]$ are two convex, lower semicontinuous and proper functions that are twice epi-differentiable in a point $v \in \operatorname{dom}(j_1) \cap \operatorname{dom}(j_2)$ for some subgradients $\varphi_1 \in \partial j_1(v)$ and $\varphi_2 \in \partial j_2(v)$, respectively, then it is in general not true that the sum $j_1 + j_2$ is twice epi-differentiable in v for $\varphi_1 + \varphi_2 \in \partial(j_1 + j_2)(v)$ with

$$Q_{j_1+j_2}^{v,\varphi_1+\varphi_2}(z) = Q_{j_1}^{v,\varphi_1}(z) + Q_{j_2}^{v,\varphi_2}(z) \quad \forall z \in V. \tag{2.10}$$

Consider, for example, the characteristic functions $j_1 := \chi_{D_1}$ and $j_2 := \chi_{D_2}$ of the sets

$$D_1 := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 \leq x_2\} \quad \text{and} \quad D_2 := \{(x_1, x_2) \in \mathbb{R}^2 \mid -x_1^2 \geq x_2\}.$$

Then, it is easy to see (either by direct calculation or by invoking Lemma 3.2.1) that both j_1 and j_2 are twice epi-differentiable in $v := (0,0)$ for $\varphi := (0,0) \in \partial j_1(v) \cap \partial j_2(v)$ with

$$Q_{j_1}^{v,\varphi} = \chi_{\mathcal{T}_{D_1}(v)} = \chi_{\mathbb{R}\times[0,\infty)} \quad \text{and} \quad Q_{j_2}^{v,\varphi} = \chi_{\mathcal{T}_{D_2}(v)} = \chi_{\mathbb{R}\times(-\infty,0]}.$$

In particular,

$$Q_{j_1}^{v,\varphi} + Q_{j_2}^{v,\varphi} = \chi_{\mathbb{R}\times\{0\}}.$$

However, from $\operatorname{dom}(j_1) \cap \operatorname{dom}(j_2) = \{v\}$, we also obtain

$$Q_{j_1+j_2}^{v,\varphi+\varphi} = Q_{\chi_{\{v\}}}^{v,\varphi} = \chi_{\{(0,0)\}} \neq \chi_{\mathbb{R}\times\{0\}}.$$

This shows that equation (2.10) can indeed be false when we do not impose additional assumptions on the involved functions j_1 and j_2 (cf. the setting of Theorem 2.2.1).

2.3 A First Chain Rule

Next, we study the second-order epi-differentiability properties of composite functions. We first consider the case where a twice epi-differentiable, convex and lower semicontinuous $j : V \to \mathbb{R}$ is composed with a twice continuously differentiable $k : \mathbb{R} \to \mathbb{R}$. In this situation, we can prove:

Theorem 2.3.1. *Let V be a Hilbert space. Suppose that a convex and non-decreasing C^2-function $k : \mathbb{R} \to \mathbb{R}$ and a convex and lower semicontinuous $j : V \to \mathbb{R}$ are given. Then, $k \circ j$ is convex and lower semicontinuous, and for every $v \in V$ it is true that*

$$\partial(k \circ j)(v) = k'(j(v))\partial j(v). \tag{2.11}$$

If, moreover, $(v, \varphi) \in \operatorname{graph}(\partial j)$ is a tuple such that $k'(j(v)) > 0$ holds and such that j is twice epi-differentiable in v for φ in a direction $z \in V$, then $k \circ j$ is twice epi-differentiable in v for $k'(j(v))\varphi$ in the same direction z and

$$Q_{k\circ j}^{v,k'(j(v))\varphi}(z) = k'(j(v))Q_j^{v,\varphi}(z) + k''(j(v))j'(v;z)^2. \tag{2.12}$$

Proof. The convexity and the lower semicontinuity of $k \circ j$ are trivial, and (2.11) follows immediately from

$$\begin{aligned}\partial(k \circ j)(v) &= \big\{\nu \in V^* \mid \langle \nu, u\rangle \leq k'(j(v))j'(v;u) \ \forall u \in V\big\} \\ &= k'(j(v))\big\{\varphi \in V^* \mid \langle \varphi, u\rangle \leq j'(v;u) \ \forall u \in V\big\} = k'(j(v))\partial j(v) \quad \forall v \in V,\end{aligned}$$

cf. [Borwein and Zhu, 2005, Proposition 4.2.5]. It remains to prove (2.12) and the second-order epi-differentiability of $k \circ j$ in v for $k'(j(v))\varphi$ in the direction z. So let us assume that v, φ, z, j and k satisfy the assumptions of the second part of the theorem. Then, for all $t > 0$ and all $u \in V$, it holds

$$\begin{aligned}&k(j(v + tu)) - k(j(v)) \\ &\quad = k'(j(v))\Big(j(v + tu) - j(v)\Big) \\ &\qquad + \Big(j(v + tu) - j(v)\Big)^2 \int_0^1 (1 - s)k''\Big((1 - s)j(v) + sj(v + tu)\Big)\mathrm{d}s,\end{aligned}$$

and we may compute that

$$\begin{aligned}
&\frac{2}{t}\left(\frac{k(j(v+tu))-k(j(v))}{t}-k'(j(v))\langle\varphi,u\rangle\right)\\
&\quad=k'(j(v))\frac{2}{t}\left(\frac{j(v+tu)-j(v)}{t}-\langle\varphi,u\rangle\right)\\
&\qquad+\frac{2}{t^2}\Big(j(v+tu)-j(v)\Big)^2\int_0^1(1-s)k''\Big((1-s)j(v)+sj(v+tu)\Big)\mathrm{d}s\quad\forall t>0\quad\forall u\in V.
\end{aligned}\tag{2.13}$$

The above implies in combination with the convexity of k, the assumption $k'(j(v))>0$, the local Lipschitz continuity of j (see [Borwein and Zhu, 2005, Theorem 4.1.3]), the dominated convergence theorem and the binomial identities that for all sequences $\{t_n\}\subset\mathbb{R}^+$ and $\{z_n\}\subset V$ with $t_n\searrow 0$ and $z_n\rightharpoonup z$ we have

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{2}{t_n}\left(\frac{k(j(v+t_nz_n))-k(j(v))}{t_n}-k'(j(v))\langle\varphi,z_n\rangle\right)\\
&\geq\liminf_{n\to\infty}k'(j(v))\frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n}-\langle\varphi,z_n\rangle\right)+k''(j(v))\left(\frac{j(v+t_nz_n)-j(v)}{t_n}\right)^2\\
&\quad+\liminf_{n\to\infty}2\left(\frac{j(v+t_nz_n)-j(v)}{t_n}\right)^2\int_0^1(1-s)\left[k''\Big((1-s)j(v)+sj(v+t_nz_n)\Big)-k''(j(v))\right]\mathrm{d}s\\
&=\liminf_{n\to\infty}k'(j(v))\frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n}-\langle\varphi,z_n\rangle\right)+k''(j(v))\left(\frac{j(v+t_nz_n)-j(v)}{t_n}\right)^2\\
&\geq\liminf_{n\to\infty}\Big(k'(j(v))+t_nk''(j(v))\langle\varphi,z_n\rangle\Big)\frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n}-\langle\varphi,z_n\rangle\right)\\
&\quad+\liminf_{n\to\infty}k''(j(v))\langle\varphi,z_n\rangle^2\\
&=k'(j(v))\liminf_{n\to\infty}\frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n}-\langle\varphi,z_n\rangle\right)+k''(j(v))\langle\varphi,z\rangle^2.
\end{aligned}\tag{2.14}$$

If we take the infimum over all possible $\{z_n\}$ and $\{t_n\}$ in (2.14), then we obtain

$$Q^{v,k'(j(v))\varphi}_{k\circ j}(z)\geq k'(j(v))Q^{v,\varphi}_j(z)+k''(j(v))\langle\varphi,z\rangle^2.\tag{2.15}$$

Note that the definition of the reduced critical cone, the Hadamard directional differentiability of j in v (cf. [Bonnans and Shapiro, 2000, Proposition 2.49]), the condition $k'(j(v))>0$ and Lemma 1.3.4 imply

$$k'(j(v))Q^{v,\varphi}_j(z)+k''(j(v))\langle\varphi,z\rangle^2=k'(j(v))Q^{v,\varphi}_j(z)+k''(j(v))j'(v;z)^2.$$

We may thus rewrite (2.15) as

$$Q^{v,k'(j(v))\varphi}_{k\circ j}(z)\geq k'(j(v))Q^{v,\varphi}_j(z)+k''(j(v))j'(v;z)^2.$$

Consider now an arbitrary but fixed $\{t_n\}\subset\mathbb{R}^+$ with $t_n\searrow 0$ and let $\{z_n\}\subset V$ be a recovery sequence for z w.r.t. $\{t_n\}$ as in Definition 1.3.6. Then, we may use exactly the same arguments as in (2.14) to compute

$$\lim_{n\to\infty}\frac{2}{t_n}\left(\frac{k(j(v+t_nz_n))-k(j(v))}{t_n}-k'(j(v))\langle\varphi,z_n\rangle\right)=k'(j(v))Q^{v,\varphi}_j(z)+k''(j(v))j'(v;z)^2.$$

If we combine the above with (2.15) and Definition 1.3.1, then the claim follows immediately. □

Observe that (2.12) has precisely the structure that one would expect from analogous results for smooth functions, cf. the formula of Faà di Bruno, [Krantz and Parks, 2012, Theorem 1.3.2].

2.4 A Second Chain Rule

In this section, we study the situation where a twice continuously Fréchet differentiable map $F : V \to U$ between two Hilbert spaces V and U is composed with a convex, lower semicontinuous, proper and twice epi-differentiable $k : U \to (-\infty, \infty]$. As before, our aim is to derive sufficient conditions for the second-order epi-differentiability of the composition $j := k \circ F : V \to (-\infty, \infty]$ and to prove a formula for the second subderivative $Q^{v,\varphi}_{k\circ F}$. Let us first make precise our standing assumptions:

Assumption 2.4.1 (Standing Assumptions for Section 2.4)**.**

- *U and V are Hilbert spaces,*
- *$k : U \to (-\infty, \infty]$ is a convex, lower semicontinuous and proper function,*
- *$F : V \to U$ is a twice continuously Fréchet differentiable function such that the composition $j := k \circ F : V \to (-\infty, \infty]$ is proper and convex.*

Note that, in contrast to our first chain rule, in the above setting the "original" convex function k and the composition j live on different Hilbert spaces. This makes it much harder to relate the second subderivatives of k and j to each other (cf. the proof of Theorem 2.3.1 where we could use the same recovery sequence $\{z_n\}$ for the original function and the composition). In what follows, the idea behind our analysis is to reduce the problem at hand to the case $j = \chi_K$ and $k = \chi_L$ (analogously to step one in the proof of Theorem 1.2.2) and to subsequently apply a variant of the generalized open mapping theorem of [Zowe and Kurcyusz, 1979] to derive an analogue of (2.12). We begin our investigation by recalling some facts about the convex subdifferential:

Lemma 2.4.2. *Let $u \in U$ be arbitrary but fixed. Then,*

$$\lambda \in \partial k(u) \quad \iff \quad (\lambda, -1) \in \mathcal{N}_{\mathrm{epi}(k)}(u, k(u)).$$

Proof. It holds

$$\begin{aligned}
\lambda \in \partial k(u) \quad &\iff \quad k(\tilde{u}) - k(u) \geq \langle \lambda, \tilde{u} - u \rangle_U && \forall \tilde{u} \in U \\
&\iff \quad 0 \geq \langle (\lambda, -1), (\tilde{u}, \alpha) - (u, k(u)) \rangle_{U\times\mathbb{R}} && \forall (\tilde{u}, \alpha) \in \mathrm{epi}(k) \\
&\iff \quad 0 \geq \langle (\lambda, -1), s \rangle_{U\times\mathbb{R}} && \forall s \in \mathcal{T}_{\mathrm{epi}(k)}(u, k(u)).
\end{aligned}$$

This proves the claim (cf. Definition 1.1.1). □

From the last lemma and classical results on the existence of Lagrange multipliers, we obtain:

Proposition 2.4.3 (Chain Rule for the Convex Subdifferential)**.** *For every $v \in V$ satisfying*

$$F'(v)V - \mathbb{R}^+\Big(\mathrm{dom}(k) - F(v)\Big) = U \tag{2.16}$$

it holds

$$\partial\left(k \circ F\right)(v) = F'(v)^* \partial k(F(v)). \tag{2.17}$$

Here, $F'(v)^* \in L(U^*, V^*)$ denotes the adjoint of the Fréchet derivative $F'(v) \in L(V, U)$.

Proof. Define

$$G : V \times \mathbb{R} \to U \times \mathbb{R}, \qquad (\tilde{v}, \tilde{\alpha}) \mapsto (F(\tilde{v}), \tilde{\alpha}),$$

and set $L := \mathrm{epi}(k) \subset U \times \mathbb{R}$, $M := \mathrm{epi}(k \circ F) \subset V \times \mathbb{R}$. Then, it holds

$$(\tilde{v}, \tilde{\alpha}) \in M \quad \iff \quad \tilde{\alpha} \geq k(F(\tilde{v})) \quad \iff \quad G(\tilde{v}, \tilde{\alpha}) \in L$$

and M is precisely the preimage of L under the function G. Consider now an arbitrary but fixed $v \in V$ satisfying (2.16). We assume w.l.o.g. that $F(v) \in \operatorname{dom}(k)$. If this is not the case, then (2.17) is trivially true since both the left- and the right-hand side are empty. Define $\alpha := k(F(v))$. Then, (2.16) and the definitions of G and L yield

$$G'(v,\alpha)\begin{pmatrix} V \\ \mathbb{R} \end{pmatrix} - \mathbb{R}^+\Big(L - G(v,\alpha)\Big) = \begin{pmatrix} U \\ \mathbb{R} \end{pmatrix} \tag{2.18}$$

and it follows from [Bonnans and Shapiro, 2000, Corollary 2.91, Proposition 2.95] and the convexity, closedness and non-emptiness of the sets L and M that

$$\mathcal{T}_M(v,\alpha) = G'(v,\alpha)^{-1}\,\mathcal{T}_L(F(v),\alpha).$$

In particular, it holds

$$\big\langle G'(v,\alpha)^*\eta, s\big\rangle_{V\times\mathbb{R}} = \big\langle \eta, G'(v,\alpha)s\big\rangle_{U\times\mathbb{R}} \leq 0 \quad \forall s \in \mathcal{T}_M(v,\alpha) \quad \forall \eta \in \mathcal{N}_L(F(v),\alpha)$$

and we may deduce that $G'(v,\alpha)^*\mathcal{N}_L(F(v),\alpha) \subset \mathcal{N}_M(v,\alpha)$. To see that the last inclusion is, in fact, an equality, we fix a $\nu \in \mathcal{N}_M(v,\alpha)$ and consider the auxiliary problem

$$\min\ \langle -\nu, h\rangle_{V\times\mathbb{R}} \quad \text{s.t. } h \in V\times\mathbb{R},\ \ G'(v,\alpha)h \in \mathcal{T}_L(G(v,\alpha)). \tag{2.19}$$

Note that, since $\langle -\nu, h\rangle_{V\times\mathbb{R}} \geq 0$ for all $h \in \mathcal{T}_M(v,\alpha) = G'(v,\alpha)^{-1}\mathcal{T}_L(G(v,\alpha))$, (2.19) is obviously solved by $\tilde{h} := 0 \in V\times\mathbb{R}$. Further, (2.18) yields

$$G'(v,\alpha)\begin{pmatrix} V \\ \mathbb{R} \end{pmatrix} - \mathbb{R}^+\Big(\mathcal{T}_L(G(v,\alpha)) - 0\Big) \supset G'(v,\alpha)\begin{pmatrix} V \\ \mathbb{R} \end{pmatrix} - \mathbb{R}^+\Big(L - G(v,\alpha)\Big) = \begin{pmatrix} U \\ \mathbb{R} \end{pmatrix}.$$

The latter implies that the constraint qualification of Zowe-Kurcyusz is satisfied for (2.19) in $\tilde{h}$, that we can find a Lagrange multiplier $\eta \in \mathcal{N}_L(G(v,\alpha))$ with $\nu = G'(v,\alpha)^*\eta$ (see [Bonnans and Shapiro, 2000, Theorem 3.9]) and that $G'(v,\alpha)^*\mathcal{N}_L(F(v),\alpha) \supset \mathcal{N}_M(v,\alpha)$ (since ν was arbitrary). Combining all of the above, we arrive at the desired identity $G'(v,\alpha)^*\mathcal{N}_L(F(v),\alpha) = \mathcal{N}_M(v,\alpha)$. Consider now an arbitrary but fixed $\lambda \in \partial k(F(v))$. Then, it follows from Lemma 2.4.2 and the definitions of G, L and M that $G'(v,\alpha)^*(\lambda,-1) = (F'(v)^*\lambda,-1) \in \mathcal{N}_M(v,\alpha)$ and we may deduce that $F'(v)^*\lambda \in \partial(k\circ F)(v)$. This proves $F'(v)^*\partial k(F(v)) \subset \partial(k\circ F)(v)$. If, conversely, we start with a $\varphi \in \partial(k\circ F)(v)$, then we know that $(\varphi,-1) \in \mathcal{N}_M(v,\alpha) = G'(v,\alpha)^*\mathcal{N}_L(F(v),\alpha)$ and there exists at least one $\lambda \in \partial k(F(v))$ with $G'(v,\alpha)^*(\lambda,-1) = (F'(v)^*\lambda,-1) = (\varphi,-1)$. Consequently, $\partial\,(k\circ F)\,(v) \subset F'(v)^*\partial k(F(v))$ and the proof of the proposition is complete. □

Remark 2.4.4. *In the literature, the chain rule for the convex subdifferential $\partial(k\circ F)(v)$ in a point $v \in V$ is often proved under the assumption that there exists an $\tilde{h} \in V$ such that k is finite and continuous in $F(v) + F'(v)\tilde{h}$. See, e.g., [Ekeland and Temam, 1976, Proposition 5.7], [Schiela and Wollner, 2011, Lemma 3.4] and [Peypouquet, 2015, Proposition 3.28]. We would like to point out that this condition is more restrictive than our assumption* (2.16)*. Indeed, if $\tilde{h}$ is such that k is continuous and finite in $F(v) + F'(v)\tilde{h}$, then there exists an $\varepsilon > 0$ with $F(v) + F'(v)\tilde{h} + B_\varepsilon(0) \subset \operatorname{dom}(k)$ (where $B_\varepsilon(0)$ denotes the closed ball of radius $\varepsilon > 0$ around the origin), and we may compute that*

$$F'(v)V - \mathbb{R}^+\Big(\operatorname{dom}(k) - F(v)\Big) \supset \alpha F'(v)\tilde{h} - \alpha\Big(F'(v)\tilde{h} + B_\varepsilon(0)\Big) = B_{\alpha\varepsilon}(0) \quad \forall \alpha > 0.$$

The latter immediately yields (2.16)*. It should be noted that the $\tilde{h}$-condition described above corresponds to a linearized Slater condition on the epigraph-level, cf. the proof of Proposition 2.4.3 and [Bonnans and Shapiro, 2000, Lemma 2.99]. We remark that, in contrast to the $\tilde{h}$-assumption,* (2.16) *can also be satisfied when the interior of the domain $\operatorname{dom}(k)$ is empty. Consider, e.g., the case where the derivative $F'(v) \in L(V,U)$ is surjective.*

Using Proposition 2.4.3, we can prove a first result on the relationship between the second subderivatives of the functions k and $j = k \circ F$:

Proposition 2.4.5. *Suppose that a $v \in V$ and a $\lambda \in \partial k(F(v))$ are given. Assume that* (2.16) *is satisfied and that the map $z \mapsto \langle \lambda, F''(v)z^2 \rangle$ is weakly lower semicontinuous. Then, $\varphi := F'(v)^*\lambda$ is an element of the subdifferential $\partial(k \circ F)(v)$ and it holds*

$$Q^{v,\varphi}_{k\circ F}(z) \geq Q^{F(v),\lambda}_k(F'(v)z) + \langle \lambda, F''(v)z^2 \rangle \qquad \forall z \in V. \tag{2.20}$$

Proof. From Proposition 2.4.3, we readily obtain that $\varphi \in \partial(k \circ F)(v)$. Consider now an arbitrary but fixed $z \in V$ and let $\{t_n\} \subset \mathbb{R}^+$ and $\{z_n\} \subset V$ be sequences with $t_n \searrow 0$ and $z_n \rightharpoonup z$ in V. Then, a second-order Taylor expansion yields

$$F(v + t_n z_n) = F(v) + t_n F'(v)z_n + \frac{1}{2}t_n^2 F''(v)z_n^2 + o(t_n^2) = F(v) + t_n y_n$$

with $y_n := F'(v)z_n + \frac{1}{2}t_n F''(v)z_n^2 + o(t_n) \rightharpoonup F'(v)z$ for $n \to \infty$ and we may compute that

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{k(F(v + t_n z_n)) - k(F(v))}{t_n} - \langle \varphi, z_n \rangle\right) \\
&\quad = \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{k(F(v) + t_n y_n) - k(F(v))}{t_n} - \left\langle \lambda, y_n - \frac{1}{2}t_n F''(v)z_n^2 - o(t_n) \right\rangle\right) \\
&\quad \geq Q^{F(v),\lambda}_k(F'(v)z) + \langle \lambda, F''(v)z^2 \rangle .
\end{aligned}$$

Taking the infimum over all $\{t_n\}$ and $\{z_n\}$ in the above, (2.20) follows immediately. □

Note that Proposition 2.4.5 does not require any second-order epi-differentiability assumptions whatsoever. To obtain an estimate reverse to (2.20), we observe the following:

Proposition 2.4.6. *Let $(u, \lambda) \in \operatorname{graph}(\partial k)$ be arbitrary but fixed and denote the characteristic function of the epigraph $\operatorname{epi}(k)$ with $\chi_{\operatorname{epi}(k)} : U \times \mathbb{R} \to \{0, \infty\}$. Then, it holds*

$$Q^{u,\lambda}_k(z) = Q^{(u,k(u)),(\lambda,-1)}_{\chi_{\operatorname{epi}(k)}}(z, \langle \lambda, z \rangle) \qquad \forall z \in U. \tag{2.21}$$

Moreover, k is twice epi-differentiable in u for λ in a direction $z \in U$ if and only if $\chi_{\operatorname{epi}(k)}$ is twice epi-differentiable in $(u, k(u))$ for $(\lambda, -1)$ in the direction $(z, \langle \lambda, z \rangle)$.

Proof. First, we note that $\partial k(u) \times \{-1\} \subset \mathcal{N}_{\operatorname{epi}(k)}(u, k(u)) = \partial \chi_{\operatorname{epi}(k)}(u, k(u))$ by Lemma 2.4.2. This shows that it makes sense to talk about the functional on the right-hand side of (2.21). Consider now an arbitrary but fixed $z \in \mathcal{K}^{red}_k(u, \lambda)$ and some sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset U$ with $t_n \searrow 0$, $z_n \rightharpoonup z$ and

$$\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{k(u + t_n z_n) - k(u)}{t_n} - \langle \lambda, z_n \rangle\right) < \infty. \tag{2.22}$$

Then, after passing over to a subsequence (unrelabeled) that realizes the limes inferior in (2.22) and that satisfies $u + t_n z_n \in \operatorname{dom}(k)$ for all n, it holds

$$\frac{1}{t_n}\Big(\operatorname{epi}(k) - (u, k(u))\Big) \ni y_n := \left(z_n, \frac{k(u + t_n z_n) - k(u)}{t_n}\right) \rightharpoonup (z, \langle \lambda, z \rangle)$$

in $U \times \mathbb{R}$ as $n \to \infty$ and

$$\begin{aligned}
&\frac{2}{t_n}\left(\frac{k(u + t_n z_n) - k(u)}{t_n} - \langle \lambda, z_n \rangle\right) \\
&\qquad = \frac{2}{t_n}\left(\frac{\chi_{\operatorname{epi}(k)}\Big((u, k(u)) + t_n y_n\Big) - \chi_{\operatorname{epi}(k)}(u, k(u))}{t_n} - \langle (\lambda, -1), y_n \rangle\right).
\end{aligned}$$

The latter implies in combination with Definition 1.3.1 that

$$Q_k^{u,\lambda}(z) \geq Q_{\chi_{\mathrm{epi}(k)}}^{(u,k(u)),(\lambda,-1)}(z,\langle\lambda,z\rangle) \quad \forall z \in \mathcal{K}_k^{red}(u,\lambda). \tag{2.23}$$

If, conversely, we start with a $z \in U$ satisfying $(z,\langle\lambda,z\rangle) \in \mathcal{K}^{red}_{\chi_{\mathrm{epi}(k)}}((u,k(u)),(\lambda-1))$ and assume that $\{t_n\} \subset \mathbb{R}^+$ and $\{y_n\} \subset U \times \mathbb{R}$ are sequences with $t_n \searrow 0$, $y_n := (z_n,\alpha_n) \rightharpoonup (z,\langle\lambda,z\rangle)$ and

$$\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{\chi_{\mathrm{epi}(k)}\Big((u,k(u))+t_n y_n\Big) - \chi_{\mathrm{epi}(k)}(u,k(u))}{t_n} - \langle(\lambda,-1),y_n\rangle\right) < \infty, \tag{2.24}$$

then, after passing over to a subsequence (again unrelabeled) that realizes the limes inferior in (2.24) and that satisfies $(u,k(u))+t_n y_n \in \mathrm{epi}(k)$ for all n, we obtain

$$\begin{aligned}
&\frac{2}{t_n}\left(\frac{\chi_{\mathrm{epi}(k)}\Big((u,k(u))+t_n y_n\Big) - \chi_{\mathrm{epi}(k)}(u,k(u))}{t_n} - \langle(\lambda,-1),y_n\rangle\right)\\
&\quad = \frac{2}{t_n}\left(\frac{k(u)+t_n\alpha_n - k(u)}{t_n} - \langle\lambda,z_n\rangle\right) \geq \frac{2}{t_n}\left(\frac{k(u+t_n z_n)-k(u)}{t_n} - \langle\lambda,z_n\rangle\right).
\end{aligned}$$

Taking the infimum over all sequences $\{t_n\}$, $\{y_n\}$ in the above, we arrive at the inequality

$$Q_{\chi_{\mathrm{epi}(k)}}^{(u,k(u)),(\lambda,-1)}(z,\langle\lambda,z\rangle) \geq Q_k^{u,\lambda}(z) \quad \forall z \in U \text{ with } (z,\langle\lambda,z\rangle) \in \mathcal{K}^{red}_{\chi_{\mathrm{epi}(k)}}((u,k(u)),(\lambda-1)). \tag{2.25}$$

If we combine (2.23) and (2.25), then (2.21) follows immediately. The proof of the second part of the lemma is completely along the same lines. □

Proposition 2.4.6 shows that it suffices to study the second-order epi-differentiability and the second subderivative of the characteristic function $\chi_{\mathrm{epi}(k)}$ to fully understand the second-order behavior of the function k. This is very advantageous because it allows to reduce the situation in Assumption 2.4.1 to a setting where classical pull-back and push-forward results are applicable (cf. also with the proof of Proposition 2.4.3 in this context). The transformation result that we will employ in the following is a corollary of the (set-valued) open mapping theorem that goes back to [Werner, 1984].

We would like to point out that Lemma 2.4.7 below and the subsequent Theorem 2.4.8 have originally been prepared (in a slightly less general format) in a joint work with Gerd Wachsmuth for the article [Christof and Wachsmuth, 2017b] (but ultimately remained unpublished).

Lemma 2.4.7 ([Werner, 1984, Corollary 5.2.4, Proof of Theorem 5.2.5])**.** *Let X and Y be Banach spaces, let $L \subset Y$ be a non-empty, closed and convex set, and let $G : X \to Y$ be a twice continuously Fréchet differentiable function. Suppose that an $x \in G^{-1}(L)$ and an $\eta \in Y^*$ are given such that*

$$G'(x)X - \mathbb{R}^+\left[L - G(x)\right] \cap \ker(\eta) = Y,$$

where $\ker(\eta)$ denotes the kernel of η. Then, there exists a $C > 0$ such that the following holds true:

(i) For every $y \in Y$ there exist a $v \in X$ and a $u \in Y$ with

$$\|v\|_X \leq C\|y\|_Y, \quad u \in C\|y\|_Y\left[B_1^Y(0) \cap (L - G(x)) \cap \ker(\eta)\right], \quad y = G'(x)v - u.$$

(ii) For every $h \in X$ there exist maps

$$\rho : \left[0, \frac{1}{C\|h\|_X}\right] \to X, \qquad \zeta : \left[0, \frac{1}{C\|h\|_X}\right] \to Y$$

with

$$\begin{aligned}
&\|\rho(t)\|_X \leq C\|G(x+th)-G(x)-tG'(x)h\|_Y,\\
&\zeta(t) \in C\|G(x+th)-G(x)-tG'(x)h\|_Y \left[B_1^Y(0)\cap(L-G(x))\cap\ker(\eta)\right],\\
&G(x)+tG'(x)h = G(x+th+\rho(t))-\zeta(t) \qquad \forall t\in\left[0,\frac{1}{C\|h\|_X}\right].
\end{aligned}$$

We are now in the position to prove the main result of this section.

Theorem 2.4.8. *Let U, V, F and k satisfy Assumption 2.4.1. Suppose that a $v \in V$, a $\lambda \in \partial k(F(v))$ and a $z \in V$ are given, such that the map $\tilde{z} \mapsto \langle \lambda, F''(v)\tilde{z}^2\rangle$ is weakly lower semicontinuous, such that*

$$F'(v)V - \mathbb{R}^+\Big\{u-F(v)\;\Big|\; u\in\operatorname{dom}(k),\ k(u)-k(F(v)) = \langle\lambda, u-F(v)\rangle\Big\} = U, \tag{2.26}$$

and such that k is twice epi-differentiable in $F(v)$ for λ in the direction $F'(v)z$. Then, $\varphi := F'(v)^*\lambda$ is an element of $\partial(k\circ F)(v)$, $k\circ F$ is twice epi-differentiable in v for φ in the direction z and it holds

$$Q^{v,\varphi}_{k\circ F}(z) = Q^{F(v),\lambda}_k(F'(v)z) + \left\langle\lambda, F''(v)z^2\right\rangle. \tag{2.27}$$

Proof. Note that (2.26) implies (2.16). This shows that Propositions 2.4.3 and 2.4.5 are applicable, that $\varphi \in \partial(k\circ F)(v)$ and that

$$Q^{v,\varphi}_{k\circ F}(z) \geq Q^{F(v),\lambda}_k(F'(v)z) + \left\langle\lambda, F''(v)z^2\right\rangle. \tag{2.28}$$

It remains to prove the converse of (2.28) and the second-order epi-differentiability of $k\circ F$ in v for φ in the direction z. To obtain the latter, we proceed in two steps:

Step 1 (Reduction to the case $k = \chi_L$): As in the proofs of Theorem 1.2.2 and Proposition 2.4.3, we first simplify the problem by rewriting it using the epigraph. Define

$$G: V\times\mathbb{R}\to U\times\mathbb{R}, \qquad (\tilde{v},\tilde{\alpha})\mapsto(F(\tilde{v}),\tilde{\alpha}),$$

and set $L := \operatorname{epi}(k)\subset U\times\mathbb{R}$, $M := \operatorname{epi}(k\circ F) = G^{-1}(L)$, $x := (v, k(F(v)))\in M$, $h := (z,\langle\varphi, z\rangle)$, $\eta := (\lambda,-1)\in U^*\times\mathbb{R}$. Then, it follows from Proposition 2.4.6 that

$$Q^{v,\varphi}_{k\circ F}(z) = Q^{(v,k(F(v))),(\varphi,-1)}_{\chi_{\operatorname{epi}(k\circ F)}}(z,\langle\varphi,z\rangle) = Q^{x,G'(x)^*\eta}_{\chi_M}(h), \tag{2.29}$$

that

$$Q^{F(v),\lambda}_k(F'(v)z) = Q^{(F(v),k(F(v))),(\lambda,-1)}_{\chi_{\operatorname{epi}(k)}}(F'(v)z,\left\langle\lambda,F'(v)z\right\rangle) = Q^{G(x),\eta}_{\chi_L}(G'(x)h), \tag{2.30}$$

and that χ_L is twice epi-differentiable in $G(x)$ for η in the direction $G'(x)h$. In what follows, our aim will be to show that

$$Q^{x,G'(x)^*\eta}_{\chi_M}(h) \leq Q^{G(x),\eta}_{\chi_L}(G'(x)h) + \left\langle\lambda, F''(v)z^2\right\rangle \tag{2.31}$$

and that χ_M is twice epi-differentiable in x for $G'(x)^*\eta$ in the direction h. If this is established, then the claim follows immediately from (2.28) and Proposition 2.4.6.

Step 2 (Proof on the epigraph-level): We may assume w.l.o.g. that $Q^{G(x),\eta}_{\chi_L}(G'(x)h) < \infty$. If this is not the case, then (2.28) yields

$$Q^{v,\varphi}_{k\circ F}(z) \geq Q^{G(x),\eta}_{\chi_L}(G'(x)h) + \left\langle\lambda, F''(v)z^2\right\rangle = \infty$$

and the claim is trivially true. The claim is further trivial if $z = 0$, so we assume w.l.o.g. that $z\neq 0$ (and thus $h\neq 0$). Consider now an arbitrary but fixed $\{t_n\}\subset\mathbb{R}^+$ with $t_n\searrow 0$ and define $s_n := t_n + t_n^2$.

Then, it follows from the second-order epi-differentiability of χ_L in $G(x)$ for η in the direction $G'(x)h$ that there exists a sequence $\{w_n\} \subset U \times \mathbb{R}$ with $s_n\, w_n \to 0$ for $n \to \infty$ and

$$\begin{aligned} &Q_{\chi_L}^{G(x),\eta}(G'(x)h) \\ &\quad = \lim_{n\to\infty} \frac{2}{s_n} \left(\frac{\chi_L\left(G(x) + s_n G'(x)h + \frac{1}{2}s_n^2 w_n\right) - \chi_L(G(x))}{s_n} - \left\langle \eta, G'(x)h + \frac{1}{2}s_n w_n \right\rangle \right). \end{aligned} \tag{2.32}$$

Note that (2.32) implies that for all large enough n we have $G(x) + s_n G'(x)h + \frac{1}{2}s_n^2 w_n \in L$ and that

$$\big\langle \eta, G'(x)h \big\rangle = \big\langle \lambda, F'(v)z \big\rangle - \langle \varphi, z \rangle = 0. \tag{2.33}$$

We may thus assume w.l.o.g. that

$$s_n w_n \to 0, \qquad G(x) + s_n G'(x)h + \frac{1}{2}s_n^2 w_n \in L, \qquad Q_{\chi_L}^{G(x),\eta}(G'(x)h) = \lim_{n\to\infty} \langle -\eta, w_n \rangle. \tag{2.34}$$

From (2.26), we further obtain that

$$G'(x)\begin{pmatrix} V \\ \mathbb{R} \end{pmatrix} - \mathbb{R}^+\Big(L - G(x)\Big) \cap \ker(\eta) = \begin{pmatrix} U \\ \mathbb{R} \end{pmatrix}.$$

The latter implies in combination with Lemma 2.4.7 that there exist a constant $C > 0$ (independent of n) and sequences $v_n \in V \times \mathbb{R}$, $u_n \in U \times \mathbb{R}$ with

$$\begin{aligned} &\|v_n\|_{V\times\mathbb{R}} \le \frac{C}{2}\|s_n w_n\|_{U\times\mathbb{R}}, \quad \frac{1}{2}s_n w_n = G'(x)v_n - u_n, \\ &u_n \in \frac{C}{2}\|s_n w_n\|_{U\times\mathbb{R}} \left[B_1^{U\times\mathbb{R}}(0) \cap (L - G(x)) \cap \ker(\eta) \right]. \end{aligned} \tag{2.35}$$

Note that v_n and u_n converge to zero for $n \to \infty$ by the properties of w_n. Define $h_n := v_n + h$. Then, it holds $h_n \to h$ in $V \times \mathbb{R}$ and we may employ part (ii) of Lemma 2.4.7 to infer that there exist maps

$$\rho_n : \left[0, \frac{1}{C\|h_n\|_{V\times\mathbb{R}}}\right] \to V \times \mathbb{R}, \qquad \zeta_n : \left[0, \frac{1}{C\|h_n\|_{V\times\mathbb{R}}}\right] \to U \times \mathbb{R}$$

with

$$\begin{aligned} &\|\rho_n(t)\|_{V\times\mathbb{R}} \le C\|G(x + th_n) - G(x) - tG'(x)h_n\|_{U\times\mathbb{R}}, \\ &\zeta_n(t) \in C\|G(x + th_n) - G(x) - tG'(x)h_n\|_{U\times\mathbb{R}} \left[B_1^{U\times\mathbb{R}}(0) \cap (L - G(x)) \cap \ker(\eta) \right], \\ &G(x) + tG'(x)h_n = G(x + th_n + \rho_n(t)) - \zeta_n(t) \qquad \forall t \in \left[0, \frac{1}{C\|h_n\|_{V\times\mathbb{R}}}\right]. \end{aligned} \tag{2.36}$$

Here, C is the same constant as in (2.35). Since $C\|h_n\|_{V\times\mathbb{R}} \to C\|h\|_{V\times\mathbb{R}} > 0$ and $t_n \to 0$ for $n \to \infty$, at least for all large enough n, we may define $r_n := \rho_n(t_n) \in V \times \mathbb{R}$ and $z_n := \zeta_n(t_n) \in U \times \mathbb{R}$. The properties of the sequences r_n, h_n, z_n, w_n, t_n and s_n now yield

$$\begin{aligned} &G(x + t_n h_n + r_n) \\ &\quad = G(x) + t_n G'(x)h_n + z_n && \text{(by (2.36))} \\ &\quad = G(x) + t_n G'(x)(h + v_n) + z_n \\ &\quad = G(x) + t_n \left(G'(x)h + \frac{1}{2}s_n w_n + u_n \right) + z_n && \text{(by (2.35))} \\ &\quad = \frac{1}{1+t_n}G(x) + \frac{s_n}{1+t_n}\left(G'(x)h + \frac{1}{2}s_n w_n \right) + \frac{t_n}{1+t_n}G(x) + z_n + \frac{s_n}{1+t_n}u_n \\ &\quad = \frac{1}{1+t_n}\left(G(x) + s_n G'(x)h + \frac{1}{2}s_n^2 w_n \right) + \frac{t_n}{1+t_n}\left(G(x) + \frac{1+t_n}{t_n}z_n + \frac{s_n}{t_n}u_n \right). \end{aligned} \tag{2.37}$$

Note that, from (2.35) and (2.36), it follows that there exist $\{y_n^u\} \subset L$ and $\{y_n^z\} \subset L$ with

$$u_n = \varepsilon_n^u(y_n^u - G(x)), \qquad \varepsilon_n^u := \frac{C}{2}\|s_n w_n\|_{U\times\mathbb{R}} = o(1),$$
$$z_n = \varepsilon_n^z\left(y_n^z - G(x)\right), \qquad \varepsilon_n^z := C\|G(x+t_nh_n) - G(x) - t_nG'(x)h_n\|_{U\times\mathbb{R}} = o(t_n).$$

The above, (2.37) and the convexity of L imply that

$$a_n := \left(1 - 2\frac{s_n}{t_n}\varepsilon_n^u\right)G(x) + \left(2\frac{s_n}{t_n}\varepsilon_n^u\right)y_n^u \in L,$$
$$b_n := \left(1 - 2\frac{1+t_n}{t_n}\varepsilon_n^z\right)G(x) + \left(2\frac{1+t_n}{t_n}\varepsilon_n^z\right)y_n^z \in L,$$

and

$$\begin{aligned} G(x+t_nh_n+r_n) &= \frac{1}{1+t_n}\left(G(x) + s_nG'(x)h + \frac{1}{2}s_n^2w_n\right) + \frac{t_n}{1+t_n}\left(G(x) + \frac{1+t_n}{t_n}z_n + \frac{s_n}{t_n}u_n\right) \\ &= \frac{1}{1+t_n}\left(G(x) + s_nG'(x)h + \frac{1}{2}s_n^2w_n\right) + \frac{t_n}{1+t_n}\left(\frac{1}{2}a_n + \frac{1}{2}b_n\right) \in L \end{aligned}$$

for all large enough n. Since $\|r_n\|_{V\times\mathbb{R}} \le C\|G(x+t_nh_n) - G(x) - t_nG'(x)h_n\|_{U\times\mathbb{R}} = o(t_n)$ and $M = G^{-1}(L)$, we now arrive at the situation

$$x + t_nh_n + r_n \in M, \qquad h_n + \frac{r_n}{t_n} \to h, \qquad t_n \to 0.$$

The above implies in combination with (2.28), (2.29), (2.30) and Definition 1.3.1 that

$$\begin{aligned} &Q_{\chi_L}^{G(x),\eta}(G'(x)h) + \left\langle \lambda, F''(v)z^2\right\rangle \\ &\quad\le Q_{\chi_M}^{x,G'(x)^*\eta}(h) \\ &\quad\le \liminf_{n\to\infty}\frac{2}{t_n}\left(\frac{\chi_M\left(x+t_nh_n+r_n\right) - \chi_M(x)}{t_n} - \left\langle G'(x)^*\eta, h_n + \frac{r_n}{t_n}\right\rangle\right) \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, G'(x)(t_nh_n+r_n)\right\rangle \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, G(x+t_nh_n+r_n) - G(x) - \frac{1}{2}G''(x)(t_nh_n+r_n)^2\right\rangle \qquad \text{(by Taylor)} \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, t_nG'(x)h_n + z_n - \frac{1}{2}G''(x)(t_nh)^2\right\rangle \qquad \text{(by (2.36))} \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, t_nG'(x)(h+v_n) - \frac{1}{2}t_n^2G''(x)h^2\right\rangle \qquad \text{(by (2.36))} \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, t_nG'(x)v_n\right\rangle + \left\langle \lambda, F''(v)z^2\right\rangle \qquad \text{(by (2.33))} \\ &\quad= \liminf_{n\to\infty}\frac{-2}{t_n^2}\left\langle \eta, t_n\frac{1}{2}s_nw_n + t_nu_n\right\rangle + \left\langle \lambda, F''(v)z^2\right\rangle \qquad \text{(by (2.35))} \\ &\quad= \liminf_{n\to\infty}\frac{t_ns_n}{t_n^2}\left\langle -\eta, w_n\right\rangle + \left\langle \lambda, F''(v)z^2\right\rangle \qquad \text{(by (2.35))} \\ &\quad= \lim_{n\to\infty}(1+t_n)\left\langle -\eta, w_n\right\rangle + \left\langle \lambda, F''(v)z^2\right\rangle \\ &\quad= Q_{\chi_L}^{G(x),\eta}(G'(x)h) + \left\langle \lambda, F''(v)z^2\right\rangle. \qquad \text{(by (2.34))} \end{aligned}$$

The above shows, on the one hand, that equality holds in (2.31), and, on the other hand, that

$$Q_{\chi_M}^{x,G'(x)^*\eta}(h) = \lim_{n\to\infty}\frac{2}{t_n}\left(\frac{\chi_M\left(x+t_nh_n+r_n\right) - \chi_M(x)}{t_n} - \left\langle G'(x)^*\eta, h_n + \frac{r_n}{t_n}\right\rangle\right).$$

This proves that χ_M is indeed twice epi-differentiable in x for $G'(x)^*\eta$ in the direction h. The claim of the theorem now follows immediately (see step one). □

Remark 2.4.9.

(i) In the case $k = \chi_L$, $L \subset U$ closed, convex, non-empty, (2.26) *takes the form*

$$F'(v)V - \mathbb{R}^+(L - F(v)) \cap \ker(\lambda) = U.$$

This Zowe/Kurcyusz-type condition is well-known, e.g., in the study of the uniqueness of Lagrange multipliers, cf. [Shapiro, 1997, Theorem 2.2]. Note that it makes sense that the condition (2.26) *in Theorem 2.4.8 is more restrictive than the assumption* (2.16) *in Proposition 2.4.3. The former result makes, after all, a statement about subderivatives of a higher order.*

(ii) A result similar to Theorem 2.4.8 can be found in [Ioffe, 1991, Section 2] (see also [Poliquin and Rockafellar, 1993] for an analysis in finite dimensions). We would like to point out that the chain rule in [Ioffe, 1991] neither yields a formula analogous to (2.27) *nor gives any information about the second-order epi-differentiability properties of the composition $k \circ F$. Theorem 2.4.8 is more precise in this regard and, as a consequence, seems to be more suitable for practical applications, cf. the examples in Chapters 3 and 4.*

2.5 Superposition with Twice Epi-Differentiable Functions

We conclude this chapter by studying the second-order epi-differentiability properties of functions that involve superposition operators. The question, that we are mainly concerned with in the present section, is that of whether (and, when yes, how) a functional of the form

$$j : L^2(\Omega, H) \to (-\infty, \infty], \qquad v \mapsto \int_\Omega k(v)\mathrm{d}\mu, \tag{2.38}$$

inherits the property of second-order epi-differentiability from its inner map $k : H \to (-\infty, \infty]$. Our standing assumptions on the quantities in (2.38) are as follows:

Assumption 2.5.1 (Standing Assumptions and Notation for Section 2.5)**.**

- *H is a separable Hilbert space,*
- *(Ω, Σ, μ) is a finite and complete measure space,*
- *$V := L^2(\Omega, H)$ (where $L^2(\Omega, H)$ is defined as in [Heinonen et al., 2015, Section 3.2]),*
- *$k : H \to (-\infty, \infty]$ is a convex, lower semicontinuous and proper function.*

Before we begin our actual analysis, let us verify that (2.38) indeed defines a function that fits into the setting of Chapter 1:

Lemma 2.5.2. *The functional j in* (2.38) *is well-defined, convex, lower semicontinuous and proper.*

Proof. We first check that we do not run into any measurability/integrability problems: If $v : \Omega \to H$ is an arbitrary but fixed Bochner measurable function (i.e., the pointwise a.e.-limit of a sequence of simple functions), then it follows from [Heinonen et al., 2015, Corollary 3.1.2] and the separability of H that $v^{-1}(D)$ is an element of Σ for all open $D \subset H$. Since the open sets in H generate the Borel σ-algebra $\mathfrak{B}(H)$, the latter implies that v is (Ω, Σ)-$(H, \mathfrak{B}(H))$-measurable. From the lower semicontinuity of the function $k : H \to (-\infty, \infty]$, we obtain further that the sets $k^{-1}([-\infty, \alpha])$ are closed for all $\alpha \in \mathbb{R}$. This yields that the preimage $k^{-1}(D)$ of any open set D in $[-\infty, \infty]$ is Borel and that the composition

$k \circ v : \Omega \to [-\infty, \infty]$ is measurable in the sense of [Evans and Gariepy, 2015, Section 1.1.4]. Consider now an arbitrary but fixed $v \in L^2(\Omega, H)$, choose an $l \in H^*$ and a $c \in \mathbb{R}$ with

$$k(h) \geq \langle l, h \rangle + c \qquad \forall h \in H \tag{2.39}$$

(such l and c always exist, cf. the proof of Theorem 1.2.2), and decompose $k \circ v$ into the functions

$$k_1 := k(v) - \langle l, v \rangle - c \quad \text{and} \quad k_2 := \langle l, v \rangle + c.$$

Then, our measurability results, the finiteness of the measure space (Ω, Σ, μ) and the property (2.39) yield that $k_1 \in L^0(\Omega, [0, \infty])$ and that $k_2 \in L^2(\Omega, \mathbb{R})$ (where, as usual, the prefix "L^0" indicates that we talk about the vector space of equivalence classes of measurable functions). The latter implies in particular that the integrals

$$\int_\Omega k_1 \mathrm{d}\mu \in [0, \infty] \quad \text{and} \quad \int_\Omega k_2 \mathrm{d}\mu \in \mathbb{R}$$

are both well-defined (the first one due to the non-negativity of the function k_1, see [Evans and Gariepy, 2015, Section 1.3], and the second one trivially). If we combine all of the above and write

$$\int_\Omega k(v) \mathrm{d}\mu = \int_\Omega k(v) - \langle l, v \rangle - c \, \mathrm{d}\mu + \int_\Omega \langle l, v \rangle + c \, \mathrm{d}\mu \in (-\infty, \infty], \tag{2.40}$$

then it follows immediately that the integral in (2.38) is sensible and that (2.38) indeed defines a function on $L^2(\Omega, H)$. It remains to prove that j is proper, convex and lower semicontinuous. Note that the first two of these properties are trivial due to the monotonicity of the integral, the properness and convexity of k, and the finiteness of the measure space (Ω, Σ, μ). To see that j is lower semicontinuous, let us consider a sequence $v_n \in L^2(\Omega, H)$ with $v_n \to v$ in $L^2(\Omega, H)$. By passing over to a subsequence (unrelabeled), we may assume w.l.o.g. that v_n converges to v pointwise almost everywhere in Ω for $n \to \infty$. Using this pointwise a.e.-convergence, the lemma of Fatou (see [Evans and Gariepy, 2015, Theorem 1.17]), (2.40) and the lower semicontinuity of k, we obtain that

$$\begin{aligned}
\liminf_{n\to\infty} j(v_n) &= \liminf_{n\to\infty} \left(\int_\Omega k(v_n) - \langle l, v_n \rangle - c \, \mathrm{d}\mu + \int_\Omega \langle l, v_n \rangle + c \, \mathrm{d}\mu \right) \\
&\geq \int_\Omega \liminf_{n\to\infty} \left(k(v_n) - \langle l, v_n \rangle - c \right) \mathrm{d}\mu + \int_\Omega \langle l, v \rangle + c \, \mathrm{d}\mu \\
&\geq \int_\Omega k(v) - \langle l, v \rangle - c \, \mathrm{d}\mu + \int_\Omega \langle l, v \rangle + c \, \mathrm{d}\mu \\
&= j(v).
\end{aligned}$$

This establishes the lower semicontinuity of j and completes the proof. □

In what follows, the main idea of our analysis is to study the differentiability properties of the solution operator $S : L^2(\Omega, H)^* \cong L^2(\Omega, H^*) \to L^2(\Omega, H)$ to the prototypical EVI

$$w \in V, \qquad (w, v - w)_V + \int_\Omega k(v) \mathrm{d}\mu - \int_\Omega k(w) \mathrm{d}\mu \geq \langle f, v - w \rangle_V \qquad \forall v \in V \tag{2.41}$$

and to subsequently establish the second-order epi-differentiability of the functional j via the equivalence in Theorem 1.4.1. To pursue this approach, we prove:

Lemma 2.5.3. *The variational inequality* (2.41) *admits a unique solution $w := S(f) \in L^2(\Omega, H)$ for all $f \in L^2(\Omega, H^*)$. This solution satisfies $S(f) = T(f)$ a.e. in Ω, where $T : H^* \to H$, $\tilde{f} \mapsto \tilde{w}$, is the solution map to the EVI*

$$\tilde{w} \in H, \qquad (\tilde{w}, \tilde{v} - \tilde{w})_H + k(\tilde{v}) - k(\tilde{w}) \geq \langle \tilde{f}, \tilde{v} - \tilde{w} \rangle_H \qquad \forall \tilde{v} \in H. \tag{2.42}$$

Proof. The unique solvability of the EVIs (2.41) and (2.42) follows immediately from Theorem 1.2.2. It remains to prove the identity $S(f) = T(f)$ a.e. in Ω. To this end, we first note that the global Lipschitz continuity of the map $T : H^* \to H$ implies that the function $L^2(\Omega, H^*) \ni f \mapsto T(f) \in L^2(\Omega, H)$ is well-defined. Consider now an arbitrary but fixed $f \in L^2(\Omega, H^*)$, the associated solution $S(f)$, and the function $T(f) \in L^2(\Omega, H)$ that is obtained by superposition. Then, it follows from (2.42) that

$$(T(f), v - T(f))_H + k(v) - k(T(f)) \geq \langle f, v - T(f) \rangle_H$$

holds a.e. in Ω for all $v \in L^2(\Omega, H)$, and we obtain by integration that $T(f)$ solves (2.41). Since (2.41) can have only one solution, we now arrive at the desired identity $S(f) = T(f)$ a.e. in Ω. This completes the proof. □

As a first consequence of Lemma 2.5.3, we obtain:

Lemma 2.5.4. *For every $v \in L^2(\Omega, H)$, it holds*

$$\varphi \in \partial j(v) \quad \Longleftrightarrow \quad \varphi \in L^2(\Omega, H^*) \quad \textit{and} \quad \varphi \in \partial k(v) \textit{ a.e. in } \Omega.$$

Proof. If we are given a $\varphi \in L^2(\Omega, H^*)$ with $\varphi \in \partial k(v)$ a.e. in Ω, then the properness of k, the finiteness of μ, the monotonicity of the integral and the inequality

$$k(u) - k(v) \geq \langle \varphi, u - v \rangle_H \quad \text{a.e. in } \Omega \quad \forall u \in L^2(\Omega, H)$$

immediately yield $(v, \varphi) \in \operatorname{graph}(\partial j)$. This proves "⇐". To obtain the reverse implication, we note that every $\varphi \in \partial j(v)$, i.e., every $\varphi \in L^2(\Omega, H^*)$ with

$$\int_\Omega k(u) \mathrm{d}\mu - \int_\Omega k(v) \mathrm{d}\mu \geq \langle \varphi, u - v \rangle_V = \int_\Omega \langle \varphi, u - v \rangle_H \, \mathrm{d}\mu \quad \forall u \in L^2(\Omega, H),$$

satisfies $v = S(\varphi + \iota(v))$, where $S : L^2(\Omega, H^*) \to L^2(\Omega, H)$ and $\iota : H \to H^*$ denote the solution operator to (2.41) and the Riesz isomorphism on H, respectively. The latter implies in combination with Lemma 2.5.3 that $v = S(\varphi + \iota(v)) = T(\varphi + \iota(v))$ holds a.e. in Ω for all $\varphi \in \partial j(v)$, i.e., we have

$$(v, \tilde{u} - v)_H + k(\tilde{u}) - k(v) \geq \langle \varphi + \iota(v), \tilde{u} - v \rangle_H \quad \forall \tilde{u} \in H \tag{2.43}$$

almost everywhere in Ω for all $\varphi \in \partial j(v)$. Note that (2.43) can be rewritten as $\varphi \in \partial k(v)$ a.e. in Ω. This establishes "⇒" and completes the proof. □

We can now state the main result of this section:

Theorem 2.5.5. *Let H, Ω, Σ, μ and k be as in Assumption 2.5.1 and suppose that k is twice epi-differentiable in all $\tilde{v} \in H$ for all $\tilde{\varphi} \in \partial k(\tilde{v})$. Then, the function*

$$j : L^2(\Omega, H) \to (-\infty, \infty], \qquad v \mapsto \int_\Omega k(v) \mathrm{d}\mu,$$

is well-defined, convex, lower semicontinuous, proper and twice epi-differentiable in all $v \in L^2(\Omega, H)$ for all $\varphi \in \partial j(v)$, and for every $v \in L^2(\Omega, H)$ and all $\varphi \in \partial j(v)$ it holds

$$\mathcal{K}_j^{red}(v, \varphi) = \left\{ z \in L^2(\Omega, H) \;\middle|\; Q_k^{v,\varphi}(z) \in L^1(\Omega, [0, \infty]) \right\} \tag{2.44}$$

and

$$Q_j^{v,\varphi}(z) = \int_\Omega Q_k^{v,\varphi}(z) \mathrm{d}\mu \qquad \forall z \in \mathcal{K}_j^{red}(v, \varphi). \tag{2.45}$$

Proof. Let S and T be defined as before. Then, (1.6) implies that there exists a constant $C > 0$ with

$$\|T(\tilde{f}_1) - T(\tilde{f}_2)\|_H \leq C\|\tilde{f}_1 - \tilde{f}_2\|_{H^*} \qquad \forall \tilde{f}_1, \tilde{f}_2 \in H^*$$

and we obtain from Lemma 2.5.3 that $S(f) = T(f)$ holds a.e. in Ω for all $f \in L^2(\Omega, H^*)$. If we combine these two properties, then we arrive at the estimate

$$\|S(f_1) - S(f_2)\|_H \leq C\|f_1 - f_2\|_{H^*} \text{ a.e. in } \Omega \quad \forall f_1, f_2 \in L^2(\Omega, H^*). \tag{2.46}$$

Note that the second-order epi-differentiability of the function k and Theorem 1.4.1 yield that the map $T : H^* \to H$ is directionally differentiable in every $\tilde{f} \in H^*$, i.e., we have

$$\lim_{t \searrow 0} \frac{T(\tilde{f} + t\tilde{g}) - T(\tilde{f})}{t} = T'(\tilde{f}; \tilde{g}) \in H \quad \forall \tilde{g} \in H^*.$$

The latter implies in tandem with (2.46), Lemma 2.5.3 and the dominated convergence theorem that the solution map $S : L^2(\Omega, H^*) \to L^2(\Omega, H)$ is directionally differentiable in all $f \in L^2(\Omega, H^*)$ in all directions $g \in L^2(\Omega, H^*)$ with $S'(f; g) = T'(f; g)$ a.e. in Ω. Consider now an arbitrary but fixed $v \in L^2(\Omega, H)$ and some $\varphi \in \partial j(v)$. Then, it holds $v = S(\varphi + \iota(v))$ (where ι again denotes the Riesz isomorphism on H, cf. the proof of Lemma 2.5.4), and we obtain from the directional differentiability of S and the equivalence in Theorem 1.4.1 that j is twice epi-differentiable in v for φ. Recall in this context that j is proper, lower semicontinuous and convex by Lemma 2.5.2. This proves the first assertion of the theorem.

To establish (2.44) and (2.45), we note that the identity $S'(\varphi + \iota(v); g) = T'(\varphi + \iota(v); g)$ a.e. in Ω and Proposition 1.3.5 yield

$$\langle g, \delta \rangle_H - \|\delta\|_H^2 = Q_k^{v,\varphi}(\delta) \text{ a.e. in } \Omega$$

for all $\delta = S'(\varphi + \iota(v); g)$, $g \in L^2(\Omega, H^*)$. This shows that $Q_k^{v,\varphi}(\delta)$ is an element of $L^1(\Omega, \mathbb{R})$ for all $\delta \in S'(\varphi + \iota(v); L^2(\Omega, H^*))$. By applying Proposition 1.3.5 once more (on both the V- and the H-level), we obtain further that

$$Q_j^{v,\varphi}(\delta) = \int_\Omega \langle g, \delta \rangle_H - \|\delta\|_H^2 \, \mathrm{d}\mu = \int_\Omega Q_k^{v,\varphi}(\delta) \mathrm{d}\mu \tag{2.47}$$

for all $\delta \in S'(\varphi + \iota(v); L^2(\Omega, H^*))$. Suppose now that a $z \in \mathcal{K}_j^{red}(v, \varphi)$ and a $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ are given, and let $\{z_n\} \subset L^2(\Omega, H)$ be a recovery sequence as in Definition 1.3.6, i.e.,

$$z_n \to z \quad \text{and} \quad \frac{2}{t_n}\left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle\right) \to Q_j^{v,\varphi}(z)$$

for $n \to \infty$. Then, after transition to a subsequence (unrelabeled), $\{z_n\}$ converges also pointwise a.e. in Ω to z, and we may use the lemma of Fatou to compute

$$\begin{aligned} \infty &> Q_j^{v,\varphi}(z) \\ &= \lim_{n\to\infty} \int_\Omega \frac{2}{t_n}\left(\frac{k(v + t_n z_n) - k(v)}{t_n} - \langle \varphi, z_n \rangle\right) \mathrm{d}\mu \\ &\geq \int_\Omega \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{k(v + t_n z_n) - k(v)}{t_n} - \langle \varphi, z_n \rangle\right) \mathrm{d}\mu. \end{aligned} \tag{2.48}$$

The above chain of inequalities can only be true if the limes inferior in the integrand on the right-hand side of (2.48) is finite almost everywhere. In particular, it has to hold $Q_k^{v,\varphi}(z) < \infty$ and $z \in \mathcal{K}_k^{red}(v, \varphi)$ a.e. in Ω (due to the pointwise convergence $z_n \to z$ a.e. in Ω).

Consider now for our arbitrary but fixed $z \in \mathcal{K}_j^{red}(v,\varphi)$ the solutions ζ_n, $n \in \mathbb{N}$, of the problems

$$\begin{aligned}&\zeta_n \in \mathcal{K}_j^{red}(v,\varphi),\\ &\left(1+\frac{n}{2}\right)(\zeta_n, u-\zeta_n)_V + \frac{1}{2}Q_j^{v,\varphi}(u) - \frac{1}{2}Q_j^{v,\varphi}(\zeta_n) \geq \left(1+\frac{n}{2}\right)(z, u-\zeta_n)_H \quad \forall u \in \mathcal{K}_j^{red}(v,\varphi).\end{aligned}$$

Then, it follows from $z \in \mathcal{K}_k^{red}(v,\varphi)$ a.e. in Ω, the identity $S'(\varphi+\iota(v);g) = T'(\varphi+\iota(v);g)$ a.e. in Ω for all $g \in L^2(\Omega, H^*)$ and the same arguments as in the proof of Proposition 1.3.15 that

$$\left(1+\frac{n}{2}\right)(\zeta_n, \tilde{u}-\zeta_n)_H + \frac{1}{2}Q_k^{v,\varphi}(\tilde{u}) - \frac{1}{2}Q_k^{v,\varphi}(\zeta_n) \geq \left(1+\frac{n}{2}\right)(z, \tilde{u}-\zeta_n)_H \quad \forall \tilde{u} \in \mathcal{K}_k^{red}(v,\varphi) \tag{2.49}$$

holds a.e. in Ω for all $n \in \mathbb{N}$ and that $\{\zeta_n\}$ satisfies

$$\begin{aligned}\zeta_n &\in S'(\varphi+\iota(v); L^2(\Omega, H^*)) \quad \forall n \in \mathbb{N},\\ \zeta_n &\to z \text{ in } L^2(\Omega,H) \quad \text{and} \quad Q_j^{v,\varphi}(\zeta_n) \nearrow Q_j^{v,\varphi}(z) \text{ in } \mathbb{R},\\ \zeta_n &\to z \text{ in } H \quad \text{and} \quad Q_k^{v,\varphi}(\zeta_n) \nearrow Q_k^{v,\varphi}(z) \text{ in } \mathbb{R} \quad \text{a.e. in } \Omega.\end{aligned}$$

Using the above properties, $Q_k^{v,\varphi}(\zeta_n) \in L^0(\Omega,[0,\infty])$, the fact that measurability is preserved under pointwise a.e.-convergence, the identity (2.47), Fatou's lemma and the dominated convergence theorem, we immediately obtain that $Q_k^{v,\varphi}(z) \in L^0(\Omega,[0,\infty])$ and that

$$Q_j^{v,\varphi}(z) = \lim_{n\to\infty} Q_j^{v,\varphi}(\zeta_n) = \lim_{n\to\infty} \int_\Omega Q_k^{v,\varphi}(\zeta_n)\mathrm{d}\mu = \int_\Omega \lim_{n\to\infty} Q_k^{v,\varphi}(\zeta_n)\mathrm{d}\mu = \int_\Omega Q_k^{v,\varphi}(z)\mathrm{d}\mu.$$

This establishes (2.45) and "$\subset$" in (2.44). To obtain "$\supset$" in (2.44), we note that (2.49) is also sensible for all $z \in L^2(\Omega,H)$ with $Q_k^{v,\varphi}(z) \in L^1(\Omega,[0,\infty])$. If we plug in such a z both on the right-hand side and as a test function almost everywhere and integrate, then we arrive at the estimate

$$\infty > \frac{1}{2}\int_\Omega Q_k^{v,\varphi}(z)\mathrm{d}\mu \geq \frac{1}{2}Q_j^{v,\varphi}(\zeta_n) + \left(1+\frac{n}{2}\right)\|z-\zeta_n\|_V^2.$$

The above yields $\zeta_n \to z$ in V and, because of the lower semicontinuity of the second subderivative, see Lemma 1.3.12, that $z \in \mathcal{K}_j^{red}(v,\varphi)$. This completes the proof. □

We again close the section with some concluding remarks:

Remark 2.5.6.

(i) For finite-dimensional spaces H, Theorem 2.5.5 has already been proved in [Do, 1992, Section 5]. In this paper, the author makes use of the notions of q-boundedness and normality to establish the formula (2.45)*, cf. [Joly and Thelin, 1976].*

(ii) The proof of Theorem 2.5.5 demonstrates that the sharpness of the differentiability criterion in Theorem 1.4.1 is not only of theoretical interest but also of practical importance. The equivalence between the conditions (I) and (II) in Theorem 1.4.1 allowed us to establish the second-order epi-differentiability of the functional j in (2.38) *by studying the differentiability properties of the solution operator to the simple auxiliary problem* (2.41) *and to completely avoid working with the unhandy Definitions 1.3.1 and 1.3.6. Note that, in this whole section, it was never necessary to construct a recovery sequence for j as in* (1.37) *or to work with the weak topology of the space $L^2(\Omega,H)$. If one does not take the detour via* (2.41) *and uses Definition 1.3.6, then this is different and the derivation of Theorem 2.5.5 becomes much more complicated. We would like to point out that EVIs that involve non-smooth terms of the form* (2.38) *are in general not as trivial as our model problem* (2.41) *suggests. See, e.g., the problem of static elastoplasticity in Section 4.3.2.*

3 Application to EVIs of the First Kind

We now turn our attention to the consequences that the results of the last two chapters have for the study of special instances of the problem (P). In what follows, we first focus on EVIs that involve functionals of the form $j = \chi_K$, where χ_K denotes the characteristic function of a closed, convex and non-empty set $K \subset V$. Such problems are also known as elliptic variational inequalities of the first kind (cf. the discussion in Section 1.2) and are considered frequently in the literature - especially in the context of metric projections. See, e.g., [Bonnans and Shapiro, 2000; Fitzpatrick and Phelps, 1982; Haraux, 1977; Holmes, 1973; Levy, 1999; Mignot, 1976; Noll, 1995; Rockafellar, 1990; Rockafellar and Wets, 1998; Shapiro, 1992, 1994b, 2016; Zarantonello, 1971]. Before we delve into our analysis, we give a quick overview of the structure and the contents of this chapter:

After some initial remarks in Section 3.1, we begin our investigation by proving a generalization of Zarantonello's lemma on the directional differentiability of metric projections in boundary points in Section 3.2. See Theorem 3.2.2 for the corresponding result. Section 3.3 is then concerned with the concepts of (extended) polyhedricity and second-order regularity employed in [Bonnans and Shapiro, 2000; Haraux, 1977; Mignot, 1976; Shapiro, 1992, 1994b, 2016]. Here, we will see that the classical results of Mignot/Haraux and Bonnans/Shapiro on the directional differentiability of metric projections follow straightforwardly from the general analysis of Chapter 1. In the subsequent Section 3.4, we illustrate by means of several examples that the notion of polyhedricity is very counterintuitive in infinite dimensions. The main result of this section, Theorem 3.4.7, demonstrates that the admissible set of the elastoplastic torsion problem with ∞-norm constraint is, contrary to popular belief, cf. [Hintermüller and Surowiec, 2011, Section 5.2], in general not polyhedric.

3.1 Generalities and Preliminaries

As already mentioned, the aim of this chapter is to apply the mathematical machinery developed so far to elliptic variational inequalities of the first kind, i.e., to problems of the form

$$w \in K, \qquad \langle A(w), v - w\rangle \geq \langle f, v - w\rangle \qquad \forall v \in K. \tag{Q}$$

Here and in what follows, the quantities in (Q) are assumed to satisfy the conditions in Assumption 1.2.1 (with $j = \chi_K$). For the convenience of the reader and the sake of readability, we briefly recall what this means:

Assumption 3.1.1 (Standing Assumptions and Notation for the Study of the Problem (Q))**.**

- *V is a Hilbert space with topological dual V^* and dual pairing $\langle \cdot, \cdot \rangle$.*
- *$f \in V^*$ is a given datum (the argument of the solution map).*
- *K is a convex, closed and non-empty subset of V.*
- *$A : K \to V^*$ is an operator with the following properties:*
  - *(i) A maps bounded subsets of K into bounded subsets of V^*.*
  - *(ii) A is strongly monotone on K, i.e., there exists a constant $c > 0$ such that*
  $$\langle A(v_1) - A(v_2), v_1 - v_2\rangle \geq c\|v_1 - v_2\|_V^2 \quad \forall v_1, v_2 \in K.$$
  - *(iii) A is Fréchet differentiable on K in the sense of Definition 1.1.2(iv).*

Note that, in the situation of Assumption 3.1.1, we trivially have the following (which we have, in fact, already used in the proof of Theorem 2.4.8):

Lemma 3.1.2.

(i) For every $v \in K$, it holds $\partial\chi_K(v) = \mathcal{N}_K(v)$.

(ii) For every $v \in K$, every $\varphi \in \mathcal{N}_K(v)$ and every $z \in V$, it holds

$$Q^{v,\varphi}_{\chi_K}(z) = \inf\left\{ \liminf_{n\to\infty} \left\langle \frac{-2\varphi}{t_n}, z_n \right\rangle \;\middle|\; \begin{array}{l} \{t_n\} \subset \mathbb{R}^+, \{z_n\} \subset V, \\ t_n \searrow 0,\ z_n \rightharpoonup z,\ v + t_n z_n \in K \end{array} \right\}.$$

Further, $\mathcal{K}^{red}_{\chi_K}(v,\varphi) \subset \mathcal{T}_K(v) \cap \ker(\varphi)$ and for every $z \in \mathcal{T}_K(v) \cap \ker(\varphi)$, it is true that

$$Q^{v,\varphi}_{\chi_K}(z) = \inf\left\{ \liminf_{n\to\infty} \langle -\varphi, r_n \rangle \;\middle|\; \begin{array}{l} \{t_n\} \subset \mathbb{R}^+, \{r_n\} \subset V, \\ t_n \searrow 0,\ t_n r_n \rightharpoonup 0,\ v + t_n z + \dfrac{1}{2} t_n^2 r_n \in K \end{array} \right\}.$$

Proof. The identity in (i) is obvious and (ii) follows immediately from Definition 1.3.1 and Lemma 1.3.4. □

We remark that Lemma 3.1.2(ii) implies that the weak second subderivative $Q^{v,\varphi}_{\chi_K}$ coincides with the "directional curvature functional" introduced in [Christof and Wachsmuth, 2017b]. Compare also with Section 6.2 in this context. To simplify the notation, in the remainder of this chapter, we frequently drop the letter χ and write

$$Q^{v,\varphi}_K(z) := Q^{v,\varphi}_{\chi_K}(z), \qquad \mathcal{K}^{red}_K(v,\varphi) := \mathcal{K}^{red}_{\chi_K}(v,\varphi),$$

for all $v \in K$, $\varphi \in \mathcal{N}_K(v)$ and $z \in V$. The solution map to (Q) (which is well-defined and globally Lipschitz by Theorem 1.2.2) is again denoted with $S : V^* \to V$, $f \mapsto w$.

3.2 Zarantonello's Lemma

In this section and the next, we derive several conditions that are sufficient for the second-order epi-differentiability of the function χ_K and, as a consequence, ensure the directional differentiability of the solution map $S : V^* \to V$ to (Q), cf. Theorem 1.4.1. We begin with the following observation:

Lemma 3.2.1. *The characteristic function $\chi_K : V \to \{0, \infty\}$ associated with K is twice epi-differentiable in all $v \in K$ for $\varphi = 0$, and it holds*

$$Q^{v,0}_K = \chi_{\mathcal{T}_K(v)} \quad \forall v \in K.$$

Proof. From $\varphi = 0$ and Lemma 3.1.2, it follows straightforwardly that $Q^{v,0}_K = \chi_{\mathcal{T}_K(v)}$ for all $v \in K$. Further, χ_K is trivially twice epi-differentiable in v for $\varphi = 0$ since, given a $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ and a $z \in \mathcal{K}^{red}_K(v,\varphi) = \mathcal{T}_K(v)$, we can always find a sequence $\{z_n\} \subset V$ with $z_n \to z$ and $v + t_n z_n \in K$ for all n, i.e., with

$$z_n \to z \quad \text{and} \quad \frac{2}{t_n}\left(\frac{\chi_K(v + t_n z_n) - \chi_K(v)}{t_n} - \langle \varphi, z_n \rangle \right) = 0 \to Q^{v,0}_K(z) = \chi_{\mathcal{T}_K(v)}(z) = 0.$$

This proves the claim. □

By combining Lemma 3.2.1 and Theorem 1.4.1, we obtain:

Theorem 3.2.2 (Non-Linear Version of Zarantonello's Lemma)**.** *Suppose that Assumption 3.1.1 is satisfied and that a right-hand side $f \in V^*$ with $f \in A(K)$ is given. Then, the solution $w := S(f)$ to* (Q) *is precisely $w = A^{-1}(f)$ and the solution map $S : V^* \to V$ associated with* (Q) *is Hadamard directionally differentiable in f in all directions $g \in V^*$. Moreover, the directional derivative $\delta := S'(f; g)$ in f in a direction $g \in V^*$ is uniquely characterized by the variational inequality*

$$\delta \in \mathcal{T}_K(w), \quad \big\langle A'(w)\delta, z - \delta \big\rangle \geq \langle g, z - \delta \rangle \quad \forall z \in \mathcal{T}_K(w).$$

Proof. Since $f \in A(K)$, there exists a unique $w \in K$ with $\varphi := f - A(w) = 0$. This w is obviously the unique solution $S(f)$ to (Q). The claim now follows immediately from Lemma 3.2.1 and Theorem 1.4.1. □

Note that, if we identify V^* with V and choose A to be the identity, then Theorem 3.2.2 expresses that the metric projection

$$P_K : V \to V, \qquad f \mapsto \operatorname*{argmin}_{v \in K} \frac{1}{2}\|v - f\|_V^2,$$

is Hadamard directionally differentiable in all points $f \in K$. This demonstrates that Theorem 3.2.2 is indeed a generalization of Zarantonello's lemma, cf. [Zarantonello, 1971, Lemma 4.6]. We would like to point out that the directional differentiability of the function P_K in all $f \in K$ is not as trivial as it seems since the perturbed points $f + tg$, $t > 0$, appearing in the difference quotients (1.18) do not necessarily have to be elements of K. In particular we do not differentiate the identity map here.

Unfortunately, even in finite dimensions, the metric projection $P_K : V \to V$ onto a closed, convex and non-empty set K does not necessarily have to be directionally differentiable in all points $f \in V \setminus K$. To see this, we define for arbitrary but fixed numbers $R_1, R_2 \in \mathbb{R}^+$ with $R_1 < R_2$ the functions

$$\psi_m : [-R_m, R_m] \to \mathbb{R}, \quad x \mapsto R_m + 1 - \sqrt{R_m^2 - x^2}, \qquad m = 1, 2,$$

and consider the sequences $\{x_n\}$, $\{y_n\}$ that are generated by the following algorithm:

(i) $x_0 := R_2$

(ii) For $n = 1, 2, 3, ...$

Choose $x_n \in (0, x_{n-1})$ such that the line connecting the points $(x_{n-1}, \psi_2(x_{n-1}))$ and $(x_n, \psi_2(x_n))$ is a tangent to the graph of the function ψ_1. Define $y_n \in (0, R_1)$ to be the first component of the point where this tangent intersects $\operatorname{graph}(\psi_1)$.

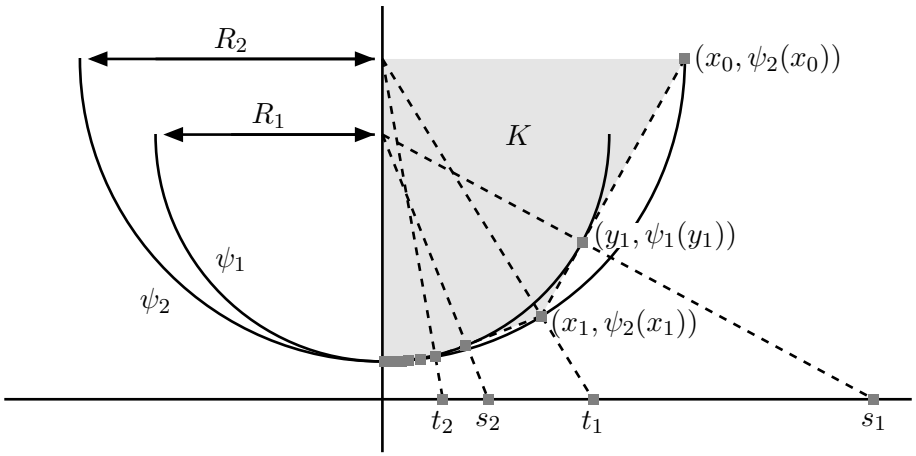

Figure 3.1: Construction of the sequences $\{x_n\}$, $\{y_n\}$, $\{s_n\}$, $\{t_n\}$ and the set K in (3.1).

Note that the above recurrence relation indeed produces infinitely many x_n and y_n (since the tangent at the point $(y_n, \psi_1(y_n))$ can never reach the slope zero) and that the iterates x_n and y_n are uniquely determined due to the strict convexity of the functions ψ_1 and ψ_2. From our construction, we obtain further that

$$0 < x_n < y_n < x_{n-1} \qquad \text{and} \qquad \frac{\psi_2(x_n) - \psi_2(x_{n-1})}{x_n - x_{n-1}} = \psi_1'(y_n)$$

holds for all $n \in \mathbb{N}$. This implies that the sequences $\{x_n\}$ and $\{y_n\}$ converge to an $\tilde{x} \in [0, R_1)$ which satisfies $\psi_1'(\tilde{x}) = \psi_2'(\tilde{x})$. Since $\tilde{x} = 0$ is the only point with the latter property, it follows that $x_n \searrow 0$ and $y_n \searrow 0$ for $n \to \infty$. Consider now the set

$$K := \operatorname{cl}\Big(\operatorname{conv}\big(\big\{(x_n, \psi_2(x_n))\big\}_{n=0}^{\infty} \cup \big\{(0, R_2+1)\big\}\big)\Big), \tag{3.1}$$

and define

$$t_n := \frac{(R_2+1)x_n}{R_2 + 1 - \psi_2(x_n)}, \qquad s_n := \frac{(R_1+1)y_n}{R_1 + 1 - \psi_1(y_n)} \qquad \forall n \in \mathbb{N}.$$

Then, K is convex, closed and non-empty, it holds $t_n \searrow 0$ and $s_n \searrow 0$, and $(t_n, 0)$ and $(s_n, 0)$ are precisely those points where the lines $\operatorname{aff}\{(x_n, \psi_2(x_n)), (0, R_2+1)\}$ and $\operatorname{aff}\{(y_n, \psi_1(y_n)), (0, R_1+1)\}$ intersect the first coordinate axis, see Figure 3.1. From the latter, we obtain that the metric projection P_K onto K satisfies $P_K(t_n, 0) = (x_n, \psi_2(x_n))$ and $P_K(s_n, 0) = (y_n, \psi_1(y_n))$ for all $n \in \mathbb{N}$ and that

$$\begin{aligned} \frac{P_K(t_n, 0) - P_K(0,0)}{t_n} &= \frac{(x_n, \psi_2(x_n)) - (0,1)}{t_n} \\ &= \left(\frac{R_2 + 1 - \psi_2(x_n)}{R_2+1}, \frac{R_2 + 1 - \psi_2(x_n)}{R_2+1}\frac{\psi_2(x_n) - 1}{x_n}\right) \\ &\to \left(\frac{R_2}{R_2+1}, 0\right) \end{aligned} \tag{3.2}$$

and

$$\frac{P_K(s_n, 0) - P_K(0,0)}{s_n} = \frac{(y_n, \psi_1(y_n)) - (0,1)}{s_n} \to \left(\frac{R_1}{R_1+1}, 0\right) \neq \left(\frac{R_2}{R_2+1}, 0\right) \tag{3.3}$$

holds for $n \to \infty$. This shows that P_K is not differentiable in $(0,0)$ in the direction $(1,0)$. Note that we could also have argued with part (iii) of Proposition 1.3.5 here since

$$\frac{2}{t_n}\left(\frac{\chi_K(x_n, \psi_2(x_n)) - \chi_K(0,1)}{t_n} - \left\langle (0,-1), \frac{(x_n, \psi_2(x_n)) - (0,1)}{t_n}\right\rangle\right) \to \frac{R_2}{(R_2+1)^2} \tag{3.4}$$

and

$$\frac{2}{s_n}\left(\frac{\chi_K(y_n, \psi_1(y_n)) - \chi_K(0,1)}{s_n} - \left\langle (0,-1), \frac{(y_n, \psi_1(y_n)) - (0,1)}{s_n}\right\rangle\right) \to \frac{R_1}{(R_1+1)^2}, \tag{3.5}$$

and that the non-differentiability of P_K in $(0,0)$ in the direction $(1,0)$ implies that χ_K is an example of a convex, lower semicontinuous and proper function that is not twice epi-differentiable everywhere.

The above construction, that essentially goes back to [Shapiro, 1994a], demonstrates that the metric projection onto a closed, convex and non-empty set K can only be expected to be directionally differentiable if the set K under consideration admits a well-defined notion of curvature. The reason for the behavior in (3.2) and (3.3) is, after all, that the set K in (3.1) behaves like a circle of radius R_2 along the sequence $(x_n, \psi_2(x_n)) \to (0,1)$ and like a circle of radius R_1 along the sequence $(y_n, \psi_1(y_n)) \to (0,1)$ and that, as a consequence, it is not quantifiable "how much" curvature K possesses at $(0,1)$, cf. (3.4) and (3.5). In the literature, several different concepts have been proposed to sensibly define what (non-) curvedness means for a subset of a Hilbert space. The most prominent are probably those of:

3.3 Polyhedricity and Second-Order Regularity

In this section, we study the relationship between the notions of second-order epi-differentiability, (extended) polyhedricity and second-order regularity. We begin by recalling several definitions that are needed for our analysis, cf. [Mignot, 1976, Section 2], [Haraux, 1977], [Wachsmuth, 2016, Section 3] and [Bonnans and Shapiro, 2000, Section 3.2.1, Definitions 3.51, 3.85].

Definition 3.3.1. *Suppose that L is a closed, convex, non-empty subset of a Banach space X.*

(i) The (inner) second-order tangent set to a tuple $(x, z) \in L \times \mathcal{T}_L(x)$ is defined by

$$\mathcal{T}_L^2(x,z) := \left\{ r \in X \;\middle|\; \operatorname{dist}\left(x + t\,z + \frac{1}{2}\,t^2\,r, L \right) = o(t^2) \text{ as } t \searrow 0 \right\}.$$

(ii) The set L is said to be polyhedric at $x \in L$ for $\varphi \in \mathcal{N}_L(x)$ if

$$\mathcal{T}_L(x) \cap \ker(\varphi) = \operatorname{cl}\big(\mathcal{T}_L^{rad}(x) \cap \ker(\varphi)\big).$$

(iii) The set L is said to be extended polyhedric at $x \in L$ for $\varphi \in \mathcal{N}_L(x)$ if

$$\mathcal{T}_L(x) \cap \ker(\varphi) = \operatorname{cl}\big(\big\{ z \in \mathcal{T}_L(x) \;\big|\; 0 \in \mathcal{T}_L^2(x,z) \big\} \cap \ker(\varphi)\big).$$

(iv) If L is (extended) polyhedric at $x \in L$ for all $\varphi \in \mathcal{N}_L(x)$, then L is called (extended) polyhedric at x. If L is (extended) polyhedric at all $x \in L$, then L is said to be (extended) polyhedric.

(v) The set L is called second-order regular at a point $x \in L$ if for all $z \in \mathcal{T}_L(x)$ and all $x_n \in L$ of the form $x_n := x + t_n z + \frac{1}{2} t_n^2 r_n$ with $t_n \searrow 0$ and $t_n r_n \to 0$ it is true that

$$\lim_{n\to\infty} \operatorname{dist}\big(r_n, \mathcal{T}_L^2(x,z)\big) = 0.$$

(vi) The set L is called strongly second-order regular at a point $x \in L$ if for all $z \in \mathcal{T}_L(x)$ and all $x_n \in L$ of the form $x_n := x + t_n z + \frac{1}{2} t_n^2 r_n$ with $t_n \searrow 0$ and $t_n r_n \rightharpoonup 0$ it is true that

$$\lim_{n\to\infty} \operatorname{dist}\big(r_n, \mathcal{T}_L^2(x,z)\big) = 0.$$

Let us give some remarks on the concepts in Definition 3.3.1:

Remark 3.3.2.

(i) It is easy to prove (see [Wachsmuth, 2016, Lemma 4.1]) that a set L is polyhedric in a point $x \in L$ if and only if

$$\mathcal{T}_L(x) \cap \ker(\varphi) = \operatorname{cl}\big(\mathcal{T}_L^{rad}(x) \cap \ker(\varphi)\big) \quad \forall \varphi \in X^*.$$

(ii) From the definition of the radial cone, it follows straightforwardly that $0 \in \mathcal{T}_L^2(x,z)$ for all $x \in L$ and all $z \in \mathcal{T}_L^{rad}(x)$. This shows that polyhedricity in a point $x \in L$ for some $\varphi \in \mathcal{N}_L(x)$ implies extended polyhedricity in $x \in L$ for the same $\varphi \in \mathcal{N}_L(x)$. We point out that the reverse implication is in general not true. The set

$$\{0\} \cup \operatorname{conv}\left\{ \left(\frac{1}{n}, \frac{1}{n^4} \right) \in \mathbb{R}^2 \;\middle|\; n \in \mathbb{Z} \right\} \subset \mathbb{R}^2,$$

for example, is extended polyhedric everywhere but not polyhedric.

(iii) Note that strong second-order regularity implies second-order regularity, and that these concepts are the same in finite-dimensional spaces.

(iv) Remarkably, (extended) polyhedricity does not imply second-order regularity. The set

$$L := \{v \in L^2(0,1) \mid v \geq 0 \text{ a.e.}\},$$

for example, is a polyhedric subset of $L^2(0,1)$ but not second-order regular at, e.g., $v \equiv 1$. We refer to [Christof and Wachsmuth, 2017b, Example 5.6] for a proof of this statement.

What the concepts in Definition 3.3.1 have in common is that, at one point or another, all of them have been proposed as sufficient criteria for the directional differentiability of metric projections, cf. [Bonnans et al., 1998; Bonnans and Shapiro, 2000; Haraux, 1977; Mignot, 1976]. In what follows, we will demonstrate that the differentiability results in the latter papers can be reproduced straightforwardly by applying the theory of Chapter 1. We begin by proving:

Proposition 3.3.3 (Second-Order Epi-Differentiability in the Case of Extended Polyhedricity)**.** *Assume that K and V satisfy the conditions in Assumption 3.1.1. Suppose that a $v \in K$ and a $\varphi \in \mathcal{N}_K(v)$ are given such that K is extended polyhedric at v for φ. Then, χ_K is twice epi-differentiable in v for φ and it holds*

$$\mathcal{K}_K^{red}(v,\varphi) = \mathcal{T}_K(v) \cap \ker(\varphi) \quad \textit{and} \quad Q_K^{v,\varphi}(z) = 0 \quad \forall z \in \mathcal{T}_K(v) \cap \ker(\varphi). \tag{3.6}$$

Proof. We use Lemma 1.3.13 to prove the claim: Define $\mathcal{Z} := \left\{z \in \mathcal{T}_K(v) \mid 0 \in \mathcal{T}_K^2(v,z)\right\} \cap \ker(\varphi)$, $\mathcal{K} := \mathcal{T}_K(v) \cap \ker(\varphi)$ and $Q : \mathcal{K} \to [0,\infty)$, $Q(z) := 0$ for all $z \in \mathcal{K}$. Then, the definition of the second-order tangent set implies that for every $z \in \mathcal{Z}$ and every $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ we can find a sequence $\{r_n\} \subset V$ with $v + t_n z + \frac{1}{2}t_n^2 r_n \in K$ and $r_n \to 0$. The latter yields

$$\begin{aligned} 0 &\leq Q_K^{v,\varphi}(z) \\ &\leq \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{\chi_K(v + t_n z + \frac{1}{2}t_n^2 r_n) - \chi_K(v)}{t_n} - \left\langle \varphi, z + \frac{1}{2}t_n r_n \right\rangle\right) = \lim_{n\to\infty} \langle -\varphi, r_n\rangle = 0, \end{aligned}$$

i.e., it holds $\mathcal{Z} \subset \mathcal{K}_K^{red}(v,\varphi)$ and for every $z \in \mathcal{Z}$ and every $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ there exists a recovery sequence $\{z_n\}$ as in (1.43). From Lemma 3.1.2 and the definition of extended polyhedricity, we obtain further that $\mathcal{K}_K^{red}(v,\varphi) \subset \mathcal{K}$ and that $\mathcal{Z}$ is dense in $\mathcal{K}$. If we combine all of the above, then we arrive precisely at the situation of Lemma 1.3.13 and the claim follows immediately. □

Note that the proof of Proposition 3.3.3 shows that Lemma 1.3.13 indeed generalizes the idea behind the concept of (extended) polyhedricity. In Chapter 4, we will see that it also covers situations where curvature is present (cf. the derivation of Theorem 4.3.16 and Remark 4.3.17). For strongly second-order regular sets, we obtain:

Proposition 3.3.4 (Second-Order Epi-Differentiability in the Case of Second-Order Regularity)**.** *Assume that K and V satisfy the conditions in Assumption 3.1.1. Suppose that a $v \in K$ and a $\varphi \in \mathcal{N}_K(v)$ are given such that K is strongly second-order regular at v. Then, χ_K is twice epi-differentiable in v for φ and it holds*

$$\mathcal{K}_K^{red}(v,\varphi) = \mathcal{T}_K(v) \cap \ker(\varphi) \quad \textit{and} \quad Q_K^{v,\varphi}(z) = \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s\rangle \quad \forall z \in \mathcal{T}_K(v) \cap \ker(\varphi). \tag{3.7}$$

Proof. Consider an arbitrary but fixed $z \in \mathcal{T}_K(v) \cap \ker(\varphi)$ and assume that $\{t_n\} \subset \mathbb{R}^+$ and $\{r_n\} \subset V$ are sequences with

$$t_n \searrow 0, \quad t_n r_n \rightharpoonup 0, \quad v + t_n z + \frac{1}{2}t_n^2 r_n \in K.$$

Then, the strong second-order regularity of K in v implies that there exists a sequence $\{s_n\} \subset \mathcal{T}_K^2(v,z)$ with $\|s_n - r_n\|_V \to 0$ for $n \to \infty$, and we may compute

$$\liminf_{n\to\infty} \langle -\varphi, r_n\rangle = \liminf_{n\to\infty} \langle -\varphi, s_n\rangle \geq \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s\rangle .$$

Taking the infimum over all $\{r_n\}$, $\{t_n\}$ and using Lemma 3.1.2, we now obtain

$$Q_K^{v,\varphi}(z) \geq \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s \rangle .$$

If, conversely, an $s \in \mathcal{T}_K^2(v,z)$ is given and $\{t_n\} \subset \mathbb{R}^+$ is an arbitrary but fixed sequence with $t_n \searrow 0$, then the definition of the second-order tangent set implies that there exists a sequence $\{r_n\} \subset V$ with $v + t_n z + \frac{1}{2} t_n^2 r_n \in K$ for all n and $r_n \to s$ for $n \to \infty$. The latter yields

$$Q_K^{v,\varphi}(z) \leq \liminf_{n \to \infty} \langle -\varphi, r_n \rangle = \langle -\varphi, s \rangle .$$

Taking the infimum over all $s \in \mathcal{T}_K^2(v,z)$, we now infer

$$Q_K^{v,\varphi}(z) = \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s \rangle .$$

It remains to prove the second-order epi-differentiability of χ_K in v for φ. To this end, we assume that some sequence $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$ is given and choose a sequence $\{s_n\} \subset \mathcal{T}_K^2(v,z)$ with

$$\langle -\varphi, s_n \rangle \to \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s \rangle$$

for $n \to \infty$. Since $\{s_n\} \subset \mathcal{T}_K^2(v,z)$, for every $n \in \mathbb{N}$, we can find a sequence $\{r_{n,m}\}$ with $r_{n,m} \to s_n$ for $m \to \infty$ and $v + t_m z + \frac{1}{2} t_m^2 r_{n,m} \in K$ for all m. Choose a strictly increasing subsequence $\{m_n\}$ such that $\|s_n - r_{n,m_n}\|_V \leq \frac{1}{n}$ and $t_{m_n} \|s_n\|_V \leq \frac{1}{n}$ for all n. Then, it holds

$$t_{m_n} \searrow 0, \quad t_{m_n} r_{n,m_n} \to 0, \quad v + t_{m_n} z + \frac{1}{2} t_{m_n}^2 r_{n,m_n} \in K,$$
$$\langle -\varphi, r_{n,m_n} \rangle \to \inf_{s \in \mathcal{T}_K^2(v,z)} \langle -\varphi, s \rangle .$$

This shows that, given an arbitrary $\{t_m\} \subset \mathbb{R}^+$ with $t_m \searrow 0$, we can always find a subsequence $\{t_{m_n}\}$ such that there exist $r_{m_n} \in V$ with

$$t_{m_n} r_{m_n} \to 0, \quad v + t_{m_n} z + \frac{1}{2} t_{m_n}^2 r_{m_n} \in K, \quad \langle -\varphi, r_{m_n} \rangle \to Q_K^{v,\varphi}(z).$$

Using the same argument as in the proof of Lemma 1.3.13, the second-order epi-differentiability of χ_K in v for φ now follows immediately. This proves the claim. □

Using the last two results and Theorem 1.4.1, we can prove:

Theorem 3.3.5. *Suppose that V, K and A satisfy the conditions in Assumption 3.1.1. Assume that $K = F^{-1}(L)$ holds for some twice continuously Fréchet differentiable map $F : V \to U$ from V into some Hilbert space U and some closed, convex, non-empty set $L \subset U$. Let $w := S(f)$ be the unique solution to* (Q) *for some arbitrary but fixed right-hand side $f \in V^*$ and denote with φ the residuum $f - A(w) \in \mathcal{N}_K(w)$. Suppose that there exists a Lagrange multiplier $\lambda \in \mathcal{N}_L(F(w))$, i.e., $\varphi = F'(w)^* \lambda$, such that L is extended polyhedric at $F(w)$ for λ and such that*

$$F'(w)V - \mathbb{R}^+\Big(L - F(w)\Big) \cap \ker(\lambda) = U \tag{3.8}$$

holds. Assume further that the map $z \mapsto \langle \lambda, F''(w) z^2 \rangle$ is weakly lower semicontinuous. Then, the solution operator $S : V^* \to V$ associated with (Q) *is Hadamard directionally differentiable in f in all directions $g \in V^*$. Moreover, the directional derivative $\delta := S'(f; g)$ in f in a direction $g \in V^*$ is uniquely characterized by the following EVI of the first kind:*

$$\begin{aligned} &\delta \in F'(w)^{-1} \left(\mathcal{T}_L(F(w)) \cap \ker(\lambda) \right), \\ &\big\langle A'(w)\delta, z - \delta \big\rangle + \big\langle \lambda, F''(w)(\delta, z - \delta) \big\rangle \geq \langle g, z - \delta \rangle \quad \forall z \in F'(w)^{-1} \left(\mathcal{T}_L(F(w)) \cap \ker(\lambda) \right). \end{aligned} \tag{3.9}$$

Proof. From Proposition 3.3.3, we obtain that χ_L is twice epi-differentiable in $F(w)$ for λ with

$$Q_L^{F(w),\lambda} = \chi_{\mathcal{T}_L(F(w))\cap\ker(\lambda)}.$$

Further, $K = F^{-1}(L)$ yields $\chi_K = \chi_L \circ F$ and (3.8) is equivalent to

$$F'(w)V - \mathbb{R}^+\Big\{u - F(w) \;\Big|\; u \in U,\ \chi_L(u) - \chi_L(F(w)) = \langle \lambda, u - F(w)\rangle\Big\} = U$$

which is precisely (2.26) for the characteristic function χ_L. As a consequence, we may employ the chain rule in Theorem 2.4.8 to deduce that χ_K is twice epi-differentiable in w for φ with

$$\begin{aligned} &Q_K^{w,\varphi}(z) \\ &\quad = Q_L^{F(w),\lambda}(F'(w)z) + \left\langle \lambda, F''(w)z^2\right\rangle \\ &\quad = \chi_{\mathcal{T}_L(F(w))\cap\ker(\lambda)}(F'(w)z) + \left\langle \lambda, F''(w)z^2\right\rangle \\ &\quad = \chi_{F'(w)^{-1}(\mathcal{T}_L(F(w))\cap\ker(\lambda))}(z) + \left\langle \lambda, F''(w)z^2\right\rangle \quad \forall z \in V. \end{aligned}$$

Using Theorem 1.4.1, it now follows immediately that S is Hadamard directionally differentiable in f in all directions $g \in V^*$, and that the directional derivative $\delta := S'(f;g)$ in f in a direction $g \in V^*$ is uniquely characterized by the EVI

$$\begin{aligned} &\delta \in F'(w)^{-1}\left(\mathcal{T}_L(F(w)) \cap \ker(\lambda)\right), \\ &\left\langle A'(w)\delta, z - \delta\right\rangle + \frac{1}{2}\left\langle \lambda, F''(w)z^2\right\rangle - \frac{1}{2}\left\langle \lambda, F''(w)\delta^2\right\rangle \geq \langle g, z - \delta\rangle \\ &\qquad\qquad \forall z \in F'(w)^{-1}\left(\mathcal{T}_L(F(w)) \cap \ker(\lambda)\right). \end{aligned}$$

Since the admissible set of the above variational inequality is convex, we may rewrite it to obtain (3.9). This completes the proof. □

Theorem 3.3.6. *Suppose that V, K and A satisfy the conditions in Assumption 3.1.1. Assume that $K = F^{-1}(L)$ holds for some twice continuously Fréchet differentiable function $F : V \to U$ from V into some Hilbert space U and some closed, convex, non-empty set $L \subset U$. Let $w := S(f)$ be the unique solution to* (Q) *for some arbitrary but fixed right-hand side $f \in V^*$ and denote with φ the residuum $f - A(w) \in \mathcal{N}_K(w)$. Suppose that L is strongly second-order regular in $F(w)$ and that there exists a Lagrange multiplier $\lambda \in \mathcal{N}_L(F(w))$, i.e., $\varphi = F'(w)^*\lambda$, such that the map $z \mapsto \left\langle \lambda, F''(w)z^2\right\rangle$ is weakly lower semicontinuous and such that*

$$F'(w)V - \mathbb{R}^+\Big(L - F(w)\Big) \cap \ker(\lambda) = U.$$

Then, the solution operator $S : V^* \to V$ associated with (Q) *is Hadamard directionally differentiable in f in all directions $g \in V^*$. Moreover, the directional derivative $\delta := S'(f;g)$ in f in a direction $g \in V^*$ is uniquely characterized by the following EVI of the second kind:*

$$\begin{aligned} &\delta \in F'(w)^{-1}\left(\mathcal{T}_L(F(w)) \cap \ker(\lambda)\right), \\ &\left\langle A'(w)\delta, z - \delta\right\rangle + \frac{1}{2}\left(\inf_{s\in\mathcal{T}_L^2(F(w),F'(w)z)} \langle -\lambda, s\rangle + \left\langle \lambda, F''(w)z^2\right\rangle\right) \\ &\qquad - \frac{1}{2}\left(\inf_{s\in\mathcal{T}_L^2(F(w),F'(w)\delta)} \langle -\lambda, s\rangle + \left\langle \lambda, F''(w)\delta^2\right\rangle\right) \geq \langle g, z - \delta\rangle \\ &\qquad\qquad \forall z \in F'(w)^{-1}\left(\mathcal{T}_L(F(w)) \cap \ker(\lambda)\right). \end{aligned} \tag{3.10}$$

Proof. Completely analogous to that of Theorem 3.3.5. □

We close this section with some remarks on Theorems 3.3.5 and 3.3.6.

Remark 3.3.7.

(i) The case $F = \mathrm{Id}$ and $U = V$ is explicitly allowed in Theorems 3.3.5 and 3.3.6. In this situation, the Zowe/Kurcyusz-type condition (3.8) *is trivially satisfied and the EVIs for the directional derivatives $S'(f;g)$ involve exactly the functionals in* (3.6) *and* (3.7)*.*

(ii) Theorem 3.3.5 generalizes, e.g., [Mignot, 1976, Théorème 2.1] and [Haraux, 1977, Theorem 1], and Theorem 3.3.6 extends, e.g., [Shapiro, 2016, Proposition 2.1, Theorem 3.1] and [Bonnans et al., 1998, Theorem 7.2]. See also [Bonnans and Shapiro, 2000; Shapiro, 1994b] for related results. We would like to point out that a variant of Theorem 3.3.6 still holds when (3.8) *is replaced with the classical Zowe-Kurcyusz constraint qualification* (2.16)*. In this case, however, the second subderivative $Q_K^{v,\varphi}$ cannot be split into a $\mathcal{T}_L^2$- and an F''-contribution as in* (3.10) *but is given by a more implicit formula, cf. [Christof and Wachsmuth, 2017b, Lemmas 5.7 and 5.8].*

(iii) If the operator $A : K \to V^*$ admits a twice Fréchet differentiable potential $k : V \to \mathbb{R}$ as in (1.4)*, then both* (3.9) *and* (3.10) *can be rewritten as*

$$\min_{z \in F'(w)^{-1}(\mathcal{T}_L(F(w)) \cap \ker(\lambda))} \frac{1}{2}\left(\partial_v^2 \mathcal{L}(w,\lambda)z^2 + Q_L^{F(w),\lambda}(F'(w)z)\right) - \langle g, z\rangle\,.$$

Here, $\partial_v^2 \mathcal{L}(w,\lambda)$ denotes the second partial Fréchet derivative of the Lagrange function

$$\mathcal{L} : V \times U^* \to \mathbb{R}, \qquad \mathcal{L}(v,\eta) := k(v) + \langle \eta, F(v)\rangle\,,$$

associated with the objective k and the constraint $v \in F^{-1}(L)$ at (w,λ).

3.4 Examples and Warning Counterexamples

As the results in Section 3.3 show, the concepts of (extended) polyhedricity and second-order regularity both provide a notion of curvature in Hilbert space that is suitable for the differential sensitivity analysis of metric projections and more general variational inequalities. However, a word of warning is in order here. In infinite dimensions, the properties in Definition 3.3.1 turn out to be very treacherous and often completely defy intuition. We have already seen this in the case of the set $\{v \in L^2(0,1) \mid v \geq 0 \text{ a.e.}\}$ which is polyhedric (i.e., in a sense, flat) but not second-order regular (a property that one would associate with the well-definedness of boundary curvature). In what follows, we collect several examples that highlight the peculiar effects that can occur when the curvature of sets in infinite-dimensional spaces is studied. Since the notion of second-order regularity is mainly suited for finite-dimensional problems and scalar constraints and with view on the analysis in Chapters 4 and 5, we focus primarily on the concept of polyhedricity. For tangible examples of second-order regular sets, we refer to [Bonnans et al., 1998, Section 7.1] and [Bonnans and Shapiro, 2000, Chapters 3 and 4].

3.4.1 Sets with Upper and Lower Bounds in Dirichlet Spaces

Let us begin with a positive example: The result that is most commonly used in applications to verify the condition of polyhedricity in infinite-dimensional spaces is Mignot's theorem on the polyhedricity of sets with upper and lower bounds in Dirichlet spaces (cf., e.g., [Betz, 2015; Jarušek et al., 2003; Müller and Schiela, 2017]). To formulate this theorem, we need:

Definition 3.4.1 (Dirichlet Space)**.** *Suppose that Ω is a locally compact and separable metric space. Denote with $\mathfrak{B}(\Omega)$ the Borel σ-algebra of Ω and assume that $\mu : \mathfrak{B}(\Omega) \to [0, \infty]$ is a measure such that $\mu(O) > 0$ holds for all non-empty open sets $O \subset \Omega$ and such that $\mu(E) < \infty$ holds for all compact sets $E \subset \Omega$. Then, a regular Dirichlet space is a subspace V of $L^2(\Omega, \mu)$ equipped with a symmetric, positive semidefinite, bilinear form $\mathcal{E} : V \times V \to \mathbb{R}$ such that the following conditions are satisfied:*

(i) V is dense in $L^2(\Omega, \mu)$ w.r.t. the norm $\| \cdot \|_{L^2}$,

(ii) V is a Hilbert space when endowed with the product $(\cdot, \cdot)_V := \mathcal{E}(\cdot, \cdot) + (\cdot, \cdot)_{L^2}$,

(iii) for all $v \in V$, it holds $u := \min(1, \max(0, v)) \in V$ with $\mathcal{E}(u, u) \leq \mathcal{E}(v, v)$,

(iv) $V \cap C_c(\Omega)$ is dense in V w.r.t. the norm $\| \cdot \|_V$,

(v) $V \cap C_c(\Omega)$ is dense in $C_c(\Omega)$ w.r.t. the norm $\| \cdot \|_{L^\infty}$.

Here,

$$C_c(\Omega) := \{v \in C(\Omega) \mid v \text{ has compact support in } \Omega\}.$$

For details on regular Dirichlet spaces and their properties, we refer to [Beurling and Deny, 1959] and [Fukushima et al., 2011]. Mignot's theorem now reads as follows:

Theorem 3.4.2 ([Mignot, 1976, Théorème 3.2])**.** *Assume that $(V, \mathcal{E})$ is a regular Dirichlet space and suppose that $\psi_1, \psi_2 : \Omega \to [-\infty, \infty]$ are Borel measurable functions such that the set*

$$K := \{v \in V \mid \psi_1 \leq v \leq \psi_2 \ \mu\text{-a.e. in } \Omega\} \tag{3.11}$$

is non-empty. Then, K is polyhedric everywhere.

We point out that the above result can also be proved in the more general setting of vector lattices, see [Wachsmuth, 2016, Theorem 4.18].

In practice, the main value of Theorem 3.4.2 lies in its applicability to sets with upper and lower bounds in H^1- and $H^{1/2}$-spaces. Such sets appear frequently in fields like contact mechanics and play a major role in many classical problems. As direct consequences of Theorems 3.3.5 and 3.4.2, we obtain, for example, the following two corollaries:

Corollary 3.4.3 (Directional Differentiability for the Classical Obstacle Problem)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a non-empty, open set (endowed with the Lebesgue measure $\mathcal{L}^d$ and the Euclidean topology) and let $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ be defined as in [Attouch et al., 2006, Section 5.2]. Assume that two Borel measurable functions $\psi_1, \psi_2 : \Omega \to [-\infty, \infty]$ are given such that*

$$K := \left\{ v \in H_0^1(\Omega) \;\middle|\; \psi_1 \leq v \leq \psi_2 \ \mathcal{L}^d\text{-a.e. in } \Omega \right\} \neq \emptyset,$$

and suppose that $A : K \to H^{-1}(\Omega)$ is an operator which satisfies the conditions in Assumption 3.1.1. Then, the solution map $S : H^{-1}(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, to the classical obstacle problem

$$w \in K, \qquad \langle A(w), v - w \rangle \geq \langle f, v - w \rangle \qquad \forall v \in K \tag{3.12}$$

is Hadamard directionally differentiable in all $f \in H^{-1}(\Omega)$, and the directional derivative $\delta := S'(f; g)$ in an $f \in H^{-1}(\Omega)$ in a direction $g \in H^{-1}(\Omega)$ is uniquely characterized by the variational inequality

$$\delta \in \mathcal{T}_K(w) \cap \ker(f - A(w)), \qquad \left\langle A'(w)\delta, z - \delta \right\rangle \geq \langle g, z - \delta \rangle \qquad \forall z \in \mathcal{T}_K(w) \cap \ker(f - A(w)). \tag{3.13}$$

Proof. From Stampacchia's lemma, see [Kinderlehrer and Stampacchia, 2000, Chapter II, Theorem A.1] or [Attouch et al., 2006, Theorem 5.8.2], and standard results, it follows straightforwardly that $H_0^1(\Omega)$ is a regular Dirichlet space with

$$\mathcal{E}(v_1, v_2) := \int_\Omega \nabla v_1 \cdot \nabla v_2 \, \mathrm{d}\mathcal{L}^d \qquad \forall v_1, v_2 \in H_0^1(\Omega).$$

Moreover, Assumption 3.1.1 is trivially satisfied in the situation of Corollary 3.4.3. We may thus employ Theorems 3.3.5 and 3.4.2 (with $F = \mathrm{Id}$ and $L = K$) to deduce that K is polyhedric everywhere, that S is directionally differentiable and that the directional derivatives $S'(f; g)$ are characterized by (3.13). This proves the claim. □

Corollary 3.4.4 (Directional Differentiability for a Signorini Problem)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a bounded domain with a $C^{1,1}$-boundary and let $H^1(\Omega)$ be defined as in [Attouch et al., 2006, Section 5.1]. Denote with $\nu : \partial\Omega \to \mathbb{R}^d$ the outward unit normal vector field on $\partial\Omega$, with $\mathrm{tr} : H^1(\Omega) \to H^{1/2}(\partial\Omega)$ the usual trace operator (cf. [Attouch et al., 2006, Proposition 5.6.3]) and with $\mathrm{tr}_\nu : H^1(\Omega)^d \to H^{1/2}(\partial\Omega)$, $v \mapsto \mathrm{tr}(v) \cdot \nu$ the normal trace. Suppose that $\partial\Omega$ is equipped with the $(d-1)$-dimensional Hausdorff measure $\mathcal{H}^{d-1}$ (scaled as in [Evans and Gariepy, 2015, Definition 2.1]) and assume that two Borel measurable functions $\psi_1, \psi_2 : \partial\Omega \to [-\infty, \infty]$ are given such that*

$$L := \left\{ u \in H^{1/2}(\partial\Omega) \;\middle|\; \psi_1 \leq u \leq \psi_2 \; \mathcal{H}^{d-1}\text{-a.e. on } \partial\Omega \right\} \neq \emptyset. \tag{3.14}$$

Define further $K := \mathrm{tr}_\nu^{-1}(L)$ and suppose that $A : K \to [H^1(\Omega)^d]^*$ is an operator which satisfies the conditions in Assumption 3.1.1. Then, the solution map $S : [H^1(\Omega)^d]^* \to H^1(\Omega)^d$, $f \mapsto w$, to the problem

$$w \in K, \qquad \langle A(w), v - w \rangle \geq \langle f, v - w \rangle \qquad \forall v \in K \tag{3.15}$$

is Hadamard directionally differentiable in all $f \in [H^1(\Omega)^d]^*$. Moreover, for every $f \in [H^1(\Omega)^d]^*$ with associated solution $w := S(f)$ there exists a unique multiplier $\lambda \in \mathcal{N}_L(\mathrm{tr}_\nu(w))$ with $f - A(w) = \mathrm{tr}_\nu^*(\lambda)$ and the directional derivatives $\delta := S'(f; g)$, $f, g \in [H^1(\Omega)^d]^*$, of the solution map S are uniquely characterized by the variational inequality

$$\delta \in \mathrm{tr}_\nu^{-1}\left(\mathcal{T}_L(\mathrm{tr}_\nu(w)) \cap \ker(\lambda)\right), \quad \left\langle A'(w)\delta, z - \delta \right\rangle \geq \langle g, z - \delta \rangle \quad \forall z \in \mathrm{tr}_\nu^{-1}\left(\mathcal{T}_L(\mathrm{tr}_\nu(w)) \cap \ker(\lambda)\right). \tag{3.16}$$

Proof. Using standard density results and the facts collected in [Musina and Nazarov, 2017], it is easy to check that $H^{1/2}(\partial\Omega)$ is a regular Dirichlet space with

$$\mathcal{E}(u_1, u_2) := \int_{\partial\Omega} \int_{\partial\Omega} \frac{(u_1(x) - u_1(y))(u_2(x) - u_2(y))}{|x - y|^d} \mathrm{d}\mathcal{H}^{d-1}(x) \mathrm{d}\mathcal{H}^{d-1}(y).$$

See also [Haraux, 1977, Example 6] for an alternative choice of the bilinear form $\mathcal{E}$. This shows that Theorem 3.4.2 is applicable and that the set L in (3.14) is polyhedric. From [Kikuchi and Oden, 1988, Theorem 5.5], we obtain further that the normal trace $\mathrm{tr}_\nu : H^1(\Omega)^d \to H^{1/2}(\partial\Omega)$ is surjective. The latter implies in combination with the linearity and the continuity of tr_ν, Proposition 2.4.3 and Lemma 3.1.2 that $\mathcal{N}_K(v) = \mathrm{tr}_\nu^* \mathcal{N}_L(\mathrm{tr}_\nu(v))$ holds for all $v \in K$ and that for every $\varphi \in \mathcal{N}_K(v)$ we can find exactly one $\eta \in \mathcal{N}_L(\mathrm{tr}_\nu(v))$ with $\varphi = \mathrm{tr}_\nu^*(\eta)$ (due to the injectivity of the adjoint). This proves the existence and the uniqueness of the Lagrange multiplier λ for all f. Note that the surjectivity of tr_ν further yields that (3.8) is satisfied for all $w = S(f)$ (with $F := \mathrm{tr}_\nu$, $V := H^1(\Omega)^d$ and $U := H^{1/2}(\partial\Omega)$). We may thus again employ Theorem 3.3.5 to deduce that S is directionally differentiable and that the directional derivatives $S'(f; g)$ satisfy (3.16). The claim now follows immediately. □

Some remarks are in order regarding Corollaries 3.4.3 and 3.4.4:

Remark 3.4.5.

(i) The obstacle problem (3.12) *with $A := -\Delta$ characterizes the equilibrium state of a membrane that is fixed at the boundary $\partial\Omega$, deformed by an external force f and enclosed from above and below by two impenetrable obstacles ψ_1 and ψ_2. Problems of the type* (3.15)*, on the other hand, can be used to model the elastic deformation of a body that collides with a rigid surface. For more details on the physical background of the variational inequalities in Corollaries 3.4.3 and 3.4.4, we refer to [Betz, 2015; Rodrigues, 1987].*

(ii) Using capacity theory, it is possible to derive more explicit formulas for the admissible sets of the variational inequalities (3.13) *and* (3.16)*. In particular, it can be shown that the appearing elements of the normal cones $\mathcal{N}_K(w)$ and $\mathcal{N}_L(\mathrm{tr}_\nu(w))$, respectively, can be identified with positive Radon measures, see [Fukushima et al., 2011, Section 2.2]. These measures can be interpreted as contact forces that are exerted by the obstacle(s).*

(iii) The results in Corollaries 3.4.3 and 3.4.4 are classical and can be found, e.g., in [Betz, 2015; Haraux, 1977; Mignot, 1976] (albeit in different variants).

Because of Mignot's theorem, sets of the form (3.11) are often regarded as the prime examples of polyhedric sets in infinite dimensions. However, one has to be very careful here, since the polyhedricity result in Theorem 3.4.2 does not hold in general when V is an arbitrary Hilbert space. Consider, e.g., the set

$$K := \{v \in L^2(-\pi,\pi) \mid -1 \leq v \leq 1 \ \mathcal{L}^1\text{-a.e. in } (-\pi,\pi)\} \tag{3.17}$$

which is trivially polyhedric in $L^2(-\pi,\pi)$ by Theorem 3.4.2. Then, K is not polyhedric anymore when we interpret it as a subset of the dual $H^{-1}(-\pi,\pi)$ of the Sobolev space $H_0^1(-\pi,\pi)$ appearing in Corollary 3.4.3. To see this, we define $v := \mathrm{sgn} \in H^{-1}(-\pi,\pi)$ and $\varphi := \sin \in H_0^1(-\pi,\pi)$, where sgn and $\sin$ denote the signum function and the sine, respectively, and identify φ with an element of $H^{-1}(-\pi,\pi)^*$ via

$$\langle \varphi, z\rangle_{H^{-1}} = \langle z, \varphi\rangle_{H_0^1} \quad \forall z \in H^{-1}(-\pi,\pi).$$

Note that the above v and φ trivially satisfy $v \in K$,

$$\mathcal{T}_K^{rad}(v) = \big\{z \in L^\infty(-\pi,\pi) \;\big|\; z \geq 0 \ \mathcal{L}^1\text{-a.e. in } (-\pi,0),\ z \leq 0 \ \mathcal{L}^1\text{-a.e. in } (0,\pi)\big\}$$

and

$$\langle \varphi, z\rangle_{H^{-1}} = \int_{-\pi}^{\pi} \varphi z \mathrm{d}\mathcal{L}^1 = -\int_{-\pi}^{\pi} |\varphi z| \mathrm{d}\mathcal{L}^1 < 0 \qquad \forall z \in \mathcal{T}_K^{rad}(v) \setminus \{0\}.$$

The function φ is thus an element of the normal cone $\mathcal{N}_K(v)$ and it holds

$$\mathcal{T}_K^{rad}(v) \cap \ker(\varphi) = \{0\}. \tag{3.18}$$

Consider now for arbitrary but fixed $\alpha, \beta > 0$ the sequence

$$z_n := \alpha n \mathbb{1}_{(-1/n,0)} - \beta n \mathbb{1}_{(0,1/n)} \qquad \forall n \in \mathbb{N},$$

where $\mathbb{1}_D : \mathbb{R} \to \{0,1\}$ denotes the indicator function of a set $D \subset \mathbb{R}$. Then, $\{z_n\}$ is obviously a subset of the radial cone $\mathcal{T}_K^{rad}(v)$ and it is easy to check that $z_n \to (\alpha-\beta)\delta_0$ holds in $H^{-1}(-\pi,\pi)$, where δ_0 is the Dirac delta at the origin. In particular, $\mathbb{R}\delta_0 \subset \mathcal{T}_K(v)$. From $\langle \varphi, z\rangle_{H^{-1}} = 0$ for all $z \in \mathbb{R}\delta_0 \subset \mathcal{T}_K(v)$ and (3.18), it now follows immediately that

$$\mathbb{R}\delta_0 \subset \mathcal{T}_K(v) \cap \ker(\varphi) \neq \mathrm{cl}\big(\mathcal{T}_K^{rad}(v) \cap \ker(\varphi)\big) = \{0\}.$$

This shows that the set K in (3.17) is indeed non-polyhedric in $H^{-1}(-\pi,\pi)$, that Theorem 3.4.2 does not hold for arbitrary V, and that the curvature properties of sets with upper and lower bounds depend largely on the structure of the ambient space.

The non-polyhedricity of the set K in (3.17) is all the more surprising when one realizes that the set

$$L := \left\{ z \in H^{-1}(-\pi,\pi) \;\middle|\; \langle z, v\rangle_{H_0^1} + \int_{-\pi}^{\pi} v \,\mathrm{d}\mathcal{L}^1 \geq 0 \quad \forall 0 \leq v \in H_0^1(-\pi,\pi) \right\}$$

is polyhedric in $H^{-1}(-\pi,\pi)$ (by Theorem 3.4.2, [Wachsmuth, 2016, Lemma 3.2] and an elementary translation argument) and that K satisfies

$$K = L \cap -L.$$

What we have constructed here is thus an example of a closed, convex, non-empty set that can be written as the intersection of two polyhedric sets but fails to be polyhedric itself (cf. [Wachsmuth, 2016, Example 4.23]). The instability of the property of polyhedricity under intersection, that becomes apparent in the above, is remarkable because it contrasts what is typically observed for sets deemed "polyhedric". If we consider, e.g., polyhedra in the classical sense, i.e., sets that are finite intersections of half-spaces, cf. [Cottle et al., 2009, Definition 2.6.1], [Bonnans and Shapiro, 2000, Definition 2.195], [Vinberg et al., 2013, Definition 3.9], then the intersection of two polyhedra is trivially again a polyhedron. Note that the property of polyhedricity is in general not even preserved when a polyhedric set is intersected with a closed subspace, cf. Section 3.4.2 and [Wachsmuth, 2016, Example 4.2.2]. Further, it should be noted that sets of the type (3.17) can be shown to be non-polyhedric in the dual of the vast majority of Dirichlet spaces. See [Christof and Wachsmuth, 2017c] for details.

We would like to point out that the non-polyhedricity of the set K in (3.17) is not just an academic curiosity but also important for practical applications. Sets of the form (3.17) appear, for example, when elliptic variational inequalities involving L^1-norms are dualized and play a major role in the sensitivity analysis of contact problems with prescribed friction, see, e.g., [Sokołowski and Zolésio, 1992, Section 4.5], [Sokołowski, 1988], [Sokołowski and Zolésio, 1988] (where, surprisingly, the difficulty with the non-polyhedricity is largely ignored). We will get back to this topic in Section 5.2.

3.4.2 Non-Polyhedricity for the Elastoplastic Torsion Problem with ∞-Norm

To give a further example of the pitfalls that can be encountered when the concept of polyhedricity is used in infinite dimensions, in what follows, we study the properties of the set

$$L := \left\{ v \in W_0^{1,1}(\Omega) \;\middle|\; \|\nabla v\|_\infty \leq 1 \ \mathcal{L}^d\text{-a.e. in } \Omega \right\} \tag{3.19}$$

in the Sobolev spaces $W_0^{1,q}(\Omega)$, $1 \leq q < \infty$. Let us first clarify our assumptions:

Assumption 3.4.6 (Standing Assumptions for the Study of the Set L in (3.19))**.**

- $\Omega \subset \mathbb{R}^d$, $d \geq 1$, *is a bounded Lipschitz domain (endowed with the Lebesgue measure $\mathcal{L}^d$),*
- *the spaces $W_{(0)}^{1,q}(\Omega)$, $1 \leq q \leq \infty$, are defined as in [Attouch et al., 2006, Section 5.1],*
- $\|x\|_\infty := \max(|x_1|, ..., |x_d|)$ *is the classical maximum norm on $\mathbb{R}^d$.*

Sets of the type (3.19) appear, for example, in the so-called elastoplastic torsion problem which models the behavior of an isotropic and homogeneous cylinder that is twisted along its axis. In the latter context, the dimension d is typically two, the domain $\Omega \subset \mathbb{R}^2$ represents the cross-section of the cylinder at hand and the pointwise gradient constraint is directly related to the employed yield criterion. See, e.g., [Rodrigues, 1987, Section 1.6] for the derivation of the corresponding EVI (with von Mises yield stress) and compare also with [Hintermüller and Surowiec, 2011, Section 5.2], [Ekeland and Temam, 1976, Section 3.4] and [Glowinski, 2015, Section 1.3.4.2].

Note that the set L in (3.19) is trivially non-empty, closed and convex in $W_0^{1,q}(\Omega)$ for all $1 \leq q < \infty$, that $\nabla : W_0^{1,q}(\Omega) \to \nabla W_0^{1,q}(\Omega) \subset L^q(\Omega, \mathbb{R}^d)$ defines an isomorphism by the inequality of Poincaré-Friedrichs (see [Attouch et al., 2006, Theorem 5.3.1]), and that

$$\nabla L = \Big\{ u \in L^q(\Omega, \mathbb{R}^d) \;\Big|\; \|u\|_\infty \leq 1 \; \mathcal{L}^d\text{-a.e. in } \Omega \Big\} \cap \nabla W_0^{1,q}(\Omega) \subset L^q(\Omega, \mathbb{R}^d)$$

holds for all $1 \leq q < \infty$. The latter implies that the set L can be identified with the intersection of a closed subspace of $L^q(\Omega, \mathbb{R}^d)$ (namely, the space $\nabla W_0^{1,q}(\Omega)$) and a polyhedric subset of $L^q(\Omega, \mathbb{R}^d)$ (namely, the set $\{u \in L^q(\Omega, \mathbb{R}^d) \mid \|u\|_\infty \leq 1 \; \mathcal{L}^d\text{-a.e. in } \Omega\}$, see [Wachsmuth, 2016, Lemma 4.20]) for all $1 \leq q < \infty$. In the remainder of this section, we will prove that, in spite of this property, the set L is in general not polyhedric in $W_0^{1,q}(\Omega)$. Our main result reads as follows:

Theorem 3.4.7. *Consider the situation in Assumption 3.4.6. Then:*

(i) The set L in (3.19) *is polyhedric in the space $W_0^{1,q}(\Omega)$ for all $1 \leq q < \infty$ if $d = 1$.*

(ii) The set L in (3.19) *is not polyhedric in the space $W_0^{1,q}(\Omega)$ if $d > 1$ and $1 \leq q < d$.*

To obtain Theorem 3.4.7, we proceed in several steps. We begin by proving that it indeed makes no difference whether we study the polyhedricity of the set L in the space $W_0^{1,q}(\Omega)$ or the polyhedricity of the set ∇L in the space $L^q(\Omega, \mathbb{R}^d)$:

Lemma 3.4.8. *Let $1 \leq q < \infty$ be arbitrary but fixed. Then, the set L in* (3.19) *is polyhedric in $W_0^{1,q}(\Omega)$ if and only if the set ∇L is polyhedric in $L^q(\Omega, \mathbb{R}^d)$.*

Proof. From Friedrichs' inequality, it follows that $\nabla W_0^{1,q}(\Omega)$ is a Banach space when endowed with the $L^q(\Omega, \mathbb{R}^d)$-norm and that $\nabla : W_0^{1,q}(\Omega) \to \nabla W_0^{1,q}(\Omega)$ is an isomorphism. The latter implies that the polyhedricity of L in $W_0^{1,q}(\Omega)$ is equivalent to the polyhedricity of ∇L in $\nabla W_0^{1,q}(\Omega)$, cf. [Wachsmuth, 2016, Lemma 3.3]. From [Christof and Wachsmuth, 2017c, Lemma 2.4], it follows further that ∇L is polyhedric in $\nabla W_0^{1,q}(\Omega)$ if and only if ∇L is polyhedric in $L^q(\Omega, \mathbb{R}^d)$. This proves the claim. □

If we combine Lemma 3.4.8 with [Wachsmuth, 2016, Theorem 4.18], then part (i) of Theorem 3.4.7 follows immediately:

Proof of Theorem 3.4.7(i). Suppose that $d = 1$ and consider an arbitrary but fixed $1 \leq q < \infty$. Then, there exist $-\infty < a < b < \infty$ with $\Omega = (a, b)$, and it holds

$$\begin{aligned} \nabla L &= L' \\ &= \big\{ u \in L^q(a,b) \;\big|\; |u| \leq 1 \; \mathcal{L}^1\text{-a.e. in } (a,b) \big\} \cap W_0^{1,q}(a,b)' \\ &= \left\{ u \in L^q(a,b) \;\middle|\; |u| \leq 1 \; \mathcal{L}^1\text{-a.e. in } (a,b) \text{ and } \int_a^b u \,\mathrm{d}\mathcal{L}^1 = 0 \right\} \\ &= \big\{ u \in L^q(a,b) \;\big|\; -1 \leq u \leq 1 \; \mathcal{L}^1\text{-a.e. in } (a,b) \big\} \cap \ker(\eta), \end{aligned} \tag{3.20}$$

where

$$\eta : L^q(a,b) \to \mathbb{R}, \qquad \eta(u) := \int_a^b u \,\mathrm{d}\mathcal{L}^1,$$

and where a prime denotes a weak derivative. From [Wachsmuth, 2016, Example 4.21], we obtain that the set $M := \{u \in L^q(a,b) \mid -1 \leq u \leq 1 \; \mathcal{L}^1\text{-a.e. in } (a,b)\}$ in (3.20) is n-polyhedric for all $n \in \mathbb{N}$, i.e., for all $\varphi_1, ..., \varphi_n \in L^q(a,b)^*$, $n \in \mathbb{N}$, and all $u \in M$ it is true that

$$\mathcal{T}_M(u) \cap \ker(\varphi_1) \cap ... \cap \ker(\varphi_n) = \mathrm{cl}\big(\mathcal{T}_M^{rad}(u) \cap \ker(\varphi_1) \cap ... \cap \ker(\varphi_n)\big).$$

The latter implies in particular that

$$\begin{aligned}\mathcal{T}_{M\cap\ker(\eta)}(u)\cap\ker(\varphi) &= \mathrm{cl}\big(\mathcal{T}^{rad}_{M\cap\ker(\eta)}(u)\big)\cap\ker(\varphi)\\ &= \mathrm{cl}\big(\mathbb{R}^+(M-u)\cap\ker(\eta)\big)\cap\ker(\varphi)\\ &\subset \mathcal{T}_M(u)\cap\ker(\eta)\cap\ker(\varphi)\\ &= \mathrm{cl}\big(\mathcal{T}^{rad}_M(u)\cap\ker(\eta)\cap\ker(\varphi)\big)\\ &= \mathrm{cl}\big(\mathcal{T}^{rad}_{M\cap\ker(\eta)}(u)\cap\ker(\varphi)\big)\\ &\subset \mathcal{T}_{M\cap\ker(\eta)}(u)\cap\ker(\varphi)\end{aligned}$$

holds for all $u \in M\cap\ker(\eta)$ and all $\varphi \in L^q(a,b)^*$. This shows that the set $L' = M\cap\ker(\eta)$ is polyhedric in the space $L^q(a,b)$. The assertion of Theorem 3.4.7(i) is now an immediate consequence of Lemma 3.4.8. □

To prove the second part of Theorem 3.4.7, we will construct a point $v \in L$ and a $\varphi \in W^{1,q}_0(\Omega)^*$ such that the condition

$$\mathcal{T}_L(v)\cap\ker(\varphi) = \mathrm{cl}\big(\mathcal{T}^{rad}_L(v)\cap\ker(\varphi)\big) \tag{3.21}$$

is violated. Recall that we do not have to ensure $\varphi \in \mathcal{N}_L(v)$ here by Remark 3.3.2(i). In what follows, the basic idea of our analysis is to exploit that the radial cone $\mathcal{T}^{rad}_L(v)$ is a subset of the continuous functions for all $v \in L$ and all $1 \le q < \infty$ (by its definition and the properties of L) while the tangent cone $\mathcal{T}_L(v)$ may contain elements that are discontinuous when $1 \le q < d$, cf. the classical Sobolev embeddings in [Adams, 1975, Theorem 5.4]. Note that, in the case of the set K in (3.17), we have used a similar discrepancy between the elements of the radial and the tangent cone, namely, that $\mathcal{T}^{rad}_K(v) \subset L^2(-\pi,\pi)$ and that $\mathcal{T}_K(v)\setminus L^2(-\pi,\pi) \neq \emptyset$. The main challenge in the construction of a tuple (v,φ) that illustrates the non-polyhedricity of the set L in (3.19) is to choose v and φ in such a way that the continuity of the functions in the radial cone indeed causes the sets $\mathcal{T}_L(v)\cap\ker(\varphi)$ and $\mathrm{cl}\big(\mathcal{T}^{rad}_L(v)\cap\ker(\varphi)\big)$ to be different. We will realize this by "delocalizing" the property of continuity.

To construct our counterexample, we have to introduce some notation:

Assumption 3.4.9 (Standing Assumptions and Notation for the Construction of the Counterexample)**.**

- $d > 1$ *and* $1 \le q < d$ *are arbitrary but fixed,*
- *the standard basis of* $\mathbb{R}^d$ *is denoted with* $e_1, ..., e_d$*,*
- $\mathbf{1} := e_1 + ... + e_d$*,*
- $\tilde{e}_1 := \mathbf{1}/\sqrt{d}$*,*
- $\tilde{e}_2, ..., \tilde{e}_d$ *are chosen such that* $\tilde{e}_1, ..., \tilde{e}_d$ *is an orthonormal basis w.r.t. the standard scalar product,*
- *the coordinates of a point w.r.t.* $\{e_n\}$ *and* $\{\tilde{e}_n\}$ *are denoted with* $\{x_n\}$ *and* $\{\alpha_n\}$*, respectively,*
- $\Omega = \big\{x \in \mathbb{R}^d \;\big|\; \|x\|_2 < 3d\big\}$*, where* $\|\cdot\|_2$ *denotes the Euclidean norm,*
- $$D := \left\{ x = \sum_{n=1}^d \alpha_n \tilde{e}_n \;\middle|\; \alpha_1 \in (-1,1), \quad \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < 1 \right\},$$
$$Z_+ := \left\{ x = \sum_{n=1}^d \alpha_n \tilde{e}_n \;\middle|\; \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < \alpha_1 < 1 \right\},$$
$$Z_- := \left\{ x = \sum_{n=1}^d \alpha_n \tilde{e}_n \;\middle|\; \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < -\alpha_1 < 1 \right\}.$$

Note that the set D is a slanted circular hypercylinder of height two with axis parallel to $\mathbf{1}$ and a base of radius $2d$, that the sets Z_+ and Z_- are the Lorentz cones in D whose bases are the $(d-1)$-dimensional flat discs of radius $2d$ in the boundary of D and whose tips are at the origin, and that $\mathrm{cl}(Z_\pm) \subset \mathrm{cl}(D) \subset \Omega$ holds by the choice of Ω. The sets D, Z_+ and Z_- thus form a "tipped hourglass" in Ω, see Figure 3.2.

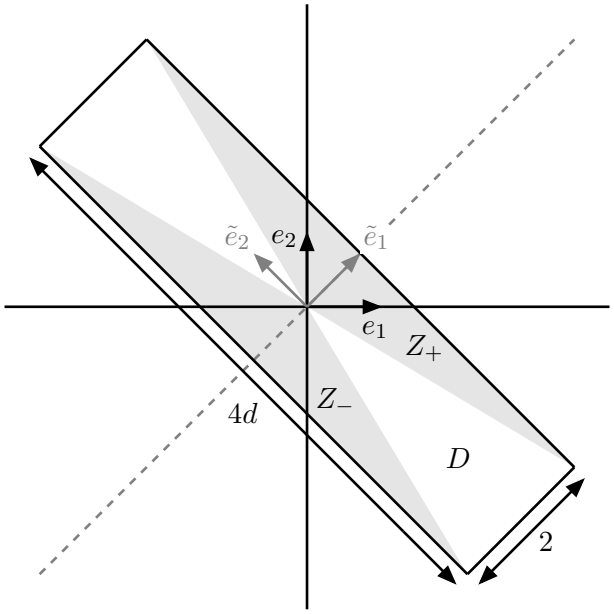

Figure 3.2: The sets D, Z_+ and Z_- for $d = 2$.

We can now introduce the function $v \in W_0^{1,q}(\Omega)$ that we use in our counterexample:

Lemma 3.4.10. *Let $v : \Omega \to \mathbb{R}$ be defined by*

$$v:\ x = \sum_{n=1}^{d} \alpha_n \tilde{e}_n \mapsto \begin{cases} \sqrt{d}\,(1 - |\alpha_1|) & \text{if } x \in Z_+ \cup Z_- \\ \sqrt{d}\left(1 - \frac{1}{2d}\sqrt{\sum_{n=2}^{d} \alpha_n^2}\right) & \text{if } x \in D \setminus (Z_+ \cup Z_-)\,. \\ 0 & \text{else} \end{cases} \tag{3.22}$$

Then, v is an element of the set L in (3.19) *and it holds*

$$\begin{aligned} &v = 0 \text{ and } \nabla v = 0 \quad \mathcal{L}^d\text{-a.e. in } \Omega \setminus \mathrm{cl}(D), \\ &\nabla v = -\operatorname{sgn}(\alpha_1)\mathbf{1} \quad \mathcal{L}^d\text{-a.e. in } Z_+ \cup Z_-, \\ &\|\nabla v\|_\infty \leq \frac{1}{2} \quad \mathcal{L}^d\text{-a.e. in } D \setminus (Z_+ \cup Z_-). \end{aligned} \tag{3.23}$$

Proof. It is easy to check that v is continuous in Ω and $C^{1,1}$ in each of the sets Z_+, Z_-, $D \setminus \mathrm{cl}(Z_+ \cup Z_-)$ and $\Omega \setminus \mathrm{cl}(D)$. This yields that v is Lipschitz in Ω and identifiable with an element of $W^{1,\infty}(\Omega)$, cf. [Ambrosio et al., 2000, Proposition 2.13]. From Rademacher's theorem, [Ambrosio et al., 2000, Theorem 2.14], it follows further that the classical gradient of v in Z_+, Z_-, $D\setminus\mathrm{cl}(Z_+\cup Z_-)$ and $\Omega\setminus\mathrm{cl}(D)$ coincides almost everywhere with the weak gradient ∇v (which we denote with $\nabla_x v$ in the following to emphasize that we employ the coordinate system $\{e_n\}$). Using the latter and a straightforward calculation, we immediately obtain that $v = 0$ and $\nabla_x v = 0$ holds a.e. in $\Omega \setminus \mathrm{cl}(D)$ and that, for almost all $x \in Z_+ \cup Z_-$, we have

$$\nabla_x v(x) = -\sqrt{d}\nabla_x\left(|\alpha_1|\right) = -\operatorname{sgn}(\alpha_1)\sqrt{d}\nabla_x(\tilde{e}_1 \cdot x) = -\operatorname{sgn}(\alpha_1)\sqrt{d}\tilde{e}_1 = -\operatorname{sgn}(\alpha_1)\mathbf{1}. \tag{3.24}$$

Here and in what follows, a dot between two vectors denotes the standard scalar product on $\mathbb{R}^d$. This proves the second line in (3.23). Analogously to (3.24), we may compute that

$$\nabla_x v(x) = -\frac{1}{2\sqrt{d}} \nabla_x \left(\sqrt{\sum_{n=2}^{d} (\tilde{e}_n \cdot x)^2} \right) = -\frac{1}{2\sqrt{d}} \frac{1}{\sqrt{\sum_{n=2}^{d} (\tilde{e}_n \cdot x)^2}} \sum_{n=2}^{d} (\tilde{e}_n \cdot x) \tilde{e}_n$$

and, consequently,

$$\|\nabla_x v(x)\|_\infty \leq \|\nabla_x v(x)\|_2 \leq \frac{1}{2\sqrt{d}} \frac{1}{\sqrt{\sum_{n=2}^{d} (\tilde{e}_n \cdot x)^2}} \sum_{n=2}^{d} |\tilde{e}_n \cdot x| \|\tilde{e}_n\|_2 \leq \frac{1}{2}$$

holds for almost all $x \in D \setminus (Z_+ \cup Z_-)$. If we combine all of the above and use that D, Z_+ and Z_- have boundaries of measure zero, then the claim follows immediately. □

In the following lemma, we construct the element z of the tangent cone $\mathcal{T}_L(v)$ that will yield a contradiction with the polyhedricity condition (3.21).

Lemma 3.4.11. *Denote with v the function in* (3.22)*, let $\psi \in C_c^\infty(\Omega)$ be an arbitrary but fixed bump function satisfying $0 \leq \psi \leq 1$ in Ω and $\psi \equiv 1$ in D, and let $z : \Omega \to \mathbb{R}$ be defined by*

$$z(x) := \begin{cases} 2\psi(x) & \text{if } \alpha_1 > \frac{1}{2d} \sqrt{\sum_{n=2}^{d} \alpha_n^2} \\ 0 & \text{if } \alpha_1 \leq -\frac{1}{2d} \sqrt{\sum_{n=2}^{d} \alpha_n^2} \\ \psi(x) \left(\frac{2d\alpha_1}{\sqrt{\sum_{n=2}^{d} \alpha_n^2}} + 1 \right) & \text{else} \end{cases} . \tag{3.25}$$

Then, z is an element of $\mathcal{T}_L(v)$ (where the closure in the definition of $\mathcal{T}_L(v)$ is taken in $W_0^{1,q}(\Omega)$).

Note that the function z constructed above is discontinuous at the origin. This is consistent with the strategy that we have outlined after Lemma 3.4.8.

Proof of Lemma 3.4.11. Extend v by zero, define

$$v_\varepsilon(x) := 3 - 4v\left(\frac{x}{\varepsilon}\right), \quad x \in \Omega, \quad \varepsilon \in (0,1),$$

and set

$$z_\varepsilon := \min(v_\varepsilon, z).$$

Then, for each $\varepsilon \in (0,1)$, there exists an open neighborhood of the origin where z_ε is identical to v_ε (since $v_\varepsilon(0) = 3 - 4\sqrt{d} < -1$, since v_ε is continuous, and since $0 \leq z \leq 2$ in Ω). This implies in particular that $z_\varepsilon \in W^{1,\infty}(\Omega)$ for all $\varepsilon \in (0,1)$ (since z is Lipschitz away from the origin, cf. the arguments in the proof of Lemma 3.4.10). Taking into account the behavior of v_ε and z at the boundary, we conclude that $z_\varepsilon \in W^{1,\infty}(\Omega) \cap W_0^{1,q}(\Omega)$ for all $\varepsilon \in (0,1)$.

In what follows, we will prove that the functions z_ε are elements of the radial cone $\mathcal{T}_L^{rad}(v)$ and that the family $\{z_\varepsilon\}$ converges to z in $W_0^{1,q}(\Omega)$ for $\varepsilon \searrow 0$. Let us again write $\nabla = \nabla_x$ and define

$$t_\varepsilon := \frac{\varepsilon}{2 + 2\operatorname{ess\,sup}_{\tilde{x} \in \Omega} \|\nabla_x z_\varepsilon(\tilde{x})\|_\infty} \in \left(0, \frac{1}{2}\varepsilon\right).$$

Then, it follows from our construction and Stampacchia's lemma, see [Kinderlehrer and Stampacchia, 2000, Chapter II, Theorem A.1] or [Attouch et al., 2006, Theorem 5.8.2], that

$$\nabla_x\big(v(x)+t_\varepsilon z_\varepsilon(x)\big)=\begin{cases}\nabla_x v(x)-\dfrac{4t_\varepsilon}{\varepsilon}(\nabla_x v)\left(\dfrac{x}{\varepsilon}\right) & \text{a.e. in } \{v_\varepsilon\le z\}\\ \nabla_x v(x)+t_\varepsilon\nabla_x z(x) & \text{a.e. in } \{v_\varepsilon> z\}\end{cases},$$

where, as usual, $\{v_\varepsilon\sim z\}$ is short for $\{x\in\Omega\mid v_\varepsilon(x)\sim z(x)\}$ (defined up to sets of measure zero), and we may use the properties in Lemma 3.4.10 to compute that

(i) for a.a. $x\in(Z_+\cup Z_-)\cap\{v_\varepsilon\le z\}$ it holds

$$\big\|\nabla_x\big(v(x)+t_\varepsilon z_\varepsilon(x)\big)\big\|_\infty=\left\|-\operatorname{sgn}(\alpha_1)\left(1-\frac{4t_\varepsilon}{\varepsilon}\right)\mathbf{1}\right\|_\infty\le 1,$$

(ii) for a.a. $x\in(Z_+\cup Z_-)\cap\{v_\varepsilon> z\}$ it holds

$$\big\|\nabla_x\big(v(x)+t_\varepsilon z_\varepsilon(x)\big)\big\|_\infty=\|\nabla_x v(x)\|_\infty=1,$$

(iii) for a.a. $x\in D\setminus(Z_+\cup Z_-)$ it holds

$$\big\|\nabla_x\big(v(x)+t_\varepsilon z_\varepsilon(x)\big)\big\|_\infty\le\|\nabla_x v(x)\|_\infty+\frac{\varepsilon\,\|\nabla_x z_\varepsilon(x)\|_\infty}{2+2\operatorname{ess\,sup}_{\tilde x\in\Omega}\|\nabla_x z_\varepsilon(\tilde x)\|_\infty}\le 1,$$

(iv) for a.a. $x\in\Omega\setminus D$ it holds

$$\big\|\nabla_x\big(v(x)+t_\varepsilon z_\varepsilon(x)\big)\big\|_\infty=\frac{\varepsilon\,\|\nabla_x z_\varepsilon(x)\|_\infty}{2+2\operatorname{ess\,sup}_{\tilde x\in\Omega}\|\nabla_x z_\varepsilon(\tilde x)\|_\infty}\le\frac12.$$

The above implies that $v+t_\varepsilon z_\varepsilon$ is an element of L for all $\varepsilon\in(0,1)$ and that $\{z_\varepsilon\}$ is indeed a subset of the radial cone $\mathcal{T}_L^{rad}(v)$. It remains to prove that the family $\{z_\varepsilon\}$ converges to z in $W_0^{1,q}(\Omega)$. Note that the dominated convergence theorem and the inclusion $\{v_\varepsilon\le z\}\subset\varepsilon D$ (up to sets of measure zero) immediately yield $\|z-z_\varepsilon\|_{L^q}\to 0$ for $\varepsilon\searrow 0$. To prove that the gradient fields ∇z_ε converge, too, we consider two arbitrary but fixed $0<\varepsilon_1<\varepsilon_2<1$ and compute (using again Stampacchia's lemma)

$$\begin{aligned}
&\left(\int_\Omega\|\nabla_x z_{\varepsilon_1}-\nabla_x z_{\varepsilon_2}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}\\
&\quad=\left(\int_{\varepsilon_2 D}\|\nabla_x z_{\varepsilon_1}-\nabla_x z_{\varepsilon_2}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}\\
&\quad\le\left(\int_{\varepsilon_2 D}\|\nabla_x z_{\varepsilon_1}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}+\left(\int_{\varepsilon_2 D}\|\nabla_x z_{\varepsilon_2}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}\\
&\quad\le\left(\int_{\varepsilon_2 D}\|\nabla_x v_{\varepsilon_1}\|_1^q\mathbb{1}_{\{v_{\varepsilon_1}<z\}}\,\mathrm{d}\mathcal{L}^d\right)^{1/q}+\left(\int_{\varepsilon_2 D}\|\nabla_x z\|_1^q\mathbb{1}_{\{v_{\varepsilon_1}\ge z\}}\,\mathrm{d}\mathcal{L}^d\right)^{1/q}\\
&\qquad+\left(\int_{\varepsilon_2 D}\|\nabla_x v_{\varepsilon_2}\|_1^q\mathbb{1}_{\{v_{\varepsilon_2}<z\}}\,\mathrm{d}\mathcal{L}^d\right)^{1/q}+\left(\int_{\varepsilon_2 D}\|\nabla_x z\|_1^q\mathbb{1}_{\{v_{\varepsilon_2}\ge z\}}\,\mathrm{d}\mathcal{L}^d\right)^{1/q}\\
&\quad\le\left(\int_{\varepsilon_1 D}\|\nabla_x v_{\varepsilon_1}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}+\left(\int_{\varepsilon_2 D}\|\nabla_x v_{\varepsilon_2}\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}+2\left(\int_{\varepsilon_2 D}\|\nabla_x z\|_1^q\,\mathrm{d}\mathcal{L}^d\right)^{1/q}.
\end{aligned}\tag{3.26}$$

Here and in what follows, $\|\cdot\|_1$ and $\mathbb{1}_D:\mathbb{R}^d\to\{0,1\}$ denote the 1-norm and the indicator function of a set D, respectively, and the expression $\nabla_x z$ on the right-hand side of (3.26) is understood as the distributional gradient of z in the punctured domain $\Omega\setminus\{0\}$ (since we do not know yet that $z\in W^{1,q}(\Omega)$). Note that the Lipschitz continuity of v yields

$$\int_{\varepsilon D}\|\nabla_x v_\varepsilon\|_1^q\,\mathrm{d}\mathcal{L}^d\le C\int_{\varepsilon D}\left(\frac{1}{\varepsilon}\right)^q\mathrm{d}\mathcal{L}^d=C\mathcal{L}^d(D)\varepsilon^{d-q}\qquad\forall\varepsilon\in(0,1)\tag{3.27}$$

with a constant $C > 0$ independent of ε. This shows that the v_ε-terms on the right-hand side of (3.26) tend to zero. To see that the z-term does the same, we calculate (with a generic constant $C > 0$, which may change from step to step but is always independent of ε, and using the chain rule, integral transformation, Fubini's theorem and spherical coordinates)

$$
\begin{aligned}
&\int_{\varepsilon D} \|\nabla_x z\|_1^q \mathrm{d}\mathcal{L}^d \\
&\quad = \int_{\varepsilon D \setminus (Z_+ \cup Z_-)} \|\nabla_x z\|_1^q \mathrm{d}\mathcal{L}^d \\
&\quad = \int_{\varepsilon D \cap \left\{ -\frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < \alpha_1 < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} \right\}} \|\nabla_x z\|_1^q \mathrm{d}\mathcal{L}^d \\
&\quad \leq C \int_{\varepsilon D \cap \left\{ -\frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < \alpha_1 < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} \right\}} \left\| \nabla_\alpha \left(\frac{2d\alpha_1}{\sqrt{\sum_{n=2}^d \alpha_n^2}} \right) \right\|_1^q \mathrm{d}\mathcal{L}^d \\
&\quad \leq C \int_{\varepsilon D \cap \left\{ -\frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < \alpha_1 < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} \right\}} \left\| \left[\frac{1}{\sqrt{\sum_{n=2}^d \alpha_n^2}}, -\alpha_1 \frac{[\alpha_2, ..., \alpha_d]}{\left(\sqrt{\sum_{n=2}^d \alpha_n^2} \right)^3} \right] \right\|_1^q \mathrm{d}\mathcal{L}^d \\
&\quad = C \int_{D \cap \left\{ -\frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < \alpha_1 < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} \right\}} \left\| \left[\frac{1}{\sqrt{\sum_{n=2}^d \alpha_n^2}}, -\alpha_1 \frac{[\alpha_2, ..., \alpha_d]}{\left(\sqrt{\sum_{n=2}^d \alpha_n^2} \right)^3} \right] \right\|_1^q \varepsilon^{d-q} \mathrm{d}\mathcal{L}^d \\
&\quad = C\varepsilon^{d-q} \int_{-1}^1 \int_{\left\{ |\alpha_1| < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < 1 \right\}} \\
&\qquad\qquad \left(\frac{1}{\sqrt{\sum_{n=2}^d \alpha_n^2}} + \sum_{n=2}^d \frac{|\alpha_1 \alpha_n|}{\left(\sqrt{\sum_{n=2}^d \alpha_n^2} \right)^3} \right)^q \mathrm{d}\mathcal{L}^{d-1}(\alpha_2, ..., \alpha_d) \mathrm{d}\mathcal{L}^1(\alpha_1) \\
&\quad \leq C\varepsilon^{d-q} \int_0^1 \int_{\left\{ \alpha_1 < \frac{1}{2d}\sqrt{\sum_{n=2}^d \alpha_n^2} < 1 \right\}} \left(\frac{1}{\sqrt{\sum_{n=2}^d \alpha_n^2}} \right)^q \mathrm{d}\mathcal{L}^{d-1}(\alpha_2, ..., \alpha_d) \mathrm{d}\mathcal{L}^1(\alpha_1) \\
&\quad \leq C\varepsilon^{d-q} \int_0^1 \int_{\alpha_1}^1 r^{d-2-q} \mathrm{d}\mathcal{L}^1(r) \mathrm{d}\mathcal{L}^1(\alpha_1) \\
&\quad \leq C\varepsilon^{d-q} \qquad \forall \varepsilon \in (0,1).
\end{aligned}
$$

The above implies in combination with (3.26) and (3.27) that there exists a constant $C > 0$ independent of ε_1 and ε_2 with

$$
\left(\int_\Omega \|\nabla_x z_{\varepsilon_1} - \nabla_x z_{\varepsilon_2}\|_1^q \mathrm{d}\mathcal{L}^d \right)^{1/q} \leq C \left(\varepsilon_1^{d/q-1} + \varepsilon_2^{d/q-1} \right).
$$

This proves that $\{\nabla_x z_\varepsilon\}$ is Cauchy in $L^q(\Omega, \mathbb{R}^d)$ for $\varepsilon \searrow 0$. Since $z_\varepsilon \to z$ in $L^q(\Omega)$, it now follows that $z_\varepsilon \to z$ in $W^{1,q}(\Omega)$, $z \in W_0^{1,q}(\Omega)$ and $z \in T_L(v)$ as claimed. □

We are now in the position to prove part (ii) of Theorem 3.4.7:

Proof of Theorem 3.4.7(ii). We consider w.l.o.g. the situation in Assumption 3.4.9 (if Ω is arbitrary, then we choose an open ball in Ω and use exactly the same argumentation). Let v and z be constructed as in Lemmas 3.4.10 and 3.4.11, and let $\varphi \in W_0^{1,q}(\Omega)^*$ be defined by

$$\varphi : W_0^{1,q}(\Omega) \to \mathbb{R}, \qquad u \mapsto \int_{Z_+} \mathbf{1} \cdot \nabla u \, \mathrm{d}\mathcal{L}^d - \int_{Z_-} \mathbf{1} \cdot \nabla u \, \mathrm{d}\mathcal{L}^d.$$

Then, it follows from

$$\nabla v = -\mathbf{1} \ \ \mathcal{L}^d\text{-a.e. in } Z_+ \quad \text{and} \quad \nabla v = \mathbf{1} \ \mathcal{L}^d\text{-a.e. in } Z_-$$

that

$$\mathcal{T}_L(v) \subset \big\{ u \in W_0^{1,q}(\Omega) \,\big|\, \nabla u \geq 0 \text{ componentwise } \mathcal{L}^d\text{-a.e. in } Z_+, \\ \nabla u \leq 0 \text{ componentwise } \mathcal{L}^d\text{-a.e. in } Z_- \big\},$$

and we obtain

$$\langle \varphi, u \rangle = \int_{Z_+} \|\nabla u\|_1 \mathrm{d}\mathcal{L}^d + \int_{Z_-} \|\nabla u\|_1 \mathrm{d}\mathcal{L}^d \geq 0 \qquad \forall u \in \mathcal{T}_L(v).$$

The latter implies in particular that every $u \in \mathcal{T}_L(v) \cap \ker(\varphi)$ satisfies $\nabla u = 0$ $\mathcal{L}^d$-a.e. in $Z_+ \cup Z_-$, i.e., for every $u \in \mathcal{T}_L(v) \cap \ker(\varphi)$ there exist constants $c_+, c_- \in \mathbb{R}$ with $u \equiv c_+$ in Z_+ and $u \equiv c_-$ in Z_-. Since the radial cone $\mathcal{T}_L^{rad}(v)$ is a subset of the continuous functions and since $\mathrm{cl}(Z_+) \cap \mathrm{cl}(Z_-) = \{0\}$, the constants c_+ and c_- have to be equal for all $u \in \mathcal{T}_L^{rad}(v) \cap \ker(\varphi)$ and it has to hold

$$\mathrm{cl}\big(\mathcal{T}_L^{rad}(v) \cap \ker(\varphi)\big) \subset \Big\{ u \in W_0^{1,q}(\Omega) \;\Big|\; \exists c \in \mathbb{R} \text{ with } u \equiv c \text{ in } Z_+ \cup Z_- \Big\}.$$

From Lemma 3.4.11 and the explicit formula (3.25) for z, we obtain, on the other hand, that

$$z \in \mathcal{T}_L(v), \quad z \equiv 2 \text{ in } Z_+ \quad \text{and} \quad z \equiv 0 \text{ in } Z_-.$$

Consequently,

$$z \in \big(\mathcal{T}_L(v) \cap \ker(\varphi)\big) \setminus \mathrm{cl}\big(\mathcal{T}_L^{rad}(v) \cap \ker(\varphi)\big).$$

This proves that the set L is indeed non-polyhedric in $W_0^{1,q}(\Omega)$. □

We conclude this section with some remarks:

Remark 3.4.12.

(i) The constantness of the functions in $\mathcal{T}_L(v) \cap \ker(\varphi)$ on the sets Z_+ and Z_-, that is obtained from the interplay between L, v and φ in the proof of Theorem 3.4.7(ii), makes the property of continuity at the origin detectable for the $W^{1,q}$-norm. This is what we meant with "delocalization" in the comments after Lemma 3.4.8.

(ii) The counterexample constructed in this section demonstrates that the constraint qualifications in [Wachsmuth, 2016, Lemma 3.3] cannot be dropped. It further illustrates that the intersection of a polyhedric set with a closed subspace does not necessarily have to be polyhedric and that the curvature properties of a functional of the form $j : W_0^{1,q}(\Omega) \to \mathbb{R}$, $v \mapsto k(\nabla v)$ with some $k : L^q(\Omega, \mathbb{R}^d) \to \mathbb{R}$ do not necessarily have to be related to those of the generating function k. Compare also with Theorem 2.4.8 in this context. The fact that all of these effects occur when the operator $\nabla : W_0^{1,q}(\Omega) \to L^q(\Omega, \mathbb{R}^d)$ is considered is rather inconvenient because gradient fields naturally play an important role in many practical applications and classical problems.

(iii) It is, at least to the author's best knowledge, currently completely unknown if the set L in (3.19) *is extended polyhedric or if the characteristic function χ_L is twice epi-differentiable. Further research is necessary here.*

4 Application to EVIs of the Second Kind

In this chapter, we apply the abstract results of Sections 1.3 and 1.4 to elliptic variational inequalities that do not fit into the setting of Assumption 3.1.1. Motivated by what is encountered in practice, we focus in particular on problems that involve seminorms and on EVIs that can be rewritten as non-smooth or degenerate partial differential equations. The structure of this chapter is as follows:

In Section 4.1, we begin our analysis with a simple corollary of Theorems 1.4.1 and 2.1.1 that allows to study partial differential equations that contain non-smooth Nemytskii operators. The usefulness of the main result of this section, Theorem 4.1.1, is illustrated by means of the simple example $w \in H_0^1(\Omega)$, $-\Delta w - \alpha \nabla \cdot \max(0, \nabla w) + \beta \max(0, w) = f$, $\alpha, \beta \geq 0$, $f \in H^{-1}(\Omega)$. The subsequent Section 4.2 is concerned with the second-order epi-differentiability of functions of the form $j(\cdot) := \sum_{m=1}^{M} k_m(|\cdot|_m)$, where k_m is C^2, where $|\cdot|_m$ is a seminorm and where $M \in \mathbb{N}$. Here, we also comment on a rather peculiar regularization effect - the so-called regularization by singular curvature - that can be exploited, e.g., in certain bilevel optimization problems, cf. Section 4.3.3. In Section 4.3, we combine the analysis of Section 4.2 with the findings of Chapter 2 to obtain two theorems, Theorem 4.3.3 and Theorem 4.3.16, that allow to prove, for instance, the directional differentiability of the solution operator to the EVI of static elastoplasticity as considered in [De los Reyes et al., 2016], and that prepare the ground for the analysis in Chapter 5, cf. the remarks in Sections 5.1.1 and 5.2.1.

4.1 Non-Smooth Partial Differential Equations

Before turning our attention to "real" elliptic variational inequalities of the second kind, we consider the case where the functional j in Assumption 1.2.1 possesses a first derivative and where the problem (P) can be rewritten as an equation. In this situation, we may combine Theorems 1.4.1 and 2.1.1 to obtain:

Theorem 4.1.1. *Suppose that V is a Hilbert space and let $A : V \to V^*$ be a Fréchet differentiable and strongly monotone operator that maps bounded subsets of V into bounded subsets of V^*. Assume that $j : V \to \mathbb{R}$ is a convex and continuously differentiable function whose first derivative $j' : V \to V^*$ is directionally differentiable everywhere. Then, the variational equality*

$$A(w) + j'(w) = f \tag{4.1}$$

has a unique solution $w \in V$ for all $f \in V^*$ and the solution operator $S : V^* \to V$, $f \mapsto w$, is Hadamard directionally differentiable everywhere. Moreover, the directional derivative $\delta := S'(f; g)$ in a point $f \in V^*$ with associated solution $w := S(f)$ in a direction $g \in V^*$ is uniquely characterized by the elliptic variational inequality

$$\delta \in V, \quad \big\langle A'(w)\delta, z - \delta \big\rangle + \frac{1}{2} \big\langle (j')'(w; z), z \big\rangle - \frac{1}{2} \big\langle (j')'(w; \delta), \delta \big\rangle \geq \langle g, z - \delta \rangle \quad \forall z \in V.$$

Proof. From the convexity of j, it follows straightforwardly that (4.1) can be rewritten in the form (P), and from Theorem 2.1.1, we obtain that j is twice epi-differentiable in all $v \in V$ for all $\varphi \in \partial j(v)$ with $Q_j^{v,\varphi}(z) = \langle (j')'(v; z), z \rangle$. The claim now follows immediately from Theorems 1.2.2 and 1.4.1. □

To illustrate the applicability of Theorem 4.1.1, let us consider the non-smooth partial differential equation

$$w \in H_0^1(\Omega), \quad -\Delta w - \alpha \nabla \cdot \max(0, \nabla w) + \beta \max(0, w) = f \in H^{-1}(\Omega), \tag{4.2}$$

where $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is a bounded domain, where the spaces $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ are defined as in [Attouch et al., 2006, Definition 5.1.4, Theorem 5.2.1], where $\max(0, \cdot)$ acts componentwise on the weak gradient $\nabla w \in L^2(\Omega, \mathbb{R}^d)$ and where $\alpha, \beta \geq 0$ are given constants. Note that the inequality of Poincaré-Friedrichs (see [Attouch et al., 2006, Theorem 5.3.1]) and the dominated convergence theorem yield that the operator $A := -\Delta : H_0^1(\Omega) \to H^{-1}(\Omega)$ is strongly monotone, that the function

$$j(v) := \frac{1}{2} \int_\Omega \alpha \| \max(0, \nabla v) \|_2^2 + \beta \max(0, v)^2 \mathrm{d}\mathcal{L}^d, \qquad \|x\|_2 := \left(\sum_{m=1}^d |x_m|^2 \right)^{1/2} \qquad x \in \mathbb{R}^d,$$

is continuously Gâteaux differentiable with derivative

$$j'(v) = -\alpha \nabla \cdot \max(0, \nabla v) + \beta \max(0, v) \in H^{-1}(\Omega) \qquad \forall v \in H_0^1(\Omega),$$

and that

$$\begin{aligned}
&\lim_{t \searrow 0} \frac{j'(v + tz) - j'(v)}{t} \\
&\quad = \alpha \sum_{m=1}^d -\partial_m \left(\mathbb{1}_{\{\partial_m v = 0\}} \max(0, \partial_m z) + \mathbb{1}_{\{\partial_m v > 0\}} \partial_m z \right) + \beta \left(\mathbb{1}_{\{v=0\}} \max(0, z) + \mathbb{1}_{\{v>0\}} z \right) \\
&\quad = -\alpha \nabla \cdot \max(0, \cdot)'(\nabla v; \nabla z) + \beta \max(0, \cdot)'(v; z)
\end{aligned}$$

holds for all $v, z \in H_0^1(\Omega)$, where $\mathbb{1}_D$ again denotes the indicator function of a set and where the expression $\max(0, \cdot)'(\nabla v; \nabla z)$ is understood as a componentwise superposition. The partial differential equation (4.2) thus fits precisely into the setting of Theorem 4.1.1, and we may deduce that the solution map $S : H^{-1}(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, to (4.2) is Hadamard directionally differentiable with derivatives $\delta = S'(f; g) \in H_0^1(\Omega)$ that satisfy

$$\begin{aligned}
&\langle -\Delta \delta, z - \delta \rangle \\
&+ \int_\Omega \frac{\alpha}{2} \sum_{m=1}^d \left(\mathbb{1}_{\{\partial_m w = 0\}} \max(0, \partial_m z)^2 + \mathbb{1}_{\{\partial_m w > 0\}} (\partial_m z)^2 \right) + \frac{\beta}{2} \left(\mathbb{1}_{\{w=0\}} \max(0, z)^2 + \mathbb{1}_{\{w>0\}} z^2 \right) \\
&- \frac{\alpha}{2} \sum_{m=1}^d \left(\mathbb{1}_{\{\partial_m w = 0\}} \max(0, \partial_m \delta)^2 + \mathbb{1}_{\{\partial_m w > 0\}} (\partial_m \delta)^2 \right) + \frac{\beta}{2} \left(\mathbb{1}_{\{w=0\}} \max(0, \delta)^2 + \mathbb{1}_{\{w>0\}} \delta^2 \right) \mathrm{d}\mathcal{L}^d \\
&\qquad \geq \langle g, z - \delta \rangle \qquad \forall z \in H_0^1(\Omega)
\end{aligned}$$

for all $f, g \in H^{-1}(\Omega)$. If we test the above variational inequality with functions of the form $z = \delta + tu$, $t > 0$, $u \in H_0^1(\Omega)$, divide by t and pass to the limit $t \searrow 0$, then we obtain that δ is uniquely characterized by the PDE

$$-\Delta \delta - \alpha \nabla \cdot \max(0, \cdot)'(\nabla w; \nabla \delta) + \beta \max(0, \cdot)'(w; \delta) = g \in H^{-1}(\Omega). \tag{4.3}$$

This shows that the solution operator $S : f \mapsto w$ behaves precisely as one would expect: The directional derivatives $S'(f; g)$ are characterized by exactly those partial differential equations that are obtained when the terms in (4.2) are differentiated w.r.t. w and f in the directions δ and g, respectively.

We would like to point out that the directional differentiability of the solution map S to (4.2) and the characterization of the derivatives $S'(f; g)$ by (4.3) are not as trivial as one might think at a first glance. To see this, let us suppose for the moment that the results of Chapter 1 are not available and that $\alpha > 0$. In this situation, one would typically try to proceed in the following steps to prove the directional differentiability of the function S (cf. the analysis in [Christof et al., 2017; De los Reyes and Meyer, 2016; Hintermüller and Surowiec, 2017; Meyer and Susu, 2017] and the proof of Proposition 1.3.10):

(i) Prove the Lipschitz continuity of the map $S : H^{-1}(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$.

(ii) Choose a sequence of difference quotients $\{\delta_{t_n}\}$ (as defined in (1.18)) that converges weakly in $H_0^1(\Omega)$ to some δ for $t_n \searrow 0$ (possible due to the boundedness of the family $\{\delta_t\}_{t>0}$).

(iii) Pass to the limit $t_n \searrow 0$ in the PDE for δ_{t_n} and prove that the weak limit of the difference quotients is itself uniquely determined by an elliptic partial differential equation that is independent of $\{t_n\}$. This yields that the function $\delta \in H_0^1(\Omega)$ in (ii) is independent of the choice of subsequences and that S is weakly directionally differentiable.

(iv) Use bootstrapping to obtain convergence in norm and strong directional differentiability.

The problem that arises when the above approach is used for the sensitivity analysis of the PDE (4.2) is that the partial differential equation for δ_t involves the term

$$\alpha \sum_{m=1}^{d} \mathbb{1}_{\{\partial_m w = 0\}} \max(0, \partial_m \delta_t)$$

whose behavior for $t \searrow 0$ is completely unclear (since $z_t \rightharpoonup z$ in $L^2(\Omega)$ for $t \searrow 0$ does not necessarily imply $\max(0, z_t) \rightharpoonup \max(0, z)$ in $L^2(\Omega)$). It is thus not possible to pass to the limit in the PDE for the difference quotients δ_t and the described procedure fails in step (iii). This shows that it makes sense to take the detour via the EVIs of the second kind when studying equations of the type (4.1). Note that the arguments that we have used for the derivation of Theorems 1.2.2 and 4.1.1 are identical to those employed in the (more concise) proof of [Christof et al., 2017, Theorem 2.2].

Before we conclude this section, we would like to mention that PDEs of the type (4.2) can also be encountered in practical applications. For $\alpha = 0$ and $\beta = 1$, (4.2) characterizes, e.g., the deflection of a membrane that is partially covered by water and (with inhomogeneous Dirichlet boundary conditions) the shape at equilibrium of a confined plasma, see [Kikuchi et al., 1984; Rappaz, 1984; Temam, 1975]. Compare also with the problems studied in [Scarpa, 2017] in this context.

4.2 EVIs Involving Seminorms

In this section, we begin our study of elliptic variational inequalities involving seminorms by proving the second-order epi-differentiability of functions of the form

$$j : V \to \mathbb{R}, \qquad j(v) := \sum_{m=1}^{M} k_m(|v|_m). \tag{4.4}$$

Our precise assumptions on the quantities in (4.4) are as follows:

Assumption 4.2.1.

- *V is a Hilbert space and $M \in \mathbb{N}$,*
- *$k_m \in C^1([0,\infty)) \cap C^2((0,\infty))$, $m = 1, ..., M$, are convex, non-decreasing functions such that the limit*

  $$\lim_{t \searrow 0} k_m''(t) \in [0, \infty]$$

  exists for all m (in what follows, we denote this limit with $k_m''(0) \in [0,\infty]$),
- *$|\cdot|_m$, $m = 1, ..., M$, are seminorms that are induced by continuous positive semidefinite symmetric bilinear forms $b_m : V \times V \to \mathbb{R}$, i.e., $|v|_m := b_m(v, v)^{1/2}$ for all $v \in V$.*

Note that the function j in (4.4) is indeed convex since the map $v \mapsto |v|_m$ is convex for all m by the Cauchy-Schwarz inequality and since the maps $k_m : [0,\infty) \to \mathbb{R}$ are convex and non-decreasing. Further, j is trivially proper and lower semicontinuous. We are thus again in the setting of Chapter 1. To prove the second-order epi-differentiability of j, we need the following lemma that provides a special version of the Taylor-like expansion (2.13) used in the proof of Theorem 2.3.1.

Lemma 4.2.2. *Let $b : V \times V \to \mathbb{R}$ be a continuous positive semidefinite symmetric bilinear form. Denote the seminorm induced by b with $|\cdot|_b$, assume that a $v \in V$ with $|v|_b > 0$ is given, and suppose that $k \in C^1([0,\infty)) \cap C^2((0,\infty))$. Then, the function $k(|\cdot|_b)$ is Fréchet differentiable in v with derivative*

$$z \mapsto k'(|v|_b)|v|_b^{-1} b(v,z),$$

and for all $t > 0$ and all $z \in V$ with $|v+tz|_b > 0$ it holds

$$\begin{aligned}
&\frac{2}{t}\left(\frac{k(|v+tz|_b) - k(|v|_b)}{t} - k'(|v|_b)|v|_b^{-1} b(v,z)\right) \\
&\quad = k'(|v|_b)\left(4\frac{|v|_b^2|z|_b^2 - b(v,z)^2}{(|v+tz|_b + |v|_b)^2|v|_b} + 2\frac{|v|_b(|v+tz|_b - |v|_b) - tb(v,z)}{(|v+tz|_b+|v|_b)^2|v|_b}|z|_b^2\right) \\
&\qquad + \left(\frac{8b(v,z)^2}{(|v|_b + |v+tz|_b)^2} + \frac{8tb(v,z) + 2t^2|z|_b^2}{(|v|_b+|v+tz|_b)^2}|z|_b^2\right) \\
&\qquad\quad \cdot \int_0^1 (1-s)k''\Big((1-s)|v|_b + s|v+tz|_b\Big)\mathrm{d}s.
\end{aligned} \tag{4.5}$$

Proof. The Fréchet differentiability of the function $k(|\cdot|_b)$ is a trivial consequence of the chain rule. To prove (4.5), we note that a Taylor expansion analogous to (2.13) yields

$$\begin{aligned}
&\frac{2}{t}\left(\frac{k(|v+tz|_b) - k(|v|_b)}{t} - k'(|v|_b)|v|_b^{-1} b(v,z)\right) \\
&\quad = k'(|v|_b)\frac{2}{t}\left(\frac{|v+tz|_b - |v|_b}{t} - |v|_b^{-1} b(v,z)\right) \\
&\qquad + \frac{2}{t^2}(|v+tz|_b - |v|_b)^2 \int_0^1 (1-s)k''\Big((1-s)|v|_b + s|v+tz|_b\Big)\mathrm{d}s.
\end{aligned}$$

Using the binomial identities, we may further compute that

$$\begin{aligned}
&\frac{1}{t}\left(\frac{|v+tz|_b - |v|_b}{t} - |v|_b^{-1} b(v,z)\right) \\
&\quad = \frac{1}{t}\left(\frac{2b(v,z) + t|z|_b^2}{|v+tz|_b + |v|_b} - \frac{b(v,z)}{|v|_b}\right) \\
&\quad = \frac{1}{t}\left(\frac{b(v,z)(|v|_b - |v+tz|_b) + t|v|_b|z|_b^2}{(|v+tz|_b+|v|_b)|v|_b}\right) \\
&\quad = \frac{1}{t}\left(\frac{b(v,z)(|v|_b^2 - |v+tz|_b^2) + t|v|_b|z|_b^2(|v+tz|_b+|v|_b)}{(|v+tz|_b+|v|_b)^2|v|_b}\right) \\
&\quad = \frac{b(v,z)(-2b(v,z) - t|z|_b^2) + |v|_b|v+tz|_b|z|_b^2 + |v|_b^2|z|_b^2}{(|v+tz|_b+|v|_b)^2|v|_b} \\
&\quad = \frac{2|v|_b^2|z|_b^2 - 2b(v,z)^2 - tb(v,z)|z|_b^2 + |v|_b|v+tz|_b|z|_b^2 - |v|_b^2|z|_b^2}{(|v+tz|_b+|v|_b)^2|v|_b} \\
&\quad = 2\frac{|v|_b^2|z|_b^2 - b(v,z)^2}{(|v+tz|_b+|v|_b)^2|v|_b} + \frac{|v|_b(|v+tz|_b - |v|_b) - tb(v,z)}{(|v+tz|_b+|v|_b)^2|v|_b}|z|_b^2
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{1}{t^2}(|v+tz|_b-|v|_b)^2\\
&\quad=\frac{1}{t^2}\left(2|v|_b^2+2tb(v,z)+t^2|z|_b^2-2|v+tz|_b|v|_b\right)\\
&\quad=\frac{1}{t^2}\left(2|v|_b(|v|_b-|v+tz|_b)+2tb(v,z)+t^2|z|_b^2\right)\\
&\quad=\frac{1}{t}\left(\frac{2|v|_b(-2b(v,z)-t|z|_b^2)+2b(v,z)(|v|_b+|v+tz|_b)+t|z|_b^2(|v|_b+|v+tz|_b)}{|v|_b+|v+tz|_b}\right)\\
&\quad=\frac{1}{t}\left(\frac{|v|_b(-2b(v,z)-t|z|_b^2)+2b(v,z)|v+tz|_b+t|z|_b^2|v+tz|_b}{|v|_b+|v+tz|_b}\right)\\
&\quad=\frac{1}{t}\left(\frac{2b(v,z)(|v+tz|_b-|v|_b)+t|z|_b^2(|v+tz|_b-|v|_b)}{|v|_b+|v+tz|_b}\right)\\
&\quad=\frac{2b(v,z)(2b(v,z)+t|z|_b^2)+|z|_b^2(2tb(v,z)+t^2|z|_b^2)}{(|v|_b+|v+tz|_b)^2}\\
&\quad=\frac{4b(v,z)^2}{(|v|_b+|v+tz|_b)^2}+\frac{4tb(v,z)+t^2|z|_b^2}{(|v|_b+|v+tz|_b)^2}|z|_b^2.
\end{aligned}
$$

If we combine the last three identities, then we arrive at (4.5). This completes the proof. □

Although technical, Lemma 4.2.2 is an important tool in the study of non-differentiabilities of the form (4.4) (in particular, when these functions appear as Nemytskii operators). It gives precise control over the behavior of the second-order difference quotients on the left-hand side of (4.5) and isolates the different curvature effects that occur when seminorms are present. Consider, e.g., the case $k(x) := x$ and $b(\cdot,\cdot) := (\cdot,\cdot)_V$, where (4.5) takes the form

$$
\begin{aligned}
&\frac{2}{t}\left(\frac{\|v+tz\|_V-\|v\|_V}{t}-\frac{(v,z)_V}{\|v\|_V}\right)\\
&\qquad=\left(4\frac{\|v\|_V^2\|z\|_V^2-(v,z)_V^2}{(\|v+tz\|_V+\|v\|_V)^2\|v\|_V}+2\frac{\|v\|_V(\|v+tz\|_V-\|v\|_V)-t(v,z)_V}{(\|v+tz\|_V+\|v\|_V)^2\|v\|_V}\|z\|_V^2\right).
\end{aligned}
\tag{4.6}
$$

In this situation, the first term on the right-hand side of (4.6) tends to the classical derivative

$$
\|\cdot\|_V''(v)z^2=\frac{\|v\|_V^2\|z\|_V^2-(v,z)_V^2}{\|v\|_V^3}
$$

for $t\searrow 0$ and is predominantly influenced by the component of z that is orthogonal to v for small $t>0$, while the second term on the right-hand side of (4.6) constitutes precisely that part of the second-order difference quotient that is - in the Nemytskii setting - relevant for distributional curvature effects in the radial direction, cf. Section 5.2.

As a first consequence of Lemma 4.2.2, we obtain the following result:

Theorem 4.2.3. *Let V, M, k_m, $|\cdot|_m$ and b_m be as in Assumption 4.2.1 and let j be defined as in* (4.4)*. Set $j_m(\cdot) := k_m(|\cdot|_m)$. Then, the convex subdifferential of j is given by*

$$
\partial j(v)=\sum_{m=1}^M \partial j_m(v)=\sum_{m=1}^M k_m'(|v|_m)\partial|v|_m \qquad \forall v\in V, \tag{4.7}
$$

and j is twice epi-differentiable in every $v\in V$ for all $\varphi=\sum_{m=1}^M k_m'(|v|_m)\varphi_m$, $\varphi_m\in\partial|v|_m$, with

$$
Q_j^{v,\varphi}(z)=\sum_{m=1}^M Q_{j_m}^{v,k_m'(|v|_m)\varphi_m}(z) \qquad \forall z\in V, \tag{4.8}
$$

where

$$Q_{j_m}^{v,k'_m(|v|_m)\varphi_m}(z) = k'_m(|v|_m)\ \left(\frac{|v|_m^2|z|_m^2 - b_m(v,z)^2}{|v|_m^3}\right) + k''_m(|v|_m)\frac{b_m(v,z)^2}{|v|_m^2} \qquad \forall z \in V \quad (4.9)$$

for all $v \in V$ with $|v|_m > 0$ and

$$Q_{j_m}^{v,k'_m(0)\varphi_m}(z) = \begin{cases} k''_m(0)|z|_m^2 + \chi_{\{u\in V\,|\,k'_m(0)(|u|_m - \langle\varphi_m,u\rangle)=0\}}(z) & \text{if } k''_m(0) < \infty \\ \chi_{\{u\in V\,||u|_m=0\}}(z) & \text{if } k''_m(0) = \infty \end{cases} \qquad \forall z \in V \quad (4.10)$$

in the case $|v|_m = 0$. Here, $\chi_D : V \to \{0,\infty\}$ again denotes the characteristic function of a set $D \subset V$.

Proof. The formula for the subdifferential $\partial j(v)$ in (4.7) is a trivial consequence of the sum rule and the chain rule for convex subdifferentials (cf. the argumentation in the proof of Theorem 2.3.1). It remains to prove the second-order epi-differentiability. To this end, we note that

$$\frac{2}{t}\left(\frac{j(v+tz)-j(v)}{t} - \langle\varphi, z\rangle\right) = \sum_{m=1}^{M}\frac{2}{t}\left(\frac{k_m(|v+tz|_m) - k_m(|v|_m)}{t} - k'_m(|v|_m)\,\langle\varphi_m, z\rangle\right)$$

holds for all $v, z \in V$, $t > 0$ and $\varphi = k'_1(|v|_1)\varphi_1 + ... + k'_M(|v|_M)\varphi_M$, $\varphi_m \in \partial|v|_m$. Suppose now that v, z and φ_m, $m = 1, ..., M$, are arbitrary but fixed, and let $\{z_n\} \subset V$, $\{t_n\} \subset \mathbb{R}^+$ be sequences with $z_n \rightharpoonup z$ and $t_n \searrow 0$. Then, the boundedness of $\{z_n\}$, the dominated convergence theorem, the properties of the functions k_m, the weak lower semicontinuity of convex lower semicontinuous functions and Lemma 4.2.2 yield that for all m with $|v|_m > 0$, it holds $\langle\varphi_m, \cdot\rangle = |v|_m^{-1}b_m(v,\cdot)$ and

$$\begin{aligned}
&\liminf_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(|v+t_nz_n|_m) - k_m(|v|_m)}{t_n} - k'_m(|v|_m)|v|_m^{-1}b_m(v,z_n)\right)\right] \\
&= \liminf_{n\to\infty}\left[k'_m(|v|_m)\ \left(4\frac{|v|_m^2|z_n|_m^2 - b_m(v,z_n)^2}{(|v+t_nz_n|_m + |v|_m)^2|v|_m}\right)\right. \\
&\qquad + k'_m(|v|_m)\ \left(2\frac{|v|_m(|v+t_nz_n|_m - |v|_m) - t_nb_m(v,z_n)}{(|v+t_nz_n|_m + |v|_m)^2|v|_m}|z_n|_m^2\right) \\
&\qquad + \left(\frac{8b_m(v,z_n)^2}{(|v|_m + |v+t_nz_n|_m)^2} + \frac{8t_nb_m(v,z_n) + 2t_n^2|z_n|_m^2}{(|v|_m + |v+t_nz_n|_m)^2}|z_n|_m^2\right) \\
&\qquad\qquad \left.\cdot\int_0^1 (1-s)k''_m\Big((1-s)|v|_m + s|v+t_nz_n|_m\Big)\mathrm{d}s\right] \\
&\geq \frac{k'_m(|v|_m)}{|v|_m^3}\liminf_{n\to\infty}\Big(|v|_m^2|z_n|_m^2 - b_m(v,z_n)^2\Big) + \frac{k''_m(|v|_m)}{|v|_m^2}\liminf_{n\to\infty}b_m(v,z_n)^2 \\
&\geq k'_m(|v|_m)\ \left(\frac{|v|_m^2|z|_m^2 - b_m(v,z)^2}{|v|_m^3}\right) + k''_m(|v|_m)\frac{b_m(v,z)^2}{|v|_m^2}.
\end{aligned} \quad (4.11)$$

For all m with $|v|_m = 0$ and $|z|_m > 0$, we obtain, on the other hand, (using Taylor's theorem, the lemma of Fatou and the fact that the weak lower semicontinuity of the seminorm $|\cdot|_m$ implies $|z_n|_m > 0$ for all sufficiently large n)

$$\begin{aligned}
&\liminf_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(|v+t_nz_n|_m) - k_m(|v|_m)}{t_n} - k'_m(|v|_m)\,\langle\varphi_m, z_n\rangle\right)\right] \\
&= \liminf_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(t_n|z_n|_m) - k_m(0)}{t_n} - k'_m(0)\,\langle\varphi_m, z_n\rangle\right)\right] \\
&= \liminf_{n\to\infty}\left[2\left(\int_0^1\int_0^1 k''_m(s_1s_2t_n|z_n|_m)\mathrm{d}s_1\,s_2\mathrm{d}s_2\right)|z_n|_m^2 + \frac{2}{t_n}\Big(k'_m(0)|z_n|_m - k'_m(0)\,\langle\varphi_m, z_n\rangle\Big)\right] \\
&\geq k''_m(0)|z|_m^2 + 2\liminf_{n\to\infty}\left[\frac{1}{t_n}\Big(k'_m(0)|z_n|_m - k'_m(0)\,\langle\varphi_m, z_n\rangle\Big)\right],
\end{aligned}$$

where the weak lower semicontinuity of $|\cdot|_m$ and the subgradient property of φ_m imply

$$\liminf_{n\to\infty}\Big(k_m'(0)|z_n|_m - k_m'(0)\,\langle\varphi_m, z_n\rangle\Big) \geq k_m'(0)\,(|z|_m - \langle\varphi_m, z\rangle) \geq 0.$$

Lastly, for all m with $|v|_m = 0$ and $|z|_m = 0$, we trivially have

$$\liminf_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(|v+t_nz_n|_m) - k_m(|v|_m)}{t_n} - k_m'(|v|_m)\,\langle\varphi_m, z_n\rangle\right)\right] \geq 0$$

and

$$k_m'(0)\,(|z|_m - \langle\varphi_m, z\rangle) = k_m'(0)\,\langle\varphi_m, z\rangle = 0$$

by the subgradient property of φ_m. Combining all of the above, we obtain

$$\begin{aligned} Q_j^{v,\varphi}(z) \geq &\sum_{m:|v|_m>0} k_m'(|v|_m)\,\left(\frac{|v|_m^2|z|_m^2 - b_m(v,z)^2}{|v|_m^3}\right) + k_m''(|v|_m)\frac{b_m(v,z)^2}{|v|_m^2} \\ &+ \sum_{m:|v|_m=0, k_m''(0)<\infty} k_m''(0)|z|_m^2 + \chi_{\{u\in V\,|\,k_m'(0)(|u|_m - \langle\varphi_m,u\rangle)=0\}}(z) \\ &+ \sum_{m:|v|_m=0, k_m''(0)=\infty} \chi_{\{u\in V\,|\,|u|_m=0\}}(z). \end{aligned} \tag{4.12}$$

In particular,

$$\begin{aligned} \mathcal{K}_j^{red}(v,\varphi) \subset \Big\{ z\in V \;\Big|\; &|z|_m = 0 \text{ for all } m \text{ with } |v|_m = 0, k_m''(0) = \infty \text{ and} \\ &k_m'(0)(|z|_m - \langle\varphi_m, z\rangle) = 0 \text{ for all } m \text{ with } |v|_m = 0, k_m''(0) < \infty\Big\}. \end{aligned} \tag{4.13}$$

Consider now an arbitrary but fixed z that is contained in the set on the right-hand side of (4.13), and assume that a sequence $\{t_n\} \subset \mathbb{R}^+$with $t_n \searrow 0$ is given. Then, for all m with $|v|_m > 0$, we obtain (by the same arguments as in (4.11))

$$\begin{aligned} &\lim_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(|v+t_nz|_m) - k_m(|v|_m)}{t_n} - k_m'(|v|_m)\,\langle\varphi_m, z\rangle\right)\right] \\ &\quad = k_m'(|v|_m)\,\left(\frac{|v|_m^2|z|_m^2 - b_m(v,z)^2}{|v|_m^3}\right) + k_m''(|v|_m)\frac{b_m(v,z)^2}{|v|_m^2}. \end{aligned}$$

Moreover, for every m with $|v|_m = 0$ and $k_m''(0) < \infty$, we may compute (using the properties of z and Taylor's formula)

$$\begin{aligned} &\lim_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(|v+t_nz|_m) - k_m(|v|_m)}{t_n} - k_m'(|v|_m)\,\langle\varphi_m, z\rangle\right)\right] \\ &\quad = \lim_{n\to\infty}\left[\frac{2}{t_n}\left(\frac{k_m(t_n|z|_m) - k_m(0)}{t_n} - k_m'(0)|z|_m\right)\right] \\ &\quad = k_m''(0)|z|_m^2, \end{aligned}$$

and for every m with $|v|_m = 0$ and $k_m''(0) = \infty$, it follows from $|z|_m = 0$ that

$$\begin{aligned} &\frac{2}{t_n}\left(\frac{k_m(|v+t_nz|_m) - k_m(|v|_m)}{t_n} - k_m'(|v|_m)\,\langle\varphi_m, z\rangle\right) \\ &\quad = \frac{2}{t_n}\left(\frac{k_m(t_n|z|_m) - k_m(0)}{t_n} - k_m'(0)|z|_m\right) \\ &\quad = 0 \qquad \forall n\in\mathbb{N}. \end{aligned}$$

This shows that equality holds in (4.12) and (4.13), that j is indeed twice epi-differentiable (with the recovery sequence $z_n = z$ for every given $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$), and that the second subderivative of j behaves as claimed. Note that the above considerations immediately yield that the addends on the right-hand side of (4.12) are precisely the second subderivatives $Q_{j_m}^{v,k'_m(|v|_m)\varphi_m}(\cdot)$. □

We point out that the last theorem cannot be obtained with the concepts of (extended) polyhedricity and second-order regularity (using, e.g., Proposition 2.4.6) because the function j in (4.4) possesses both non-zero curvature and non-differentiabilities.

4.2.1 Pointwise Maxima of Smooth Functions

To develop intuition for the formulas (4.8), (4.9) and (4.10), in what follows, we explore some of the consequences of Theorem 4.2.3. We begin with a simple result on the second-order epi-differentiability of pointwise maxima:

Corollary 4.2.4. *Let V be a Hilbert space and suppose that $j_1, j_2 : V \to \mathbb{R}$ are two convex C^2-functions. Assume that a $v \in V$ is given such that the Fréchet derivatives $j_1'(v), j_2'(v) \in V^*$ are linearly independent and such that the map $V \ni z \mapsto (j_1''(v)z^2, j_2''(v)z^2) \in \mathbb{R}^2$ is completely continuous. Then, the function $j := \max(j_1, j_2) : V \to \mathbb{R}$ is convex and continuous, it holds*

$$\partial j(v) = j_2'(v) + (j_1'(v) - j_2'(v))\partial \max(0,\cdot)(j_1(v) - j_2(v)),$$

and j is twice epi-differentiable in v for all $\varphi = \lambda j_1'(v) + (1-\lambda)j_2'(v)$, $\lambda \in \partial \max(0,\cdot)(j_1(v) - j_2(v))$, with

$$Q_j^{v,\varphi}(z) = \begin{cases} j_1''(v)z^2 & \text{if } j_1(v) > j_2(v) \\ j_2''(v)z^2 & \text{if } j_1(v) < j_2(v) \\ \lambda j_1''(v)z^2 + (1-\lambda)j_2''(v)z^2 + \chi_{\{u\in\mathbb{R}\,|\,\max(0,u)=\lambda u\}}(j_1'(v)z - j_2'(v)z) & \text{if } j_1(v) = j_2(v) \end{cases} \tag{4.14}$$

for all $z \in V$.

Proof. The convexity and the continuity of j are trivial. To prove the second-order epi-differentiability, we proceed in several steps using the chain and sum rules of Chapter 2: First of all, we know from Theorem 4.2.3 that the absolute value function $\mathbb{R} \ni x \mapsto |x| \in \mathbb{R}$ is twice epi-differentiable in all $x \in \mathbb{R}$ for all $\eta \in \partial|x|$ with

$$Q_{|\cdot|}^{x,\eta}(z) = \begin{cases} 0 & \text{if } x \neq 0 \\ \chi_{\{u\in\mathbb{R}\,|\,|u|=\eta u\}}(z) & \text{if } x = 0 \end{cases} \qquad \forall z \in \mathbb{R}. \tag{4.15}$$

The above and Theorem 2.2.1 yield that the max-function $x \mapsto \max(x) := \max(0,x) = (|x| + x)/2$ is twice epi-differentiable in all $x \in \mathbb{R}$ for all $\lambda \in \partial \max(x)$ with

$$Q_{\max}^{x,\lambda}(z) = \frac{1}{2}Q_{|\cdot|}^{x,2\lambda-1}(z) = \begin{cases} 0 & \text{if } x \neq 0 \\ \chi_{\{u\in\mathbb{R}\,|\,\max(u)=\lambda u\}}(z) & \text{if } x = 0 \end{cases} \qquad \forall z \in \mathbb{R}. \tag{4.16}$$

Consider now the map $F_1 : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto x_1 - x_2$. Then, F_1 is clearly linear and surjective, and we may use Proposition 2.4.3 and Theorem 2.4.8 to obtain that the subdifferential of the function $k_1 := \max \circ F_1 : \mathbb{R}^2 \to \mathbb{R}$ satisfies

$$\partial k_1(x_1, x_2) = (1, -1)\partial \max(x_1 - x_2) \qquad \forall (x_1, x_2) \in \mathbb{R}^2,$$

and that k_1 is twice epi-differentiable in all (x_1, x_2) for all $\nu = (1, -1)\lambda$, $\lambda \in \partial \max(x_1 - x_2)$, with

$$Q_{k_1}^{(x_1,x_2),\nu}(z_1, z_2) = Q_{\max}^{x_1-x_2,\lambda}(z_1 - z_2) \qquad \forall (z_1, z_2) \in \mathbb{R}^2.$$

Defining $k_2(x_1, x_2) := x_2 + \max(0, x_1 - x_2)$ and using again Theorem 2.2.1, we obtain in the next step that $\partial k_2(x_1, x_2) = (0, 1) + (1, -1)\partial \max(x_1 - x_2)$ holds for all $(x_1, x_2) \in \mathbb{R}^2$, and that k_2 is twice epi-differentiable in all (x_1, x_2) for all $\zeta = (0, 1) + (1, -1)\lambda$, $\lambda \in \partial \max(x_1 - x_2)$, with

$$Q_{k_2}^{(x_1,x_2),\zeta}(z_1, z_2) = Q_{\max}^{x_1-x_2,\lambda}(z_1 - z_2) \qquad \forall (z_1, z_2) \in \mathbb{R}^2.$$

If we now finally define $F_2 : V \to \mathbb{R}^2$, $z \mapsto (j_1(z), j_2(z))$, then it holds $j = k_2 \circ F_2$, and we know from our assumptions that $F_2'(v) = (j_1'(v), j_2'(v)) : V \to \mathbb{R}^2$ is surjective. This allows us to again employ Proposition 2.4.3 and Theorem 2.4.8, to deduce that

$$\begin{aligned}\partial j(v) &= F_2'(v)^* \partial k_2(F_2(v)) \\ &= F_2'(v)^* \Big((0, 1) + (1, -1)\partial \max(j_1(v) - j_2(v)) \Big) \\ &= j_2'(v) + (j_1'(v) - j_2'(v))\partial \max(j_1(v) - j_2(v))\end{aligned}$$

holds, and to obtain that j is twice epi-differentiable in v for all subgradients $\varphi = \lambda j_1'(v) + (1 - \lambda) j_2'(v)$, $\lambda \in \partial \max(j_1(v) - j_2(v))$, with

$$\begin{aligned}Q_j^{v,\varphi}(z) &= Q_{k_2}^{F_2(v),(0,1)+(1,-1)\lambda}(F_2'(v)z) + \big\langle (0, 1) + (1, -1)\lambda, F_2''(v)z^2 \big\rangle \\ &= Q_{\max}^{j_1(v)-j_2(v),\lambda}(j_1'(v)z - j_2'(v)z) + \lambda j_1''(v)z^2 + (1 - \lambda) j_2''(v)z^2\end{aligned}$$

for all $z \in V$. If we plug in the formula for the second subderivative of the $\max$-function on the right-hand side of the last estimate, then the claim follows immediately. □

Remark 4.2.5.

(i) A finite-dimensional version of Corollary 4.2.4 can be found in [Rockafellar, 1990, Corollary 3.5]. In this paper, the author proved the second-order epi-differentiability of the function $(x_1, x_2) \mapsto \max(x_1, x_2)$ *directly (using the concept of polyhedrality) and subsequently argued with a variant of Theorem 2.4.8. The advantage of this approach is that it also works when the maximum of several functions* j_m *is considered. Note that we could have used a similar line of reasoning here by employing Proposition 2.4.6, by proving the polyhedricity of the epigraph of* $(x_1, ..., x_d) \mapsto \max(x_1, x_2, ..., x_d)$ *and by subsequently invoking Proposition 3.3.3 and Theorem 2.4.8. We leave it to the interested reader to work out the details of this alternative proof.*

(ii) Note that the domains of the characteristic functions on the right-hand sides of (4.14)*,* (4.15) *and* (4.16) *are precisely the critical cones of* $|\cdot|$ *and* $\max(\cdot)$ *at the origin, cf. Lemma 1.3.4.*

(iii) As Corollary 4.2.4 demonstrates, Theorem 4.2.3 may serve as a point of departure for the study of more complicated non-smooth functionals when we combine it with the calculus rules of Chapter 2. Compare also with the analysis in Section 4.3 in this context.

4.2.2 Regularization by Singular Curvature

An important observation, following from Theorem 4.2.3, is that the blowup of the second derivative k_m'' of one of the functions k_m in (4.4) at zero can compensate the non-smooth behavior that the remaining addends $k_m(|\cdot|_m)$ with $k_m''(0) < \infty$ may exhibit at the origin. Consider, for example, the function

$$j(v) := \Big(\|v\|_V^{\beta_1} + \|v\|_V^{\beta_2} \Big), \tag{4.17}$$

where V is still assumed to be a Hilbert space and where $\beta_m \geq 1$, $m = 1, 2$. Then, Theorem 4.2.3 yields:

Corollary 4.2.6. *The function $j : V \to \mathbb{R}$ in* (4.17) *is twice epi-differentiable everywhere and for all $v \in V$ and all $\varphi = \varphi_1 + \varphi_2 \in \partial\|v\|_V^{\beta_1} + \partial\|v\|_V^{\beta_2} = \partial j(v)$, we have $Q_j^{v,\varphi} = Q^{v,\varphi_1}_{\|\cdot\|_V^{\beta_1}} + Q^{v,\varphi_2}_{\|\cdot\|_V^{\beta_2}}$ with*

$$Q^{v,\varphi_m}_{\|\cdot\|_V^{\beta_m}}(z) = \begin{cases} \beta_m \dfrac{\|v\|_V^2\|z\|_V^2 - (v,z)_V^2}{\|v\|_V^{4-\beta_m}} + \beta_m(\beta_m - 1)\dfrac{(v,z)_V^2}{\|v\|_V^{4-\beta_m}} & \text{if } v \neq 0,\ \beta_m \geq 1 \\ \chi_{\{u \in V \,|\, \|u\|_V = \langle \varphi_m, u\rangle\}}(z) & \text{if } v = 0,\ \beta_m = 1 \\ \chi_{\{0\}}(z) & \text{if } v = 0,\ \beta_m \in (1,2) \\ 2\|z\|_V^2 & \text{if } v = 0,\ \beta_m = 2 \\ 0 & \text{if } v = 0,\ \beta_m \in (2,\infty) \end{cases}.$$

Proof. Choose $M := 2$, $k_m(x) := x^{\beta_m}$ and $b_m(\cdot,\cdot) := (\cdot,\cdot)_V$, $m = 1, 2$, in Theorem 4.2.3, then the claim follows immediately. □

Since $\chi_{D_1} + \chi_{D_2} = \chi_{D_1}$ for all $D_1 \subset D_2 \subset V$, Corollary 4.2.6 implies that j satisfies

$$Q_j^{0,\varphi} = Q^{0,\varphi_1}_{\|\cdot\|_V^{\beta_1}} + Q^{0,\varphi_2}_{\|\cdot\|_V^{\beta_2}} = \chi_{\{0\}}$$

whenever one of the exponents β_m is contained in the interval $(1, 2)$. This shows that the characteristic function of the critical cone $\{u \in V \mid \|u\|_V = \langle \varphi_m, u\rangle\}$ that occurs in the case $\beta_m = 1$ and that would normally prevent the second subderivative from being quadratic and the reduced critical cone from being a subspace is removed from the functional $Q_j^{0,\varphi}$ when a term with singular curvature is present. In particular, it holds

$$Q_j^{v,\varphi}(z) = q_j^{v,\varphi}(z,z) \quad \forall z \in \mathcal{K}_j^{red}(v,\varphi) \quad \forall \varphi \in \partial j(v) \quad \forall v \in V \tag{4.18}$$

for all $(\beta_1, \beta_2) \in [1,\infty) \times [1,\infty) \setminus Z$, $Z := \{(1,1)\} \cup \{1\} \times [2,\infty) \cup [2,\infty) \times \{1\}$, with:

(i) $\mathcal{K}_j^{red}(v,\varphi) = V$ and

$$q_j^{v,\varphi}(z_1,z_2) = \sum_{m=1}^{2} \beta_m \frac{\|v\|_V^2(z_1,z_2)_V - (v,z_1)_V(v,z_2)_V}{\|v\|_V^{4-\beta_m}} + \beta_m(\beta_m-1)\frac{(v,z_1)_V(v,z_2)_V}{\|v\|_V^{4-\beta_m}}$$

for all $v \neq 0$, $\varphi \in \partial j(v)$ and $(\beta_1,\beta_2) \in [1,\infty) \times [1,\infty) \setminus Z$,

(ii) $\mathcal{K}_j^{red}(0,\varphi) = \{0\}$ and $q_j^{0,\varphi} \equiv 0$ for all $\varphi \in \partial j(0)$ and $(\beta_1,\beta_2) \in [1,\infty) \times (1,2) \cup (1,2) \times [1,\infty)$,

(iii) $\mathcal{K}_j^{red}(0,\varphi) = V$ and

$$q_j^{0,\varphi}(z_1,z_2) = \sum_{m:\beta_m=2} 2(z_1,z_2)_V$$

for all $\varphi \in \partial j(0)$ and $(\beta_1,\beta_2) \in [2,\infty) \times [2,\infty)$.

If we combine the above findings with Corollary 1.4.4, then we obtain:

Corollary 4.2.7. *Suppose that V is a Hilbert space and that $A : V \to V^*$ is a Fréchet differentiable and strongly monotone operator that maps bounded subsets of V into bounded subsets of V^*. Then, the solution map $S : V^* \to V$, $f \mapsto w$, associated with the elliptic variational inequality*

$$w \in V, \quad \langle A(w), v - w\rangle + \left(\|v\|_V^{\beta_1} + \|v\|_V^{\beta_2}\right) - \left(\|w\|_V^{\beta_1} + \|w\|_V^{\beta_2}\right) \geq \langle f, v - w\rangle \quad \forall v \in V \tag{4.19}$$

is well-defined and Hadamard directionally differentiable for all tuples $(\beta_1,\beta_2) \in [1,\infty) \times [1,\infty)$, and Hadamard-Gâteaux differentiable for precisely those $(\beta_1,\beta_2) \in [1,\infty) \times [1,\infty)$ that are not contained in the set $\{(1,1)\} \cup \{1\} \times [2,\infty) \cup [2,\infty) \times \{1\}$.

Proof. The unique solvability of (4.19) and the Hadamard directional differentiability of the solution map S are direct consequences of Theorem 1.2.2, Theorem 1.4.1 and Corollary 4.2.6, and the assertion on the Hadamard-Gâteaux differentiability follows straightforwardly from Corollaries 1.4.4 and 4.2.6 and our observation (4.18). □

The last result is remarkable for several reasons:

Remark 4.2.8.

(i) Corollary 4.2.7 demonstrates that EVIs and minimization problems involving non-differentiable terms can still possess solution operators that are (Hadamard-)Gâteaux differentiable everywhere. If we consider, e.g., the problem

$$w \in V, \quad \langle A(w), v - w\rangle + \Big(\|v\|_V + \|v\|_V^\beta\Big) - \Big(\|w\|_V + \|w\|_V^\beta\Big) \geq \langle f, v - w\rangle \quad \forall v \in V$$

with V and A as before and an exponent $\beta \geq 1$, then Corollary 4.2.7 yields that the solution map $S : f \mapsto w$ is Hadamard-Gâteaux differentiable for precisely those β with $\beta \notin \{1\} \cup [2, \infty)$. Note that the structure of the latter set (e.g., its disconnectedness) is rather counterintuitive, and that it is rather surprising that, from the viewpoint of sensitivity analysis, the function $\|v\|_V + \|v\|_V^{3/2}$ is better behaved than the map $\|v\|_V + \|v\|_V^2$.

(ii) As we will see in Section 4.3.3, the regularization effect explored above readily carries over to situations where functions of the form (4.4) *appear as superposition operators. In this context, it can be exploited, e.g., when discretized problems from fluid dynamics are considered, cf. Section 5.1.6.*

4.3 EVIs Involving Seminorms as Superposition Operators

We now turn our attention to EVIs of the second kind whose functionals j contain terms of the form (4.4) as superposition operators. To be more precise, in what follows, we are interested in the case

$$j : V \to (-\infty, \infty], \qquad j(v) := \int_\Omega \sum_{m=1}^M k_m(|Gv|_m)\mathrm{d}\mu, \tag{4.20}$$

where k_m, G, μ etc. satisfy:

Assumption 4.3.1.

- *V and H are Hilbert spaces and H is separable,*
- *(Ω, Σ, μ) is a finite and complete measure space,*
- *$L^2(\Omega, H)$ is (again) defined as in [Heinonen et al., 2015, Section 3.2]),*
- *$G \in L(V, L^2(\Omega, H))$ (for simplicity),*
- *$M \in \mathbb{N}$,*
- *$k_m \in C^1([0, \infty)) \cap C^2((0, \infty))$, $|\cdot|_m : H \to \mathbb{R}$ and $b_m : H \times H \to \mathbb{R}$, $m = 1, ..., M$, satisfy the conditions in Assumption 4.2.1 (with the same convention for $k_m''(0)$).*

Note that functions of the type (4.20) appear very frequently in variational inequalities that describe real-world phenomena. They emerge, for example, when non-Newtonian fluids are studied, see [Fuchs and Seregin, 1998; Huilgol and You, 2005; Mosolov and Miasnikov, 1965, 1967], when the deformation of elastoplastic materials is considered, see [De los Reyes et al., 2016; Han and Reddy, 1999], in the

context of superconductivity, see [Yousept, 2017], and in glaciology, see [Lindqvist, 1987]. We will discuss some of these applications in more detail in Sections 4.3.2 and 5.1.

When studying the second-order epi-differentiability of functions of the form (4.20), we have to distinguish between two different situations: The first one is that where the operator $G : V \to L^2(\Omega, H)$ is surjective. In this case, it suffices to apply the superposition result of Section 2.5 and the chain rule in Theorem 2.4.8 to prove that the composition (4.20) is twice epi-differentiable, see Theorem 4.3.3 below. The second case is that where G is non-surjective. In this situation (which appears, e.g., when G is the weak gradient ∇) the constraint qualifications in Section 2.4 are typically violated and we have to argue with Lemma 1.3.13 to be able to invoke Theorem 1.4.1, see Section 4.3.4.

Before we delve into the analysis of the above two scenarios, we prove:

Proposition 4.3.2. *Assume that one of the following conditions is satisfied:*

(i) The functions

$$L^2(\Omega, H) \ni h \mapsto \int_\Omega k_m(|h|_m)\mathrm{d}\mu \tag{4.21}$$

are real-valued and continuous for all $m = 1, ..., M$.

(ii) $k_m'(0) = 0$ for all $m = 2, ..., M$ and G is surjective.

Then, for every $v \in V$, it holds

$$\partial j(v) = G^*\left\{ \lambda \in L^2(\Omega, H^*) \;\middle|\; \exists \lambda_m \in L^2(\Omega, H^*) \text{ s.t. } \lambda_m \in \partial|\cdot|_m(Gv) \ \mu\text{-a.e.}, \lambda = \sum_{m=1}^M k_m'(|Gv|_m)\lambda_m \right\}. \tag{4.22}$$

Proof. We first consider the case (i): If the functions in (4.21) are continuous, then we may employ the sum rule and the chain rule for convex subdifferentials to deduce that

$$\partial j(v) = \sum_{m=1}^M G^* \partial \left(\int_\Omega k_m(|\cdot|_m)\mathrm{d}\mu \right)(Gv) \qquad \forall v \in V,$$

where the functions in the brackets are understood as maps from $L^2(\Omega, H)$ to $\mathbb{R}$. From Lemma 2.5.4, we obtain further that

$$\eta \in \partial \left(\int_\Omega k_m(|\cdot|_m)\mathrm{d}\mu \right)(h) \quad \iff \quad \eta \in L^2(\Omega, H^*) \quad \text{and} \quad \eta \in \partial k_m(|\cdot|_m)(h) \ \mu\text{-a.e. in } \Omega$$

holds for all $h \in L^2(\Omega, H)$ and all $m = 1, ..., M$. If we combine the above with (4.7), then the claim follows immediately. Note that, in case (ii), the assumptions of the classical sum rule for the convex subdifferential are typically not satisfied on the V-level. As a consequence, we have to argue more carefully to prove the second part of the proposition: First, we observe that, if $\varphi = G^*\lambda$ is contained in the set on the right-hand side of (4.22), then it holds

$$\begin{aligned} \langle \lambda, Gu - Gv \rangle_H &= \sum_{m=1}^M \langle k_m'(|Gv|_m)\lambda_m, Gu - Gv \rangle_H \\ &\leq \sum_{m=1}^M k_m(|Gu|_m) - k_m(|Gv|_m) \qquad \mu\text{-a.e. in } \Omega \quad \forall u \in V. \end{aligned}$$

If we integrate the above (using that $\lambda \in L^2(\Omega, H^*)$), then we obtain

$$\langle \varphi, u - v \rangle_V = \int_\Omega \langle \lambda, Gu - Gv \rangle_H \,\mathrm{d}\mu \leq j(u) - j(v) \qquad \forall u \in \mathrm{dom}(j).$$

This implies that $v \in \mathrm{dom}(j)$, that $\varphi \in \partial j(v)$ and that "$\supset$" holds in (4.22). If, conversely, we start with a $\varphi \in \partial j(v)$, then the surjectivity of G and the chain rule yield

$$\varphi \in G^* \partial \left(\int_\Omega \sum_{m=1}^M k_m(|\cdot|_m) \mathrm{d}\mu \right) (Gv),$$

and we obtain from Lemma 2.5.4 and the sum rule (on the H-level) that

$$\begin{aligned} &\lambda \in \partial \left(\int_\Omega \sum_{m=1}^M k_m(|\cdot|_m) \mathrm{d}\mu \right) (Gv) \\ &\iff \quad \lambda \in L^2(\Omega, H^*) \quad \text{and} \quad \lambda \in \sum_{m=1}^M \partial k_m(|\cdot|_m)(Gv) \ \mu\text{-a.e. in } \Omega. \end{aligned} \tag{4.23}$$

Consider now an arbitrary but fixed λ as in (4.23) satisfying $\varphi = G^*\lambda$ and define

$$\lambda_m := \begin{cases} |\cdot|_m'(Gv) & \mu\text{-a.e. where } |Gv|_m \neq 0 \\ 0 & \mu\text{-a.e. where } |Gv|_m = 0 \end{cases}, \qquad m = 2, ..., M,$$

$$\lambda_1 := \begin{cases} |\cdot|_1'(Gv) & \mu\text{-a.e. where } |Gv|_1 \neq 0 \\ 0 & \mu\text{-a.e. where } |Gv|_1 = 0 \text{ if } k_1'(0) = 0 \\ \dfrac{1}{k_1'(0)} \left(\lambda - \displaystyle\sum_{m=2}^M k_m'(|Gv|_m)\lambda_m \right) & \mu\text{-a.e. where } |Gv|_1 = 0 \text{ if } k_1'(0) > 0 \end{cases}.$$

Then, the λ_m are obviously in $L^0(\Omega, H^*)$, and the smoothness properties of the functions k_m, (4.23), and the fact that the case $k_m'(0) > 0$ can appear at most for $m = 1$ imply that $\lambda = \sum_{m=1}^M k_m'(|Gv|_m)\lambda_m$ holds μ-a.e. in Ω. Note that, since the seminorms $|\cdot|_m$ are Lipschitz and since $\lambda_m \in L^0(\Omega, H^*)$, we trivially have $\lambda_m \in L^2(\Omega, H^*)$. This proves (4.22) in the second case. □

We would like to point out that the main problem in the second case of Proposition 4.3.2 is to obtain the L^0-regularity of the multiplier λ_1, and that the condition $\lambda \in L^2(\Omega, H^*)$ in (4.22) is necessary and cannot be dropped in general (cf. the example in Section 4.3.3).

4.3.1 Second-Order Epi-Differentiability in the Presence of Surjectivity

As already mentioned, in the case that G is surjective, we may apply the results of Chapter 2 to prove the second-order epi-differentiability of the function j in (4.20). This leads to:

Theorem 4.3.3. *Consider the situation in Assumption 4.3.1 and let j be defined as in* (4.20)*. Suppose that G is surjective and that one of the two conditions in Proposition 4.3.2 is satisfied. Then,* (4.22) *holds for all $v \in V$ and j is twice epi-differentiable in all $v \in V$ for all $\varphi \in \partial j(v)$ with*

$$\begin{aligned} &\mathcal{K}_j^{red}(v, \varphi) \\ &= \Bigg\{ z \in V \;\Bigg|\; \int_{\{|Gv|_m > 0\}} k_m'(|Gv|_m) \left(\frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^3} \right) \mathrm{d}\mu < \infty \text{ for all } m = 1, ..., M, \\ &\qquad \int_{\{|Gv|_m > 0\}} k_m''(|Gv|_m) \frac{b_m(Gv, Gz)^2}{|Gv|_m^2} \mathrm{d}\mu < \infty \text{ for all } m = 1, ..., M, \\ &\qquad k_m'(0)(|Gz|_m - \langle \lambda_m, Gz \rangle) = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } k_m''(0) < \infty, \\ &\qquad |Gz|_m = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } k_m''(0) = \infty \Bigg\} \end{aligned} \tag{4.24}$$

and

$$
\begin{aligned}
Q_j^{v,\varphi}(z) = &\sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} k_m'(|Gv|_m) \left(\frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv,Gz)^2}{|Gv|_m^3} \right) \mathrm{d}\mu \\
&+ \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} k_m''(|Gv|_m) \frac{b_m(Gv,Gz)^2}{|Gv|_m^2} \mathrm{d}\mu \\
&+ \sum_{m:k_m''(0)<\infty} \int_{\{|Gv|_m=0\}} k_m''(0) |Gz|_m^2 \mathrm{d}\mu \qquad \forall z \in \mathcal{K}_j^{red}(v,\varphi).
\end{aligned} \tag{4.25}
$$

Here, $\lambda_1, ..., \lambda_M \in L^2(\Omega, H^*)$ denote the multipliers associated with φ as appearing in (4.22)*.*

Proof. Let v be fixed and assume that a $\varphi \in \partial j(v)$ with associated $\lambda, \lambda_1, ..., \lambda_M$ as in (4.22) is given. Then, the surjectivity of the operator G, Theorem 2.4.8, Theorem 2.5.5 and Theorem 4.2.3 imply that j is twice epi-differentiable in v for φ with

$$
\mathcal{K}_j^{red}(v,\varphi) = \left\{ z \in V \;\middle|\; \sum_{m=1}^{M} Q_{k_m(|\cdot|_m)}^{Gv, k_m'(|Gv|_m)\lambda_m}(Gz) \in L^1(\Omega, [0,\infty]) \right\}
$$

and

$$
Q_j^{v,\varphi}(z) = \int_\Omega \sum_{m=1}^{M} Q_{k_m(|\cdot|_m)}^{Gv, k_m'(|Gv|_m)\lambda_m}(Gz) \mathrm{d}\mu \qquad \forall z \in \mathcal{K}_j^{red}(v,\varphi).
$$

The explicit formulas in Theorem 4.2.3 now yield (4.24) and (4.25). This proves the claim. □

4.3.2 Application to Static Elastoplasticity

To demonstrate that the findings of Section 4.3.1 are not only of theoretical interest but also of relevance in practice, in what follows, we apply Theorem 4.3.3 to the EVI of static small strain elastoplasticity in primal formulation with linear kinematic hardening and von Mises yield criterion. This variational inequality arises, e.g., when the time discretization of a quasi-static elastoplastic process is considered or when instantaneous control strategies are used to optimize the response of materials that are subject to high external loads. For details on the latter topics and the physical background, we refer to [De los Reyes et al., 2016] where the problem of static elastoplasticity is studied with mollification techniques.

We again begin by clarifying our assumptions:

Assumption 4.3.4 (Setting of Static Elastoplasticity)**.**

- $\Omega \subset \mathbb{R}^3$ *is a bounded (strong) Lipschitz domain (the body under consideration),*
- Γ_D *is a relatively open and non-empty subset of the boundary $\partial\Omega$ (the Dirichlet boundary),*
- $H_D^1(\Omega, \mathbb{R}^3) := \left\{ u \in H^1(\Omega)^3 \;\middle|\; \mathrm{tr}(u) = 0 \; \mathcal{H}^2\text{-a.e. on } \Gamma_D \right\}$ *(the space of displacements),*
- $H := \mathbb{R}^{3\times3}_{dev} := \left\{ p \in \mathbb{R}^{3\times3}_{sym} \;\middle|\; p_{11} + p_{22} + p_{33} = 0 \right\}$ *(the space of plastic strain matrices),*
- $\|\cdot\|_F : \mathbb{R}^{3\times3} \to \mathbb{R}$ *is the Frobenius norm,*
- $V := H_D^1(\Omega, \mathbb{R}^3) \times L^2(\Omega, H)$ *(the space of unknowns),*
- $a : V \times V \to \mathbb{R}$ *is the bilinear form defined by*

$$
a\Big((u_1,p_1),(u_2,p_2)\Big) := \int_\Omega (\varepsilon(u_1) - p_1) : \mathbb{C}(\varepsilon(u_2) - p_2) + p_1 : \mathbb{H}p_2 \, \mathrm{d}\mathcal{L}^3,
$$

where $\varepsilon(u) := \frac{1}{2}(\nabla u + (\nabla u)^T)$ is the strain tensor associated with the displacement u, where ":" denotes the inner product induced by $\|\cdot\|_F$, and where $\mathbb{C} \in L^\infty(\Omega, L(\mathbb{R}^{3\times3}_{sym}, \mathbb{R}^{3\times3}_{sym}))$ and $\mathbb{H} \in L^\infty(\Omega, L(H,H))$ denote the elasticity tensor and the hardening modulus, respectively,

- $\mathbb{C}$ *and* $\mathbb{H}$ *are uniformly elliptic, i.e., there exists a constant* $c > 0$ *with*

$$\varepsilon : \mathbb{C}(x)\varepsilon \geq c\|\varepsilon\|_F^2 \quad \forall \varepsilon \in \mathbb{R}^{3\times 3}_{sym} \quad \textit{and} \quad p : \mathbb{H}(x)p \geq c\|p\|_F^2 \quad \forall p \in H \quad \textit{for a.a. } x \in \Omega,$$

- $P : V \to L^2(\Omega, H)$ *is the projection onto the second component, and* $j : V \to \mathbb{R}$ *is defined by*

$$j(v) := \sigma_0 \int_\Omega \|Pv\|_F \mathrm{d}\mathcal{L}^3 \qquad \forall v \in V \tag{4.26}$$

with some $\sigma_0 > 0$ *(the yield stress),*

- $f \in V^*$ *is a given datum (the external volume/boundary loads).*

Recall that tr, $\mathcal{H}^2$ and $\mathcal{L}^3$ denote the trace operator, the two-dimensional Hausdorff measure and the three-dimensional Lebesgue measure, respectively, see Section 3.4.1.

With the definitions of Assumption 4.3.4, the EVI of static elastoplasticity takes the following form (cf. [De los Reyes et al., 2016] and [Han and Reddy, 1999, Chapter 7]):

$$w \in V, \qquad a(w, v - w) + j(v) - j(w) \geq \langle f, v - w\rangle \qquad \forall v \in V. \tag{4.27}$$

Note that the bilinear form $a : V \times V \to \mathbb{R}$ in (4.27) is trivially continuous (by the L^∞-assumption on $\mathbb{C}$ and $\mathbb{H}$) and that the inequalities of Young, Korn and Poincaré (see [Schweizer, 2013, Theorem 4.22, Korollar 25.6]) imply the existence of a constant $\tilde{c} > 0$ with

$$\begin{aligned} a\Big((u,p),(u,p)\Big) &\geq c\|\varepsilon(u) - p\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} + c\|p\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} \\ &\geq c\|\varepsilon(u)\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} + 2c\|p\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} - 2c\|\varepsilon(u)\|_{L^2(\Omega,\mathbb{R}^{3\times 3})}\|p\|_{L^2(\Omega,\mathbb{R}^{3\times 3})} \\ &\geq \frac{c}{3}\|\varepsilon(u)\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} + \frac{c}{2}\|p\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})} \\ &\geq \tilde{c}\left(\|u\|^2_{H^1(\Omega)^3} + \|p\|^2_{L^2(\Omega,\mathbb{R}^{3\times 3})}\right). \end{aligned}$$

This shows that the map $A : V \to V^*$, $(u,p) \mapsto a((u,p),\cdot)$, satisfies the conditions in Assumption 1.2.1 and that (4.27) is a problem of the type (P). Since the functional j in (4.26) has exactly the structure (4.20) and since the projection $P \in L(V, L^2(\Omega, H))$ is surjective, (4.27) is further covered by Theorem 4.3.3. We may thus deduce:

Corollary 4.3.5. *The EVI* (4.27) *is uniquely solvable for all* $f \in V^*$*, and for every right-hand side* f *with associated solution* $w \in V$ *there exists a unique* $\lambda \in L^2(\Omega, H^*)$ *with* $\lambda \in \partial\|\cdot\|_F(Pw)$ *a.e. in* Ω *and* $\sigma_0 P^*\lambda = \varphi$*, where* $\varphi(\cdot) := f(\cdot) - a(w,\cdot) \in V^*$*. Further, the solution operator* $S : f \mapsto w$ *associated with* (4.27) *is globally Lipschitz continuous and Hadamard directionally differentiable, and the directional derivatives* $\delta := S'(f;g) \in V$*,* $f, g \in V^*$*, are uniquely characterized by the variational inequalities*

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w,\varphi), \\ &a(\delta, z - \delta) + \frac{\sigma_0}{2}\int_{\{Pw\neq 0\}} \frac{\|Pw\|_F^2\|Pz\|_F^2 - (Pw : Pz)^2}{\|Pw\|_F^3}\,\mathrm{d}\mathcal{L}^3 \\ &\qquad - \frac{\sigma_0}{2}\int_{\{Pw\neq 0\}} \frac{\|Pw\|_F^2\|P\delta\|_F^2 - (Pw : P\delta)^2}{\|Pw\|_F^3}\,\mathrm{d}\mathcal{L}^3 \geq \langle g, z - \delta\rangle \quad \forall z \in \mathcal{K}_j^{red}(w,\varphi), \end{aligned} \tag{4.28}$$

where $w := S(f)$*, where* φ *and* λ *are defined as before, and where*

$$\mathcal{K}_j^{red}(w,\varphi) = \left\{ z \in V \;\middle|\; \int_{\{Pw\neq 0\}} \frac{\|Pw\|_F^2\|Pz\|_F^2 - (Pw : Pz)^2}{\|Pw\|_F^3}\mathrm{d}\mathcal{L}^3 < \infty \right. \\ \left. \textit{and } \langle \lambda, Pz\rangle = \|Pz\|_F \textit{ a.e. in } \{Pw = 0\} \right\}. \tag{4.29}$$

Proof. The unique solvability and the global Lipschitz continuity of the solution operator are immediate consequences of Theorem 1.2.2, and from Proposition 4.3.2 and the surjectivity of G it follows straightforwardly that there exists a unique $\lambda \in L^2(\Omega, H^*)$ with $\sigma_0 P^* \lambda = \varphi$ and $\lambda \in \partial \|\cdot\|_F(Pw)$ a.e. in Ω. To obtain the remaining assertions of the theorem, it suffices to invoke Theorems 1.4.1 and 4.3.3. This completes the proof. □

Because of the convexity of the critical cone $\mathcal{K}_j^{red}(w, \varphi)$ and since the σ_0-terms in (4.28) are quadratic, the variational inequality for the directional derivatives $S'(f; g)$ in Corollary 4.3.5 can be reformulated as follows:

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w, \varphi), \\ &a(\delta, z - \delta) + \sigma_0 \int_{\{Pw \neq 0\}} \frac{\|Pw\|_F^2 \big(P\delta : P(z - \delta)\big) - (Pw : P\delta)\big(Pw : P(z - \delta)\big)}{\|Pw\|_F^3} \, \mathrm{d}\mathcal{L}^3 \geq \langle g, z - \delta \rangle \\ &\qquad\qquad \forall z \in \mathcal{K}_j^{red}(w, \varphi). \end{aligned} \tag{4.30}$$

The above implies that (4.28) is equivalent to an EVI of the first kind in the Hilbert space $(U, \|\cdot\|_U)$ defined by

$$U := H_D^1(\Omega, \mathbb{R}^3) \times \left\{ p \in L^2(\Omega, H) \;\middle|\; \int_{\{Pw \neq 0\}} \frac{\|Pw\|_F^2 \|p\|_F^2 - (Pw : p)^2}{\|Pw\|_F^3} \, \mathrm{d}\mathcal{L}^3 < \infty \right\}$$

and

$$\|v\|_U := \left(\|v\|_V^2 + \int_{\{Pw \neq 0\}} \frac{\|Pw\|_F^2 \|Pv\|_F^2 - (Pw : Pv)^2}{\|Pw\|_F^3} \, \mathrm{d}\mathcal{L}^3 \right)^{1/2},$$

cf. Section 3.1 and the discussion subsequent to Assumption 1.2.1. Note that the reduced critical cone $\mathcal{K}_j^{red}(w, \varphi)$ in (4.29) is indeed closed, convex and non-empty as a subset of U but typically not closed in the Hilbert space V due to the additional integrability condition in (4.29). This confirms the comments that we have made after Lemma 1.3.4.

We would like to point out that the above behavior is what is typically observed when the directional differentiability of the solution map $S : f \mapsto w$ to an EVI of the second kind involving a non-smooth Nemytskii operator is studied. The structure of the non-differentiable terms in (P) causes the directional derivatives $S'(f; g)$ to be characterized by (generalized) projections in Hilbert spaces which depend on the state $w = S(f)$ and whose norms are stronger than that of the original Hilbert space V due to the appearance of an additional bilinear form that contains curvature information. We will encounter the same effect, e.g., in Theorem 4.3.16, Section 5.1 and Section 5.2 (compare also with Theorem 4.3.3 and Corollary 2.1.2 in this regard).

What is remarkable in the context of the variational inequality (4.27) is that the properties of the EVIs for the directional derivatives $S'(f; g)$ depend heavily on the $\mathbb{R}^{3\times3}$-norm appearing in (4.26) (and thus on the yield criterion that governs the underlying elastoplastic process, cf. [Han and Reddy, 1999, Chapter 7]). If we define, for example,

$$\tilde{j}(v) := \sigma_0 \int_\Omega \|Pv\|_1 \mathrm{d}\mathcal{L}^3 \qquad \forall v \in V, \tag{4.31}$$

where

$$\|\cdot\|_1 : \mathbb{R}^{3\times3} \to \mathbb{R}, \quad p \mapsto \sum_{m,n} |p_{mn}|,$$

denotes the 1-norm on $\mathbb{R}^{3\times3}$, and consider the EVI

$$w \in V, \quad a(w, v - w) + \tilde{j}(v) - \tilde{j}(w) \geq \langle f, v - w \rangle \quad \forall v \in V, \tag{4.32}$$

then we obtain:

Corollary 4.3.6. *The EVI* (4.32) *is uniquely solvable for all $f \in V^*$, and for every right-hand side f with associated solution $w \in V$ there exist functions $\lambda_{mn} \in L^2(\Omega, H^*)$ such that $\lambda_{mn} \in \partial|\cdot|_{mn}(Pv)$ holds a.e. in Ω (where $|p|_{mn} := |p_{mn}|$ for all $p \in H$) and such that*

$$\langle \varphi, z\rangle_V := \langle f, z\rangle_V - a(w,z) = \sigma_0 \sum_{m,n} \int_\Omega \langle \lambda_{mn}, Pz\rangle \,\mathrm{d}\mathcal{L}^3 \qquad \forall z \in V.$$

Further, the solution operator $S : f \mapsto w$ associated with (4.32) *is globally Lipschitz continuous and Hadamard directionally differentiable, and the directional derivatives $\delta := S'(f;g) \in V$, $f, g \in V^*$, are uniquely characterized by the variational inequalities*

$$\delta \in \mathcal{K}_{\tilde{j}}^{red}(w,\varphi), \qquad a(\delta, z-\delta) \geq \langle g, z-\delta\rangle \qquad \forall z \in \mathcal{K}_{\tilde{j}}^{red}(w,\varphi), \tag{4.33}$$

where $w := S(f)$, where φ and λ_{mn} are defined as before, and where

$$\mathcal{K}_{\tilde{j}}^{red}(w,\varphi) = \{z \in V \mid \langle \lambda_{mn}, Pz\rangle = |(Pz)_{mn}| \text{ a.e. in } \{(Pw)_{mn} = 0\}\ \forall m,n\}.$$

Proof. The unique solvability of (4.32) and the Lipschitz continuity of the solution operator $S : f \mapsto w$ follow immediately from Theorem 1.2.2, and the existence of the functions λ_{mn} is obtained from case (i) in Proposition 4.3.2. The remaining claims are straightforward consequences of Theorem 1.4.1 and Theorem 4.3.3. □

Several things are noteworthy regarding Corollary 4.3.6:

Remark 4.3.7.

(i) The EVI (4.33) *for the directional derivatives of the solution map S to* (4.32) *is exceptional in the sense that it does not contain an additional curvature term. (As we have already pointed out such terms normally appear when functionals of the type* (4.20) *are studied.) This is due to the surjectivity of the operator G and the fact that the Nemytskii operators in* (4.20) *are absolute value functions. The reader should be warned that the emergence of additional terms has to be expected when one of the latter conditions is violated. This can be seen, e.g., in Section 5.2, where the second subderivative of the function $j : H_0^1(\Omega) \to \mathbb{R}$, $v \mapsto \|v\|_{L^1}$, involving the (obviously non-surjective) embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ is studied, cf. Theorems 5.2.14 and 5.2.15.*

(ii) Note that the yield stress σ_0 enters the EVI (4.33) *only indirectly via the multipliers λ_{mn}. This is different in* (4.28) *where σ_0 appears explicitly.*

(iii) We remark that non-differentiable terms of the form (4.31) *also appear in the study of hyperbolic Maxwell variational inequalities, cf. [Yousept, 2017].*

4.3.3 Regularization by Singular Curvature for an Optimal Control Problem

Before we turn our attention to functionals of the form (4.20) with non-surjective operators G, we would like to point out that the regularization effect studied in Section 4.2.2 can also be exploited in the setting of Assumption 4.3.1. Consider, for example, the primitive tracking-type optimal control problem

$$\min_{v \in L^2(\Omega)} \int_\Omega (T(v) - f)^2 \mathrm{d}\mathcal{L}^d + \sum_{m=1}^4 \alpha_m \|v\|_{L^{\beta_m}}^{\beta_m} \tag{4.34}$$

with a bounded domain $\Omega \subset \mathbb{R}^d$, the Lebesgue measure $\mathcal{L}^d$, $\beta_1 = 2$, $\beta_2 = 1$, $\beta_3 \in (1,2)$, $\beta_4 \in (2,\infty)$, $\alpha_m > 0$ for $m = 1, ..., 4$, $f \in L^2(\Omega)$, and a linear and continuous operator $T : L^2(\Omega) \to L^2(\Omega)$. Then, it is easy to see that (4.34) is equivalent to the EVI

$$w \in L^2(\Omega), \qquad 2\alpha_1 (w, v-w)_{L^2} + j(v) - j(w) \geq (2T^* f, v-w)_{L^2} \quad \forall v \in L^2(\Omega),$$

where

$$j : L^2(\Omega) \to (-\infty, \infty], \qquad j(v) := \int_\Omega |T(v)|^2 + \sum_{m=2}^{4} \alpha_m |v|^{\beta_m} \mathrm{d}\mathcal{L}^d. \tag{4.35}$$

Note that the domain of the above j is precisely $L^{\beta_4}(\Omega)$ and that $\mathrm{int}(\mathrm{dom}(j)) = \emptyset$ in $L^2(\Omega)$. Further, the results of the last sections yield:

Proposition 4.3.8. *Let j, Ω, β_m etc. be as above and identify $L^2(\Omega)^*$ with $L^2(\Omega)$. Then, it holds*

$$\begin{aligned} \partial j(v) = \Big\{ \varphi \in L^2(\Omega) \;\Big|\; & \exists \lambda \in L^2(\Omega) \text{ s.t. } \lambda \in \partial|v| \text{ a.e. in } \Omega, \\ & \varphi = 2T^*T(v) + \alpha_2\lambda + \alpha_3\beta_3|v|^{\beta_3-2}v + \alpha_4\beta_4|v|^{\beta_4-2}v \Big\} \end{aligned} \tag{4.36}$$

for all $v \in L^2(\Omega)$, and the functional j is twice epi-differentiable in every $v \in L^2(\Omega)$ for all $\varphi \in \partial j(v)$ with

$$\mathcal{K}_j^{red}(v,\varphi) = \left\{ z \in L^2(\Omega) \;\middle|\; \int_{\{v\neq 0\}} \left(|v|^{\beta_3-2} + |v|^{\beta_4-2} \right) z^2 \mathrm{d}\mathcal{L}^d < \infty,\ z = 0 \text{ a.e. in } \{v = 0\} \right\} \tag{4.37}$$

and

$$Q_j^{v,\varphi}(z) = \int_\Omega 2(Tz)^2 \mathrm{d}\mathcal{L}^d + \int_{\{v\neq 0\}} \sum_{m=3}^{4} \alpha_m\beta_m(\beta_m - 1)|v|^{\beta_m-2} z^2 \mathrm{d}\mathcal{L}^d \qquad \forall z \in \mathcal{K}_j^{red}(v,\varphi).$$

Proof. From the sum rule for convex subdifferentials, we obtain

$$\partial j(v) = 2T^*T(v) + \partial \left(\int_\Omega \sum_{m=2}^{4} \alpha_m |\cdot|^{\beta_m} \mathrm{d}\mathcal{L}^d \right)(v)$$

for all $v \in L^2(\Omega)$. If we apply case (ii) of Proposition 4.3.2 to the second term on the right-hand side of the above equation (with $G = \mathrm{Id}$), then (4.36) follows immediately. Using Theorems 2.2.1 and 4.3.3, we now obtain that j is twice epi-differentiable in every $v \in L^2(\Omega)$ for all $\varphi \in \partial j(v)$ with $\mathcal{K}_j^{red}(v,\varphi)$ and $Q_j^{v,\varphi}(z)$ as claimed. This completes the proof. □

The above implies:

Theorem 4.3.9. *Let j, Ω, β_m etc. be as before. Then, for every $f \in L^2(\Omega)$, there exists one and only one solution $w \in L^2(\Omega)$ to* (4.34)*. Further, the solution map $S : L^2(\Omega) \to L^2(\Omega), f \mapsto w$, associated with* (4.34) *is globally Lipschitz continuous and Hadamard-Gâteaux differentiable and the directional derivative in a point $f \in L^2(\Omega)$ with associated solution $w := S(f)$ and $\varphi := 2T^*f - 2\alpha_1 w$ in a direction $g \in L^2(\Omega)$ is uniquely characterized by the variational equality*

$$\begin{aligned} \delta \in \mathcal{K}_j^{red}(w,\varphi), \quad 2\big(\alpha_1\delta + T^*T\delta - T^*g, z\big)_{L^2} + \int_{\{w\neq 0\}} \sum_{m=3}^{4} \alpha_m\beta_m(\beta_m-1)|w|^{\beta_m-2}\delta z \mathrm{d}\mathcal{L}^d = 0 \\ \forall z \in \mathcal{K}_j^{red}(w,\varphi). \end{aligned} \tag{4.38}$$

Here, $\mathcal{K}_j^{red}(w,\varphi)$ is the subspace defined in (4.37)*.*

Proof. Combine Theorem 1.2.2, Proposition 4.3.8, Corollary 1.4.4 and the chain rule for Hadamard directionally differentiable functions as found in [Bonnans and Shapiro, 2000, Proposition 2.47]. □

We can now make the following observations:

Remark 4.3.10.

(i) Theorem 4.3.9 shows that the solution map $S : f \mapsto w$ associated with (4.34) *is Hadamard-Gâteaux differentiable in spite of the appearing non-differentiable function $\alpha_2\|\cdot\|_{L^1}$. The reason for this behavior is again the singular curvature of the term $\alpha_3\|\cdot\|_{L^{\beta_3}}^{\beta_3}$ in* (4.34)*, cf. the effects in Section 4.2.2 and Theorem 4.3.3. Note that S is not Gâteaux anymore when $\alpha_3 = 0$.*

(ii) The functional $\sum_{m=1}^4 \alpha_m\|v\|_{L^{\beta_m}}^{\beta_m}$ in the minimization problem (4.34) *can be interpreted as a non-standard Tikhonov regularization term whose components serve different purposes:*

- *the term $\alpha_1\|\cdot\|_{L^2}^2$ ensures the Lipschitz continuity and ultimately the Hadamard directional differentiability of the solution map S (cf. the example in* (1.12)*),*
- *the term $\alpha_2\|\cdot\|_{L^1}$ promotes sparsely supported optimal controls, see [Casas et al., 2012],*
- *the term $\alpha_3\|\cdot\|_{L^{\beta_3}}^{\beta_3}$ ensures the Gâteaux differentiability of S,*
- *the term $\alpha_4\|\cdot\|_{L^{\beta_4}}^{\beta_4}$ increases the regularity of the optimal control.*

We expect that the regularizing effect that the functional $\sum_{m=1}^4 \alpha_m\|v\|_{L^{\beta_m}}^{\beta_m}$ has on the solution operator S and the possibility to adjust the coefficients α_m according to one's preferences are handy tools, e.g., in the study of bilevel optimal control problems. We further believe that the above observations are also relevant when more complicated objective functions are considered. Since our analysis is tailored to the study of classical EVIs, exploring the latter issues would go beyond the scope of this work.

(iii) It should be noted that, although the solution operator S maps into $\mathrm{dom}(j) = L^{\beta_4}(\Omega)$, we only obtain Gâteaux differentiability in $L^2(\Omega)$ in Theorem 4.3.9. The reason for this is that the norm in the ellipticity estimate is decisive for the strength of the obtained differentiability result.

(iv) Analogously to (4.30)*,* (4.38) *is a variational problem in a weighted Hilbert space.*

4.3.4 Second-Order Epi-Differentiability in the Absence of Surjectivity

In what follows, we consider the case where the operator $G \in L(V, L^2(\Omega, H))$ in (4.20) is non-surjective. Note that this is what is typically encountered when EVIs in Sobolev spaces are studied since, e.g., the functions $H_0^1(\Omega) \ni v \mapsto v \in L^2(\Omega)$ and $H_0^1(\Omega) \ni v \mapsto \nabla v \in L^2(\Omega, \mathbb{R}^d)$ clearly lack the property of surjectivity, cf. Sections 4.3.5, 5.1 and 5.2. For the sake of simplicity and with view on the analysis in Chapter 5, in the remainder of this chapter, we restrict our attention to functionals j of the form

$$j(v) = \int_\Omega \sum_{m=1}^M |Gv|_m^{\beta_m} \mathrm{d}\mu. \tag{4.39}$$

To be more precise, we suppose that j satisfies:

Assumption 4.3.11.

- *j is of the type* (4.20)*,*
- *V, H, Ω, Σ, μ, G, M, $|\cdot|_m$ and b_m are as in Assumption 4.3.1,*
- *$k_m(x) = x^{\beta_m}$ with $\beta_m \in [1, 2)$ for $m = 1, ..., M$.*

As already mentioned, the main difficulty in the study of functionals of the form (4.39) with non-surjective operators G is that the chain rule in Theorem 2.4.8 is typically inapplicable. To circumvent this problem, we will use an argumentation that exploits Lemmas 1.3.13 and 4.2.2.

We begin by noting that Hölder's inequality and our assumptions on (Ω, Σ, μ), β_m and b_m imply that the functions

$$L^2(\Omega, H) \ni u \mapsto \int_\Omega |u|_m^{\beta_m} \mathrm{d}\mu$$

are continuous for all $m = 1, ..., M$. This yields that the condition in case (i) of Proposition 4.3.2 is satisfied and that

$$\begin{aligned} &\partial j(v) \\ &= G^* \left\{ \lambda \in L^2(\Omega, H^*) \;\middle|\; \exists \lambda_m \in L^2(\Omega, H^*) \text{ s.t. } \lambda_m \in \partial |\cdot|_m(Gv) \text{ a.e., } \lambda = \sum_{m=1}^M \beta_m |Gv|_m^{\beta_m - 1} \lambda_m \right\} \end{aligned} \tag{4.40}$$

holds for all $v \in V$, where we use the convention $0^0 := 1$. From the above formula and Theorem 4.3.3, we obtain a first estimate for the second subderivative of j:

Lemma 4.3.12. *Let $v \in V$ be arbitrary but fixed, and let $\varphi = G^*\lambda$ be an element of $\partial j(v)$ with associated λ_m, $m = 1, ..., M$, as in* (4.40)*. Then, it holds*

$$\begin{aligned} \mathcal{K}_j^{red}(v, \varphi) \subset \Bigg\{ z \in V \;\Bigg|\; & \int_{\{|Gv|_m \neq 0\}} \frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^3} \mathrm{d}\mu < \infty \text{ for all } m \text{ with } \beta_m = 1, \\ & \int_{\{|Gv|_m \neq 0\}} \frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}} \mathrm{d}\mu < \infty \text{ for all } m \text{ with } \beta_m \in (1,2), \\ & |Gz|_m = \langle \lambda_m, Gz \rangle \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m = 1, \\ & |Gz|_m = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m \in (1,2) \Bigg\} \end{aligned} \tag{4.41}$$

and for every $z \in \mathcal{K}_j^{red}(v, \varphi)$ it is true that

$$\begin{aligned} &Q_j^{v,\varphi}(z) \\ &\geq \sum_{m=1}^M \int_{\{|Gv|_m \neq 0\}} \beta_m \frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^{4-\beta_m}} \mathrm{d}\mu + \int_{\{|Gv|_m \neq 0\}} (\beta_m^2 - \beta_m) \frac{b_m(Gv, Gz)^2}{|Gv|_m^{4-\beta_m}} \mathrm{d}\mu. \end{aligned} \tag{4.42}$$

Proof. Define

$$k : L^2(\Omega, H) \to \mathbb{R}, \qquad u \mapsto \int_\Omega \sum_{m=1}^M |u|_m^{\beta_m} \mathrm{d}\mu.$$

Then, by Definition 1.3.1 and the relationship between j and k, we have

$$\begin{aligned} Q_j^{v,\varphi}(z) &= \inf \left\{ \liminf_{n\to\infty} \frac{2}{t_n} \left(\frac{j(v + t_n z_n) - j(v)}{t_n} - \langle \varphi, z_n \rangle \right) \;\middle|\; \begin{array}{l} \{t_n\} \subset \mathbb{R}^+, \{z_n\} \subset V, \\ t_n \searrow 0, \ z_n \rightharpoonup z \text{ in } V \end{array} \right\} \\ &\geq \inf \left\{ \liminf_{n\to\infty} \frac{2}{t_n} \left(\frac{k(Gv + t_n u_n) - k(Gv)}{t_n} - \langle \lambda, u_n \rangle \right) \;\middle|\; \begin{array}{l} \{t_n\} \subset \mathbb{R}^+, \{u_n\} \subset L^2(\Omega, H), \\ t_n \searrow 0, \ u_n \rightharpoonup Gz \text{ in } L^2(\Omega, H) \end{array} \right\} \\ &= Q_k^{Gv,\lambda}(Gz). \end{aligned} \tag{4.43}$$

Using the explicit formulas for $Q_k^{Gv,\lambda}(Gz)$ in Theorem 4.3.3, we now immediately obtain the claim. □

Remark 4.3.13. *The argument that we have used in the above proof also holds in general: If we are given a convex, lower semicontinuous and proper function $k : U \to (-\infty, \infty]$ on some Hilbert space U and know that V is continuously embedded into U, then the second subderivative of k as a function on U always provides a lower bound for the second subderivative of k as a function on V.*

In what follows, our aim will be to find a condition that is sufficient for equality in (4.41) and (4.42). To obtain such a criterion, we need:

Lemma 4.3.14. *Let $v \in V$ be arbitrary but fixed, and let $\varphi = G^*\lambda$ be an element of $\partial j(v)$ with associated λ_m, $m = 1, ..., M$, as in* (4.40)*. Then, for every $z \in V$ and every $t > 0$, it holds*

$$\begin{aligned}
&\frac{2}{t}\left(\frac{j(v+tz)-j(v)}{t} - \langle \varphi, z\rangle\right) \\
&\quad= \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \beta_m \left(4\frac{|Gv|_m^2|Gz|_m^2 - b_m(Gv,Gz)^2}{(|Gv+tGz|_m + |Gv|_m)^2|Gv|_m^{2-\beta_m}}\right) \mathrm{d}\mu \\
&\quad+ \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \beta_m \left(2\frac{|Gv|_m(|Gv+tGz|_m - |Gv|_m) - tb_m(Gv,Gz)}{(|Gv+tGz|_m + |Gv|_m)^2|Gv|_m^{2-\beta_m}}|Gz|_m^2\right) \mathrm{d}\mu \\
&\quad+ \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \left(\frac{8b_m(Gv,Gz)^2}{(|Gv|_m + |Gv+tGz|_m)^2} + \frac{8tb_m(Gv,Gz) + 2t^2|Gz|_m^2}{(|Gv|_m + |Gv+tGz|_m)^2}|Gz|_m^2\right) \\
&\qquad\qquad \cdot \int_0^1 \frac{(\beta_m^2-\beta_m)(1-s)}{\Big((1-s)|Gv|_m + s|Gv+tGz|_m\Big)^{2-\beta_m}} \mathrm{d}s\mathrm{d}\mu \\
&\quad+ \sum_{m:\beta_m=1} \int_{\{|Gv|_m=0\}} 2\frac{|Gz|_m - \langle \lambda_m, Gz\rangle}{t}\mathrm{d}\mu \\
&\quad+ \sum_{m:\beta_m\in(1,2)} \int_{\{|Gv|_m=0\}} 2\frac{|Gz|_m^{\beta_m}}{t^{2-\beta_m}}\mathrm{d}\mu.
\end{aligned} \tag{4.44}$$

Proof. To obtain the lengthy formula (4.44), we just have to apply Lemma 4.2.2 pointwise and use (4.40). (Note that we can safely ignore the condition $|Gv+tGz|_m > 0$ in Lemma 4.2.2 here since the s-integral is also defined in the case $|Gv + tGz|_m = 0$.) This proves the claim. □

From the dominated convergence theorem, we may now deduce:

Lemma 4.3.15. *Let $v \in V$ be arbitrary but fixed, and let $\varphi = G^*\lambda$ be an element of $\partial j(v)$ with associated λ_m, $m = 1, ..., M$, as in* (4.40)*. Define*

$$\begin{aligned}
\mathcal{Z} := \Bigg\{ z \in V \;\Bigg|\; & \int_{\{|Gv|_m\neq 0\}} \frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}}\mathrm{d}\mu < \infty \text{ for all } m, \\
& |Gz|_m = \langle \lambda_m, Gz\rangle \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m = 1, \\
& |Gz|_m = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m \in (1,2) \Bigg\}.
\end{aligned} \tag{4.45}$$

Then, $\mathcal{Z}$ is a subset of the reduced critical cone $\mathcal{K}_j^{red}(v,\varphi)$ and j is twice epi-differentiable in v for φ in all directions $z \in \mathcal{Z}$ with

$$\begin{aligned}
&Q_j^{v,\varphi}(z) \\
&= \sum_{m=1}^{M} \int_{\{|Gv|_m\neq 0\}} \beta_m \frac{|Gv|_m^2|Gz|_m^2 - b_m(Gv,Gz)^2}{|Gv|_m^{4-\beta_m}}\mathrm{d}\mu + \int_{\{|Gv|_m\neq 0\}} (\beta_m^2-\beta_m)\frac{b_m(Gv,Gz)^2}{|Gv|_m^{4-\beta_m}}\mathrm{d}\mu.
\end{aligned}$$

The set in (4.45) will later on play exactly the role of the set $\mathcal{Z}$ in Lemma 1.3.13 (hence the name).

Proof of Lemma 4.3.15. Consider an arbitrary but fixed $z \in \mathcal{Z}$. Then, the properties of z and (4.44) yield

$$\begin{aligned}
&\frac{2}{t}\left(\frac{j(v+tz)-j(v)}{t} - \langle \varphi, z\rangle\right) \\
&\quad= \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \beta_m \left(4\frac{|Gv|_m^2|Gz|_m^2 - b_m(Gv,Gz)^2}{(|Gv+tGz|_m+|Gv|_m)^2|Gv|_m^{2-\beta_m}}\right) \mathrm{d}\mu \\
&\quad+ \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \beta_m \left(2\frac{|Gv|_m(|Gv+tGz|_m-|Gv|_m) - tb_m(Gv,Gz)}{(|Gv+tGz|_m+|Gv|_m)^2|Gv|_m^{2-\beta_m}}|Gz|_m^2\right) \mathrm{d}\mu \qquad (4.46) \\
&\quad+ \sum_{m=1}^{M} \int_{\{|Gv|_m>0\}} \left(\frac{8b_m(Gv,Gz)^2}{(|Gv|_m+|Gv+tGz|_m)^2} + \frac{8tb_m(Gv,Gz)+2t^2|Gz|_m^2}{(|Gv|_m+|Gv+tGz|_m)^2}|Gz|_m^2\right) \\
&\qquad\qquad \cdot \int_0^1 \frac{(\beta_m^2-\beta_m)(1-s)}{\big((1-s)|Gv|_m+s|Gv+tGz|_m\big)^{2-\beta_m}} \mathrm{d}s\mathrm{d}\mu.
\end{aligned}$$

Note that the integrands in the above integrals satisfy

$$\begin{aligned}
&\left|\beta_m\left(4\frac{|Gv|_m^2|Gz|_m^2 - b_m(Gv,Gz)^2}{(|Gv+tGz|_m+|Gv|_m)^2|Gv|_m^{2-\beta_m}}\right)\right| \le 8\beta_m\frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}}, \\
&\left|\beta_m\left(2\frac{|Gv|_m(|Gv+tGz|_m-|Gv|_m) - tb_m(Gv,Gz)}{(|Gv+tGz|_m+|Gv|_m)^2|Gv|_m^{2-\beta_m}}|Gz|_m^2\right)\right| \le 4\beta_m\frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}}
\end{aligned}$$

and

$$\begin{aligned}
&\left|\left(\frac{8b_m(Gv,Gz)^2}{(|Gv|_m+|Gv+tGz|_m)^2} + \frac{8tb_m(Gv,Gz)+2t^2|Gz|_m^2}{(|Gv|_m+|Gv+tGz|_m)^2}|Gz|_m^2\right)\right. \\
&\qquad\qquad \left.\cdot \int_0^1 \frac{(\beta_m^2-\beta_m)(1-s)}{\Big((1-s)|Gv|_m+s|Gv+tGz|_m\Big)^{2-\beta_m}}\mathrm{d}s\right| \\
&\le \left|18|Gz|_m^2 \cdot \int_0^1 \frac{(\beta_m^2-\beta_m)(1-s)}{((1-s)|Gv|_m)^{2-\beta_m}}\mathrm{d}s\right| \\
&\le 18(\beta_m-1)\frac{|Gz|_m^2}{|Gv|_m^{2-\beta}}
\end{aligned}$$

a.e. on $\{|Gv|_m > 0\}$ for all $t > 0$. This shows that the dominated convergence theorem is applicable and allows us to pass to the limit $t \searrow 0$ in (4.46). Using (4.42) and Definition 1.3.1, the claim now follows immediately. □

If we combine the above with Lemma 1.3.13, then we arrive at:

Theorem 4.3.16. *Suppose that a function of the form* (4.39) *is given and that Assumption 4.3.11 is satisfied. Let $v \in V$ be arbitrary but fixed and let $\varphi = G^*\lambda$ be an element of $\partial j(v)$ with associated λ_m, $m = 1, ..., M$, as in* (4.40)*. Define*

$$\begin{aligned}
\mathcal{Z} := \Bigg\{ z \in V \;\Bigg|\; &\int_{\{|Gv|_m\neq 0\}} \frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}}\mathrm{d}\mu < \infty \text{ for all } m, \\
&|Gz|_m = \langle \lambda_m, Gz\rangle \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m = 1, \qquad (4.47) \\
&|Gz|_m = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m \in (1,2)\Bigg\},
\end{aligned}$$

$$\mathcal{K} := \left\{ z \in V \;\middle|\; \begin{aligned} &\int_{\{|Gv|_m \neq 0\}} \frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^3} \mathrm{d}\mu < \infty \text{ for all } m \text{ with } \beta_m = 1, \\ &\int_{\{|Gv|_m \neq 0\}} \frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}} \mathrm{d}\mu < \infty \text{ for all } m \text{ with } \beta_m \in (1,2), \\ &|Gz|_m = \langle \lambda_m, Gz \rangle \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m = 1, \\ &|Gz|_m = 0 \ \mu\text{-a.e. in } \{|Gv|_m = 0\} \text{ for all } m \text{ with } \beta_m \in (1,2) \end{aligned} \right\}, \tag{4.48}$$

and

$$\|z\|_{\mathcal{K}} := \left(\|z\|_V^2 + \int_{\{|Gv|_m \neq 0\}} \sum_{m:\beta_m = 1} \frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^3} + \sum_{m:\beta_m \in (1,2)} \frac{|Gz|_m^2}{|Gv|_m^{2-\beta_m}} \, \mathrm{d}\mu \right)^{1/2}.$$

Suppose further that

$$\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}. \tag{4.49}$$

Then, j is twice epi-differentiable in v for φ, it holds $\mathcal{K} = \mathcal{K}_j^{red}(v, \varphi)$, and for all $z \in \mathcal{K}$ it is true that

$$Q_j^{v,\varphi}(z) = \sum_{m=1}^{M} \int_{\{|Gv|_m \neq 0\}} \beta_m \frac{|Gv|_m^2 |Gz|_m^2 - b_m(Gv, Gz)^2}{|Gv|_m^{4-\beta_m}} + (\beta_m^2 - \beta_m) \frac{b_m(Gv, Gz)^2}{|Gv|_m^{4-\beta_m}} \mathrm{d}\mu. \tag{4.50}$$

Proof. Let us denote the right-hand side of (4.50) with Q. Then, Lemma 4.3.12, Lemma 4.3.15, (4.43) and the density $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$ yield that $\mathcal{Z} \subset \mathcal{K}_j^{red}(v, \varphi) \subset \mathcal{K}$, that $Q_j^{v,\varphi}(z) \geq Q(z)$ for all $z \in \mathcal{K}$, that j is twice epi-differentiable in v for φ in all directions $z \in \mathcal{Z}$, and that for every $z \in \mathcal{K}$ there exists a sequence $z_n \in \mathcal{Z}$ with $z_n \to z$ and

$$Q(z) = \lim_{n \to \infty} Q(z_n) = \lim_{n \to \infty} Q_j^{v,\varphi}(z_n).$$

If we combine the above with Lemma 1.3.13, then the claim of the theorem follows immediately. □

Some remarks are in order regarding the sufficient criterion for second-order epi-differentiability in Theorem 4.3.16:

Remark 4.3.17.

(i) For all $m = 1, ..., M$ with $\beta_m \in (1,2)$, the integrability conditions in $\mathcal{Z}$ and $\mathcal{K}$ are exactly the same. This means in particular that the density $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$ and, as a consequence, the second-order epi-differentiability of j are granted if there are no m with $\beta_m = 1$. Note that the case $\beta_m \in (1,2)$ is already non-trivial as it is, e.g., not covered by Theorem 2.1.1. Note further that in the case $\beta_m = 1$ and $|\cdot|_m = \|\cdot\|_H$, the integrability conditions in $\mathcal{K}$ and $\mathcal{Z}$ differ therein that the condition in $\mathcal{K}$ is a condition on the component of Gz orthogonal to Gv, i.e.,

$$\int_{\{\|Gv\|_H \neq 0\}} \frac{\|Gv\|_H^2 \|Gz\|_H^2 - (Gv, Gz)_H^2}{\|Gv\|_H^3} \mathrm{d}\mu < \infty,$$

while the integrability condition in $\mathcal{Z}$ is a condition on the whole of Gz, i.e.,

$$\int_{\{\|Gv\|_H \neq 0\}} \frac{\|Gv\|_H^2 \|Gz\|_H^2}{\|Gv\|_H^3} \mathrm{d}\mu < \infty.$$

(ii) Using the bounded inverse theorem, it is easy to check that the density condition $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$ is always satisfied when G is surjective. As a consequence, Theorem 4.3.16 may be used to obtain an alternative proof of Theorem 4.3.3 (for k_m chosen as in (4.39)*).*

(iii) Using Taylor-like expansions analogous to (2.13)*, it is possible to extend the analysis of this section to more general functions k_m, cf. the calculations in [Christof and Meyer, 2016]. We do not pursue this approach here for the sake of readability.*

(iv) The density condition $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$ in Theorem 4.3.16 is an exact analogue of the condition

$$\mathcal{T}_K(v) \cap \ker(\varphi) = \mathrm{cl}\big(\{z \in \mathcal{T}_K(v) \;\big|\; 0 \in \mathcal{T}_K^2(v,z)\} \cap \ker(\varphi)\big)$$

appearing in the definition of extended polyhedricity, see Definition 3.3.1(iii). The difference here is that in the case of the functional j in (4.39)*, curvature effects have to be taken into account. Because of these effects, we have to demand density w.r.t. the stronger norm $\|\cdot\|_{\mathcal{K}}$ and cannot work with the topology of the original space V anymore.*

4.3.5 Application to PDEs Involving Singular Terms

To illustrate the usefulness of Theorem 4.3.16, we consider a bounded Lipschitz domain $\Omega \subset \mathbb{R}^3$, the spaces $V = H_0^1(\Omega)$ and $V^* = H^{-1}(\Omega)$ (defined as usual), and the equation

$$w \in H_0^1(\Omega), \quad -\Delta w - \nabla \cdot \left(\frac{\nabla w}{\|\nabla w\|_2^{1/2}} \right) + w^5 = f \in H^{-1}(\Omega). \tag{4.51}$$

Note that the above PDE is equivalent to the EVI

$$\begin{aligned} &w \in H_0^1(\Omega), \\ &\int_\Omega \nabla w \cdot \nabla(v-w) \mathrm{d}\mathcal{L}^3 + \int_\Omega \frac{2}{3}\|\nabla v\|_2^{3/2} + \frac{1}{6}v^6 \mathrm{d}\mathcal{L}^3 - \int_\Omega \frac{2}{3}\|\nabla w\|_2^{3/2} + \frac{1}{6}w^6 \mathrm{d}\mathcal{L}^3 \geq \langle f, v-w\rangle \\ &\qquad\qquad \forall v \in H_0^1(\Omega), \end{aligned}$$

where $\mathcal{L}^3$ again denotes the Lebesgue measure. This shows that (4.51) can be rewritten as a problem of the form (P) with

$$j(v) := \int_\Omega \frac{2}{3}\|\nabla v\|_2^{3/2} + \frac{1}{6}v^6 \mathrm{d}\mathcal{L}^3. \tag{4.52}$$

From Theorem 4.3.16, we may now deduce:

Corollary 4.3.18. *The solution operator $S : H^{-1}(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, to* (4.51) *is well-defined, globally Lipschitz and Hadamard-Gâteaux differentiable, and the directional derivative $\delta := S'(f;g)$ in a point $f \in H^{-1}(\Omega)$ with associated solution $w \in H_0^1(\Omega)$ and $\varphi := f + \Delta w \in H^{-1}(\Omega)$ in a direction $g \in H^{-1}(\Omega)$ is uniquely characterized by the variational equality*

$$\begin{aligned} &\delta \in \mathcal{K}_j^{red}(w,\varphi) \\ &\int_\Omega \nabla\delta \cdot \nabla z + 5w^4\delta z \,\mathrm{d}\mathcal{L}^3 + \int_{\{\nabla w \neq 0\}} \frac{\|\nabla w\|_2^2(\nabla\delta\cdot\nabla z) - \frac{1}{2}(\nabla w\cdot\nabla\delta)(\nabla w\cdot\nabla z)}{\|\nabla w\|_2^{5/2}} \mathrm{d}\mathcal{L}^3 = \langle g, z\rangle \\ &\qquad\qquad \forall z \in \mathcal{K}_j^{red}(w,\varphi) \end{aligned}$$

with

$$\mathcal{K}_j^{red}(w,\varphi) = \left\{ z \in H_0^1(\Omega) \;\middle|\; \int_{\{\nabla w\neq 0\}} \frac{\|\nabla z\|_2^2}{\|\nabla w\|_2^{1/2}} \mathrm{d}\mathcal{L}^3 < \infty, \nabla z = 0 \text{ a.e. in } \{\nabla w = 0\} \right\}. \tag{4.53}$$

Proof. Since the function $H_0^1(\Omega) \ni v \mapsto \|v\|_{L^6}^6 \in \mathbb{R}$ is twice Fréchet differentiable (cf. the Sobolev embeddings, [Adams, 1975, Theorem 5.4 Part A]) and since $\nabla \in L(H_0^1(\Omega), L^2(\Omega,\mathbb{R}^3))$, we may invoke Theorems 2.2.1 and 4.3.16 to obtain that the functional j in (4.52) is twice epi-differentiable. The claim now follows immediately from Theorem 1.4.1 and Theorem 1.2.2. □

Note that (4.53) is again a weighted space. This has to be taken into account when, e.g., stationarity conditions for optimal control problems governed by (4.51) are considered, cf. Section 6.1.

5 EVIs of the Second Kind in Sobolev Spaces

As the reader might have noticed, up to now, we have only considered H_0^1-elliptic variational inequalities that can be rewritten as partial differential equations or (generalized) projections, cf. Corollaries 3.4.3 and 3.4.4 and Sections 3.4.2, 4.1 and 4.3.5. In what follows, we change this and analyze in detail the differentiability properties of the solution operators to two "proper" EVIs of the second kind in the space $H_0^1(\Omega)$. We hope that the subsequent analysis gives an idea of the peculiar effects that can be encountered, when non-smooth problems in Sobolev spaces are studied, and that we have already glimpsed in Section 3.4.2. Let us again begin with a short overview of the chapter:

Section 5.1 is concerned with the so-called Mosolov problem which arises in non-Newtonian fluid dynamics and which contains the TV-seminorm as a non-smooth functional. Here, we demonstrate that Theorem 4.3.16 can also be used to study EVIs that are more complicated than those in Corollaries 4.3.5, 4.3.6 and 4.3.18, cf. Section 5.1.1. In Sections 5.1.2 to 5.1.5, we further analyze the geometric meaning that the abstract density criterion (4.49) has in the case of the functional $j(v) = \|\nabla v\|_{L^1}$ and derive a sufficient condition for second-order epi-differentiability that is more tangible than (4.49) and that only uses regularity information about the tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ under consideration, see Theorem 5.1.38. The instruments used in the latter context may also be of independent interest, cf., e.g., the Hardy-type result in Corollary 5.1.13. In Section 5.1.6, we conclude our analysis of Mosolov's problem with some comments on FE-approximations.

Section 5.2 is devoted to the study of the functional $j : H_0^1(\Omega) \to \mathbb{R}$, $v \mapsto \|v\|_{L^1}$. After proving that this j falls under the scope of neither Theorem 4.3.3 nor Theorem 4.3.16, cf. Section 5.2.1, we demonstrate in Section 5.2.2 that it is nevertheless possible to obtain second-order epi-differentiability results for the L^1-norm on $H_0^1(\Omega)$ by invoking Lemma 1.3.13 (at least under appropriate regularity assumptions). Here, we will see that distributional curvature effects have to be taken into account when the differential stability of EVIs involving functions of the type $H_0^1(\Omega) \ni v \mapsto \int_\Omega k(v)\mathrm{d}\mathcal{L}^d$ is analyzed. Some of the results proved in this section, e.g., the non-standard Taylor expansion in Corollary 5.2.9, are again also interesting for their own sake.

Lastly, in Section 5.3, we close our discussion with some remarks on the relationship between the findings of Sections 3.4, 5.1 and 5.2.

Note that, throughout this chapter, we make frequent use of standard notations and abbreviations from the theory of Sobolev spaces (most of which we have already encountered in Chapters 1 to 4). For details on the appearing concepts and the precise definitions of the occurring spaces, norms etc., we refer to [Adams, 1975; Attouch et al., 2006].

5.1 Mosolov's Problem and the TV-Seminorm on $\mathrm{H}_0^1(\Omega)$

As a first example of a "proper" elliptic variational inequality of the second kind in the space $H_0^1(\Omega)$, we consider the so-called Mosolov problem

$$w \in H_0^1(\Omega), \quad \int_\Omega \nabla w \cdot \nabla(v-w)\mathrm{d}\mathcal{L}^2 + \int_\Omega \|\nabla v\|_2 \mathrm{d}\mathcal{L}^2 - \int_\Omega \|\nabla w\|_2 \mathrm{d}\mathcal{L}^2 \geq \langle f, v-w\rangle \quad \forall v \in H_0^1(\Omega), \tag{M}$$

that has also been studied, e.g., in [Brezis, 1971; Carstensen et al., 2016; Dean et al., 2007; Ekeland and Temam, 1976; Fuchs and Seregin, 2000; Huilgol and You, 2005; Mosolov and Miasnikov, 1965, 1966, 1967]. Our standing assumptions on the quantities in (M) are as follows:

Assumption 5.1.1 (Standing Assumptions for the Study of the EVI (M))**.**

- $\Omega \subset \mathbb{R}^2$ *is a bounded simply connected (strong) Lipschitz domain,*
- *the spaces* $H_0^1(\Omega)$ *and* $H^{-1}(\Omega)$ *are defined as in [Attouch et al., 2006, Chapter 5],*
- $\mathcal{L}^2$ *is the two-dimensional Lebesgue measure,*
- $\|\cdot\|_2$ *is the Euclidean norm and* $a \cdot b$ *denotes the standard scalar product on* $\mathbb{R}^2$*,*
- $f \in H^{-1}(\Omega)$ *is a given datum.*

Note that the EVI (M) clearly falls under the scope of Chapter 1. This implies in particular that the solution operator $S : H^{-1}(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, associated with (M) is well-defined and globally Lipschitz continuous and that Theorem 1.4.1 is applicable.

5.1.1 Physical Background and Basic Idea Behind the Sensitivity Analysis

Before we turn our attention to the differentiability properties of the solution operator $S : f \mapsto w$, let us briefly comment on possible applications of Mosolov's problem:

An example of a process that is governed by the EVI (M) is the steady-state motion of a viscoplastic medium in a cylindrical pipe of cross-section Ω under no-slip boundary conditions. In this context, the right-hand side $f \in H^{-1}(\Omega)$ represents the pressure gradient/volume force in the direction of the pipe axis that drives the substance under consideration (an example would be the force of gravity), and the solution $w \in H_0^1(\Omega)$ is the velocity orthogonal to the cross-section Ω. Recall that the characteristic feature of a viscoplastic medium is that it behaves like a viscous fluid whenever the internal shear stress exceeds some threshold (the so-called yield point) and that it behaves like a solid otherwise (cf. the EVI of static elastoplasticity studied in Section 4.3.2). In the case of a flow through a pipe, the regions where rigid material behavior occurs are called stagnation zones or nuclei depending on whether they border the boundary $\partial\Omega$ (and thus have velocity zero) or not. If w solves (M) and is continuously differentiable, then the rigid zones are precisely the components of the set $\{\nabla w = 0\}$, cf. Figure 5.1 below.

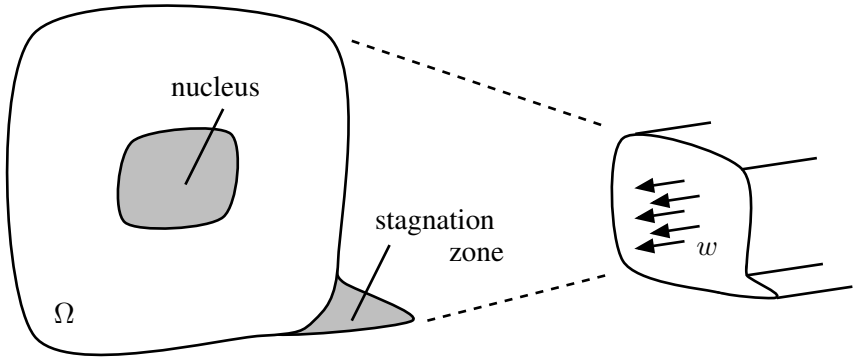

Figure 5.1: Typical flow behavior in the situation of Mosolov's problem with a constant pressure drop f. The viscoplastic medium "sticks" to the boundary in the notch on the bottom right and forms a solid nucleus in the middle of the flow domain that moves with a constant velocity along the pipe axis (cf. [Dean et al., 2007; Mosolov and Miasnikov, 1965, 1966, 1967]).

For details on the fluid-mechanical background of the problem (M), some results on the properties of the stagnation zones and nuclei (for a constant pressure gradient f), and numerical experiments, we refer to [Dean et al., 2007; Fuchs and Seregin, 2000; Mosolov and Miasnikov, 1965, 1966, 1967].

Alternatively to the above interpretation, the EVI (M) can also be identified, e.g., with a regularized TV-denoising problem from image processing. (Recall that $\|\nabla v\|_{L^1} = |v|_{BV}$ holds for all $v \in W^{1,1}(\Omega)$, where $|\cdot|_{BV}$ denotes the total variation seminorm, cf. [Ambrosio et al., 2000, Section 3.1].) Details on this topic may be found in [Calatroni et al., 2015] and [Chan and Shen, 2005]. In the following sections, we will stick to the viewpoint of fluid dynamics to interpret our findings.

To study the differentiability properties of the solution map $S : H^{-1}(\Omega) \to H^1_0(\Omega)$, $f \mapsto w$, to (M), we use the density condition in Theorem 4.3.16 and our main result, Theorem 1.4.1. Note that, due to the continuity of the map

$$L^2(\Omega, \mathbb{R}^2) \to \mathbb{R}, \quad h \mapsto \int_\Omega \|h\|_2 \mathrm{d}\mathcal{L}^2,$$

and since $\nabla \in L(H^1_0(\Omega), L^2(\Omega, \mathbb{R}^2))$, the functional

$$j : H^1_0(\Omega) \to \mathbb{R}, \qquad v \mapsto \int_\Omega \|\nabla v\|_2 \mathrm{d}\mathcal{L}^2, \tag{5.1}$$

fits precisely into the setting of Section 4.3. In particular, Proposition 4.3.2(i) yields

$$\partial j(v) = \left\{ \varphi \in H^{-1}(\Omega) \;\middle|\; \langle \varphi, z \rangle = \int_\Omega \lambda \cdot \nabla z \mathrm{d}\mathcal{L}^2 \;\; \forall z \in H^1_0(\Omega) \text{ for some } \lambda \in L^2(\Omega, \mathbb{R}^2) \right. \\ \left. \text{with } \lambda = \frac{\nabla v}{\|\nabla v\|_2} \text{ a.e. in } \{\nabla v \neq 0\},\ \|\lambda\|_2 \leq 1 \text{ a.e. in } \{\nabla v = 0\} \right\} \tag{5.2}$$

for all $v \in H^1_0(\Omega)$. From Theorems 1.2.2, 1.4.1 and 4.3.16 and the identity of Pythagoras, we may now deduce:

Theorem 5.1.2 (Abstract Differentiability Criterion for Mosolov's Problem)**.** *Suppose that the conditions in Assumption 5.1.1 are satisfied and that the functional j is defined as in* (5.1)*. Then, the solution map $S : H^{-1}(\Omega) \to H^1_0(\Omega)$, $f \mapsto w$, to* (M) *is well-defined and globally Lipschitz continuous, and for every right-hand side $f \in H^{-1}(\Omega)$ with associated solution $w := S(f)$ there exists a multiplier $\lambda \in L^2(\Omega, \mathbb{R}^2)$ such that $\varphi := f + \Delta w = \nabla^* \lambda \in \partial j(w)$ and $\lambda \in \partial \|\cdot\|_2(\nabla w)$ a.e. in Ω. Further, the solution operator S is Hadamard directionally differentiable in all points $f \in H^{-1}(\Omega)$ whose state w and subgradient $\varphi = \nabla^* \lambda$ satisfy*

$$\mathrm{cl}_{\|\cdot\|_\mathcal{K}}(\mathcal{Z}) = \mathcal{K} \tag{5.3}$$

with

$$\mathcal{Z} := \left\{ z \in H^1_0(\Omega) \;\middle|\; \int_{\{\nabla w \neq 0\}} \frac{\|\nabla z\|_2^2}{\|\nabla w\|_2} \mathrm{d}\mathcal{L}^2 < \infty,\ \|\nabla z\|_2 = \lambda \cdot \nabla z \text{ a.e. in } \{\nabla w = 0\} \right\},$$

$$\mathcal{K} := \left\{ z \in H^1_0(\Omega) \;\middle|\; \int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla z)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 < \infty,\ \|\nabla z\|_2 = \lambda \cdot \nabla z \text{ a.e. in } \{\nabla w = 0\} \right\}$$

and

$$\|z\|_\mathcal{K} := \left(\|z\|^2_{H^1} + \int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla z)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \right)^{1/2}, \tag{5.4}$$

and the directional derivatives $\delta := S'(f; g)$, $g \in H^{-1}(\Omega)$, in a point $f \in H^{-1}(\Omega)$ with (5.3) *are uniquely characterized by the EVI*

$$\delta \in \mathcal{K}, \qquad \int_\Omega \nabla \delta \cdot \nabla(z - \delta) \mathrm{d}\mathcal{L}^2 + \int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla \delta)(\nabla w^\perp \cdot \nabla(z - \delta))}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \geq \langle g, z - \delta \rangle \\ \forall z \in \mathcal{K}.$$

Here and in what follows, we use the shorthand notation

$$\begin{pmatrix} a \\ b \end{pmatrix}^{\perp} := \begin{pmatrix} b \\ -a \end{pmatrix} \quad \forall a, b \in \mathbb{R}. \tag{5.5}$$

As Theorem 5.1.2 shows, every condition that is sufficient for the density (5.3) is also sufficient for the Hadamard directional differentiability of the solution map S to (M) in f. The question that arises at this point is, of course, whether (5.3) is a realistic assumption and whether we can find tangible criteria that ensure the equality $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$. Note that the easiest way to guarantee (5.3) would be to assume that $\|\nabla w\|_2 > \varepsilon > 0$ holds a.e. in $\{\nabla w \neq 0\}$ for some $\varepsilon > 0$ (since, in this case, $\mathcal{K}$ and $\mathcal{Z}$ are exactly the same). Doing so, however, turns out to be not very sensible. To see this, we consider:

5.1.2 Mosolov's Problem in the Rotationally Symmetric Case

In this section, we study Mosolov's problem in the rotationally symmetric case to develop some intuition for the abstract differentiability criterion (5.3) in Theorem 5.1.2 and the properties of solutions to the EVI (M). Our precise assumptions are as follows:

Assumption 5.1.3 (Standing Assumptions and Notation for Section 5.1.2)**.**

- $\Omega := \{x \in \mathbb{R}^2 \mid \|x\|_2 < 1\}$,
- $H_0^1(\Omega)$, $H^{-1}(\Omega)$, $\mathcal{L}^2$, $\|\cdot\|_2$ *etc. are defined as before,*
- $r := r(x_1, x_2) := \sqrt{x_1^2 + x_2^2}$ *for all* $x_1, x_2 \in \mathbb{R}^2$,
- *the right-hand side* $f \in H^{-1}(\Omega)$ *in* (M) *satisfies* $f(x_1, x_2) := \mathfrak{f}(r)$ *for some* $\mathfrak{f} \in L^\infty(0,1)$.

The reason why we consider the above setting is that it allows to obtain an explicit formula for the solution w to (M):

Theorem 5.1.4. *Suppose that Assumption 5.1.3 holds, let* $\mathfrak{h} \in C^{0,1}(0,1)$ *be defined by*

$$\mathfrak{h}(r) := \frac{1}{r} \int_0^r s\mathfrak{f}(s)\mathrm{d}s, \tag{5.6}$$

and let $\mathfrak{w}$ *be the unique element of* $C^{1,1}(0,1)$ *with*

$$\mathfrak{w}(1) = 0, \qquad \mathfrak{w}'(r) = \max(0, -\mathfrak{h}(r) - 1) + \min(0, -\mathfrak{h}(r) + 1) \quad \text{in } (0,1). \tag{5.7}$$

Then, the solution $w = S(f)$ *to the problem* (M) *is given by* $w(x_1, x_2) = \mathfrak{w}(r)$.

Proof. Since the right-hand side f and the domain Ω are invariant under orthogonal transformations and since the solution to (M) is unique, the function $w = S(f)$ has to satisfy $w(Rx) = w(x)$ for a.a. $x \in \Omega$ and all $R \in \mathrm{O}(2)$, and from [Brezis, 1971, Example 2, Theorem 15] and the Sobolev embeddings, we readily obtain that $w \in H^2(\Omega) \subset C(\mathrm{cl}(\Omega))$. Using both these properties of w, we may deduce that $w(x_1, x_2) = w(r, 0) =: \mathfrak{w}(r)$ holds for all $(x_1, x_2) \in \Omega$ (where w denotes the continuous representative). Since $w(r, 0)$ is precisely the trace of the H^2-function w on the manifold $(0,1) \times \{0\}$, we may invoke, e.g., [Adams, 1975, Theorem 5.4 Part IA] to infer that $\mathfrak{w} \in H^1(0,1)$. Further, a simple mollification argument and the definition of $\mathfrak{w}$ yield

$$(\nabla w)(x_1, x_2) = [w(r, 0)]' e_r = \mathfrak{w}'(r) e_r \quad \text{for a.a. } (x_1, x_2) \in \Omega, \tag{5.8}$$

where a prime denotes the weak derivative w.r.t. r and where $e_r := (x_1, x_2)/\|(x_1, x_2)\|_2$ is the vector in radial direction. Recall now that the function w is also the unique solution of the minimization problem

$$\min_{v \in H_0^1(\Omega)} \int_\Omega \frac{1}{2}\|\nabla v\|_2^2 + \|\nabla v\|_2 - fv\mathrm{d}\mathcal{L}^2, \tag{5.9}$$

cf. the comments in Section 1.2. This implies in combination with (5.8) that $\mathfrak{w} \in H^1(0,1)$ is the (due to the convexity necessarily unique) solution to

$$\min_{\mathfrak{v}\in H^1(0,1),\ \mathfrak{v}(1)=0} \int_0^1 \left(\frac{1}{2}\mathfrak{v}'(r)^2 + |\mathfrak{v}'(r)| - \mathfrak{f}(r)\mathfrak{v}(r)\right) r\mathrm{d}r. \tag{5.10}$$

If we rewrite (5.10) as a problem in $\mathfrak{u} := \mathfrak{v}' \in L^2(0,1)$ and use integration by parts, then we obtain that

$$\mathfrak{w}' = \operatorname*{argmin}_{\mathfrak{u}\in L^2(0,1)} \int_0^1 \left(\frac{1}{2}\mathfrak{u}(r)^2 + |\mathfrak{u}(r)| + \mathfrak{u}(r)\mathfrak{h}(r)\right) r\mathrm{d}r,$$

where $\mathfrak{h}$ is defined as in (5.6). Since the above minimization problem is entirely pointwise, we may minimize pointwise to deduce that

$$\mathfrak{w}'(r) = \begin{cases} 0 & \text{a.e. in } \{|\mathfrak{h}| \le 1\} \\ -\mathfrak{h}(r) + \operatorname{sgn}(\mathfrak{h}(r)) & \text{a.e. in } \{|\mathfrak{h}| > 1\}. \end{cases} \tag{5.11}$$

The formula (5.7) now follows immediately. To see that the function $\mathfrak{h}$ is Lipschitz and that $\mathfrak{w}$ is in $C^{1,1}(0,1)$, it suffices to note that $\mathfrak{h}$ possesses a bounded weak derivative by (5.6) and to invoke (5.7). This completes the proof. □

From the explicit formula (5.7) and the identities in the proof of Theorem 5.1.4, we obtain, e.g.:

Corollary 5.1.5. *In the situation of Theorem 5.1.4, the gradient ∇w of the solution $w = S(f)$ to* (M) *vanishes almost everywhere in a ball around the origin of radius*

$$r(\mathfrak{f}) := \sup\left\{ r \in (0,1) \;\middle|\; \left|\frac{1}{\rho}\int_0^\rho s\mathfrak{f}(s)\mathrm{d}s\right| \le 1 \text{ for all } \rho \in (0,r] \right\} \ge \min\left(1, \frac{2}{\|\mathfrak{f}\|_{L^\infty}}\right) > 0.$$

Proof. From (5.7), (5.8) and Hölder's inequality, we obtain that $\mathfrak{w}' = 0$ holds a.e. in $\{|\mathfrak{h}| \le 1\}$, that the gradient of w satisfies $(\nabla w)(x_1,x_2) = \mathfrak{w}'(r)e_r$, and that $|\mathfrak{h}(\rho)| \le \|\mathfrak{f}\|_{L^\infty}\rho/2$ for all $\rho \in (0,1)$. The claim now follows immediately. □

Corollary 5.1.6. *In the situation of Theorem 5.1.4, the solution $w = S(f)$ is an element of $C^{1,1}(\Omega)$. Moreover, for every $\mathfrak{f} \in C([0,1])$ with $w \not\equiv 0$ and $\|\mathfrak{f}\|_{L^\infty}/2 < \mathfrak{f}$ in $(0,1)$, it holds $w \in C^{1,1}(\Omega) \setminus C^2(\Omega)$.*

Proof. The $C^{1,1}$-regularity of the solution w follows straightforwardly from $(\nabla w)(x_1,x_2) = \mathfrak{w}'(r)e_r$ and $\mathfrak{w}' \in C^{0,1}(0,1)$ (note that we do not have to worry about the derivatives of $r = r(x_1,x_2)$ at the origin here since $\mathfrak{w}'$ vanishes everywhere in $(0, r(\mathfrak{f}))$). It remains to prove that w is not in $C^2(\Omega)$ for right-hand sides $\mathfrak{f}$ that satisfy the conditions in the second part of the corollary. To see this, we note that $w \not\equiv 0$ implies $\tilde{r} := r(\mathfrak{f}) \in (0,1)$, that our assumption $\mathfrak{f} \in C([0,1])$ yields $\mathfrak{h} \in C^1(0,1)$, and that w can only be twice continuously differentiable if $\mathfrak{h}'(\tilde{r}) = 0$, i.e., if

$$0 = -\frac{1}{\tilde{r}^2}\int_0^{\tilde{r}} s\mathfrak{f}(s)\mathrm{d}s + \mathfrak{f}(\tilde{r}) = -\frac{1}{\tilde{r}}\mathfrak{h}(\tilde{r}) + \mathfrak{f}(\tilde{r}), \tag{5.12}$$

cf. (5.11). Since $\tilde{r}$ satisfies $|\mathfrak{h}(\tilde{r})| = 1$ and $\tilde{r} \ge 2/\|\mathfrak{f}\|_{L^\infty}$ by Corollary 5.1.5, (5.12) entails

$$|\mathfrak{f}(\tilde{r})| = \frac{1}{\tilde{r}} \le \frac{1}{2}\|\mathfrak{f}\|_{L^\infty}.$$

This contradicts our assumption $\|\mathfrak{f}\|_{L^\infty}/2 < \mathfrak{f}$ in $(0,1)$ and proves that w cannot be a C^2-function for right-hand sides $\mathfrak{f}$ that satisfy the conditions $\mathfrak{f} \in C([0,1])$, $S(f) \not\equiv 0$ and $\|\mathfrak{f}\|_{L^\infty}/2 < \mathfrak{f}$. □

We can now make the following observations:

Remark 5.1.7.

(i) In the context of fluid mechanics, Corollary 5.1.5 expresses that a viscoplastic medium in a cylindrical pipe with a circular cross-section of radius one can only flow when the pressure gradient/ volume force $\mathfrak{f}$ is greater than two somewhere in the fluid domain Ω and that there is always a solid nucleus in the middle of the pipe when $\mathfrak{f}$ is bounded. We remark that the existence of such a nucleus can also be proved for pipes whose cross-sections Ω are strongly symmetric in the sense of [Mosolov and Miasnikov, 1965], see ibid. Theorem 2.6.

(ii) Corollary 5.1.6 shows that the solution w to (M) *can be expected to be (Lipschitz) continuously differentiable when the domain Ω and the right-hand side f are sufficiently smooth. (Compare also with the regularity results in [Brezis, 1971; Fuchs and Seregin, 1998, 2000] in this context.) This implies in particular that it does not make much sense to work with an assumption of the form $\|\nabla w\|_2 > \varepsilon > 0$ a.e. in $\{\nabla w \neq 0\}$ to check the density condition* (5.3) *in Theorem 5.1.2 and that the weight function $\|\nabla w\|_2^{-1}$ in the definition of the norm $\|\cdot\|_{\mathcal{K}}$ in* (5.4) *typically blows up in the vicinity of the set $\{\nabla w = 0\}$. Here and in what follows, when using the shorthand*

$$\{v \text{ satisfies a pointwise condition } P\} := \{x \in \Omega \mid v \text{ satisfies } P \text{ at } x\} \tag{5.13}$$

for a function $v \in C(\Omega, \mathbb{R}^d) \cap L^1(\Omega, \mathbb{R}^d)$, we always assume that the continuous representative of v is chosen so that (5.13) *is defined in the classical sense (and not just up to sets of measure zero) and so that the notions of vicinity, neighborhood, distance and boundary are well-defined.*

(iii) The second part of Corollary 5.1.6 is remarkable because it yields that the solution w to (M) *does not(!) have a certain regularity when the right-hand side $\mathfrak{f}$ is sufficiently well-behaved. (Note that the condition $\|\mathfrak{f}\|_{L^\infty}/2 < \mathfrak{f}$ in $(0,1)$ is an assumption on the maximum oscillation of $\mathfrak{f}$.) Corollary 5.1.6 further implies that we cannot reasonably expect more than $C^{1,1}$-regularity when we study the solution w to* (M)*. We point out that this behavior accords very well with the intuition that the term $\int_\Omega \|\nabla v\|_2 \mathrm{d}\mathcal{L}^2$ causes the solution of the minimization problem* (5.9) *to have "kinks" in the first derivative, cf. [Casas et al., 2012].*

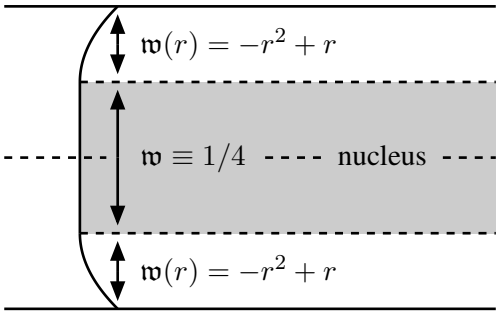

Figure 5.2: Flow profile in the situation of Assumption 5.1.3 for the constant pressure drop $\mathfrak{f} \equiv 4$.

5.1.3 The Integrability Condition in the Two-Dimensional Setting

As the results of the last section show, in the situation of Assumption 5.1.1 and Theorem 5.1.2, we have to expect that the solution w to (M) is in $C^{1,1}(\Omega)$ and that the term $\|\nabla w\|_2^{-1}$ satisfies

$$\frac{1}{\|\nabla w\|_2} \geq \frac{1}{\|w\|_{C^{1,1}} \operatorname{dist}(\cdot, \{\nabla w = 0\})} \quad \text{in } \{\nabla w \neq 0\},$$

where dist denotes the Euclidean distance to a set and where ∇w is the continuous representative of the gradient. This is rather unfortunate because it implies that $\|\cdot\|_{\mathcal{K}}$ is a "proper" weighted norm which includes a singular term and that verifying (5.3) amounts to proving the density of the set $\mathcal{Z}$ in $\mathcal{K}$ in the weighted Sobolev space

$$\big\{z \in H^1_0(\Omega) \;\big|\; \|z\|_{\mathcal{K}} < \infty\big\}.$$

Checking such density conditions is in general hard and, at least to the author's best knowledge, there is currently no closed theory that can be used to prove an equality of the type (5.3) in an arbitrary weighted Sobolev space. Even the study of necessary and sufficient conditions for the famous identity $H = W$ in weighted $W^{1,q}$-spaces is nowadays still an active field of research, cf. [Ambrosio et al., 2014; Kufner, 1980; Piat and Cassano, 1994; Surnachev, 2014; Zhikov, 1998, 2013]. (Note that classical instruments like the standard "mollification by convolution" technique of Friedrichs are typically inapplicable in this context because they do not necessarily preserve additional integrability conditions.) Moreover, the overwhelming majority of the literature focuses on norms of the type

$$\|z\| := \left(\int_\Omega \big(|z|^q + \|\nabla z\|_2^q\big)\rho \mathrm{d}\mathcal{L}^d\right)^{1/q},$$

where ρ is a non-negative function in $L^1(\Omega)$. Weights that affect only a component of the gradient field as in (5.4) have apparently not been considered so far.

However, the special geometric structure of the norm $\|\cdot\|_{\mathcal{K}}$ in Theorem 5.1.2 can also be turned into an advantage. To see this, we note that the integrability condition in (5.4) can be rewritten as

$$\int_{\{\nabla w \neq 0\}} \frac{1}{\|\nabla w\|_2}\left(\frac{\nabla w^\perp}{\|\nabla w\|_2}\cdot \nabla z\right)^2 \mathrm{d}\mathcal{L}^2 < \infty. \tag{5.14}$$

The above implies that, for a C^1-solution w, the condition in our weighted Sobolev space penalizes gradients that are not parallel or anti-parallel to the unit vector field $\nabla w/\|\nabla w\|_2$ in the vicinity of the set $\{\nabla w = 0\}$. Since $\nabla w/\|\nabla w\|_2$ is precisely the normal to the level sets $\{w = c\}$, $c \in \mathbb{R}$, in $\{\nabla w \neq 0\}$, this means that functions z which satisfy (5.14) can be expected to have level sets which locally resemble those of w near the rigid zone $\{\nabla w = 0\}$ and traces which are (locally) constant on the boundary $\partial\{\nabla w = 0\}$, cf. Figure 5.3.

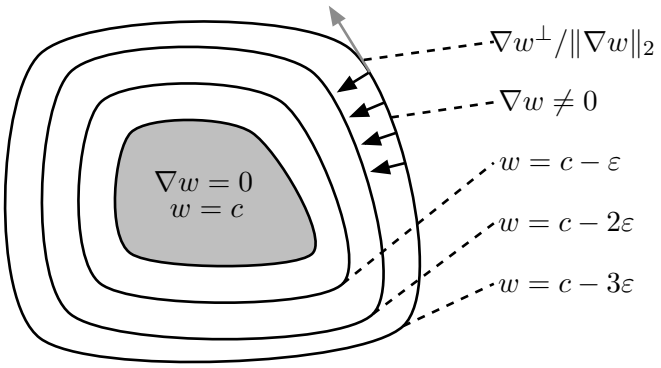

Figure 5.3: The situation near the set $\{\nabla w = 0\}$.

Note that the condition "$\mathrm{tr}(z) = \mathrm{const}$ on all components of $\partial\{\nabla w = 0\}$ for all $z \in \mathcal{K}$" is necessary for the density of the set

$$\Big\{z \in H^1_0(\Omega) \;\Big|\; \nabla z = 0 \text{ a.e. in } \Omega \setminus (\{\nabla w = 0\} \cup E) \text{ for some compact set } E \subset \{\nabla w \neq 0\},\\ \|\nabla z\|_2 = \lambda \cdot \nabla z \text{ a.e. in } \{\nabla w = 0\}\Big\} \subset \mathcal{Z} \tag{5.15}$$

in $\mathcal{K}$ w.r.t. $\|\cdot\|_{\mathcal{K}}$ for a C^1-solution w and that the set on the left-hand side of (5.15) is far easier to handle than the set $\mathcal{Z}$ itself. Proving the constantness of the traces of functions z with (5.14) on $\partial\{\nabla w = 0\}$ is thus indeed a first step towards verifying the differentiability criterion (5.4).

In the remainder of this section, our aim will be to make the above informal argumentation rigorous. To this end, we recall the following (variants of) classical results:

Lemma 5.1.8 (Integration by Parts on Level Sets)**.** *Suppose that $D \subset \mathbb{R}^2$ is a non-empty, open, bounded set and assume that a $v \in C(D)$ and a $c \in \mathbb{R}$ with $\{v = c\} \neq \emptyset$ are given such that there exists an $\varepsilon > 0$ with*

$$v \in C^1(\{|v - c| < \varepsilon\}) \quad \textit{and} \quad \|\nabla v\|_2 > 0 \textit{ in } \{|v - c| < \varepsilon\}.$$

Then, for all $z_1 \in C^1_c(D)$, $z_2 \in C^1(D)$, it holds

$$\int_{\{v=c\}} z_1 \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_2 \right) \mathrm{d}\mathcal{H}^1 = - \int_{\{v=c\}} z_2 \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_1 \right) \mathrm{d}\mathcal{H}^1.$$

Here, $\mathcal{H}^1$ is the one-dimensional Hausdorff measure and $\nabla v^\perp$ is defined as in (5.5)*.*

Proof. Under the assumptions of the lemma, the set $\{|v - c| < \varepsilon\} = \{x \in D \mid |v(x) - c| < \varepsilon\}$ is an open subset of D, and the set $\{v = c\}$ is a one-dimensional (embedded) C^1-submanifold of $\mathbb{R}^2$, cf. [Holm et al., 2009, Theorem 2.32]. The latter implies that for every $p \in \{v = c\}$ we can find an open set $O \subset D$, an open interval I and a C^1-chart $\theta : I \to O$ such that

$$p \in \theta(I) = \{v = c\} \cap O \quad \text{and} \quad \|\theta'(t)\|_2 \geq 1 \quad \forall t \in I.$$

Note that

$$\frac{\theta'}{\|\theta'\|_2} \cdot \frac{(\nabla v^\perp)(\theta)}{\|(\nabla v)(\theta)\|_2} = \text{const} =: \sigma \in \{\pm 1\} \quad \text{in } I.$$

Consider now two arbitrary but fixed functions $z_1 \in C^1_c(D)$, $z_2 \in C^1(D)$. Since we can always use a partition of unity to localize the problem (using that the function z_1 has compact support), we may assume w.l.o.g. that $\mathrm{supp}(z_1)$ is contained in an open set O with an associated parametrization $\theta : I \to O$. In this case, we may calculate

$$\begin{aligned}
\int_{\{v=c\}} z_1 \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_2 \right) \mathrm{d}\mathcal{H}^1 &= \int_I z_1(\theta(t)) \left(\frac{(\nabla v^\perp)(\theta(t))}{\|(\nabla v)(\theta(t))\|_2} \cdot (\nabla z_2)(\theta(t)) \right) \|\theta'(t)\|_2 \mathrm{d}t \\
&= \sigma \int_I z_1(\theta(t))(\nabla z_2)(\theta(t)) \cdot \theta'(t) \mathrm{d}t \\
&= -\sigma \int_I z_2(\theta(t))(\nabla z_1)(\theta(t)) \cdot \theta'(t) \mathrm{d}t \\
&= - \int_{\{v=c\}} z_2 \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_1 \right) \mathrm{d}\mathcal{H}^1.
\end{aligned}$$

This proves the claim. □

Lemma 5.1.9 (Integration by Parts in $H(\mathrm{div}; D)$)**.** *Suppose that $D \subset \mathbb{R}^d$ is a bounded Lipschitz domain and assume that a vector field $F \in H(\mathrm{div}; D)$ is given such that there exists a compact set $E \subset D$ with $F|_{D \setminus E} \in H^1(D \setminus E, \mathbb{R}^d) \cap C(\mathrm{cl}(D \setminus E), \mathbb{R}^d)$. Then, for all $z \in H^1(D)$, it is true that*

$$\int_D z(\nabla \cdot F) + \nabla z \cdot F \mathrm{d}\mathcal{L}^d = \int_{\partial D} \mathrm{tr}(z)(F \cdot \nu) \mathrm{d}\mathcal{H}^{d-1}, \tag{5.16}$$

where $\nu : \partial D \to \mathbb{R}^d$ denotes the outward unit normal vector field on ∂D.

Here and in what follows, $H(\mathrm{div}; D)$ denotes the space of all L^2-vector fields whose distributional divergence is in $L^2(D)$, i.e.,

$$H(\mathrm{div}; D) := \Big\{ v \in L^2(D, \mathbb{R}^d) \;\Big|\; \nabla \cdot v \in L^2(D) \Big\}. \tag{5.17}$$

Note that $H(\mathrm{div}; D)$ is a Hilbert space when equipped with the norm

$$\|v\|_{H(\mathrm{div};D)} := \left(\|v\|^2_{L^2(D,\mathbb{R}^d)} + \|\nabla \cdot v\|^2_{L^2(D)} \right)^{1/2}.$$

See, e.g., [Girault and Raviart, 1986, Section 2.2] for details.

Proof of Lemma 5.1.9. Choose $\psi_1, \psi_2 \in C_c^\infty(\mathbb{R}^d)$ such that $\mathrm{supp}(\psi_1) \subset D$, $\mathrm{supp}(\psi_2) \subset \mathbb{R}^d \setminus E$, and $\psi_1 + \psi_2 \equiv 1$ in a neighborhood of $\mathrm{cl}(D)$. Then, it holds

$$\begin{aligned} &\int_D z(\nabla \cdot F) + \nabla z \cdot F \mathrm{d}\mathcal{L}^d \\ &\quad = \int_D z(\nabla \cdot (\psi_1 F))\mathrm{d}\mathcal{L}^d + \int_D \nabla z \cdot (\psi_1 F)\mathrm{d}\mathcal{L}^d + \int_D z(\nabla \cdot (\psi_2 F))\mathrm{d}\mathcal{L}^d + \int_D \nabla z \cdot (\psi_2 F)\mathrm{d}\mathcal{L}^d \end{aligned} \tag{5.18}$$

for all $z \in H^1(D)$. Since $\psi_2 F$ is an element of $H^1(D, \mathbb{R}^d) \cap C(\mathrm{cl}(D), \mathbb{R}^d)$ by our assumptions, we may use the density of the set $\{\phi|_D \mid \phi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)\}$ in $H^1(D, \mathbb{R}^d)$ and [Evans and Gariepy, 2015, Theorem 4.6] to obtain

$$\int_D z(\nabla \cdot (\psi_2 F))\mathrm{d}\mathcal{L}^d + \int_D \nabla z \cdot (\psi_2 F)\mathrm{d}\mathcal{L}^d = \int_{\partial D} \mathrm{tr}(z)(F \cdot \nu)\mathrm{d}\mathcal{H}^{d-1}. \tag{5.19}$$

From the density of the set $\{\phi|_D \mid \phi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)\}$ in $H(\mathrm{div}; D)$ (see [Girault and Raviart, 1986, Theorem 2.4]), the fact that the map $H(\mathrm{div}; D) \ni G \mapsto \nu \cdot G \in H^{-1/2}(\partial D)$ is continuous (see [Girault and Raviart, 1986, Theorem 2.5]), and again [Evans and Gariepy, 2015, Theorem 4.6], it follows further that

$$\int_D z(\nabla \cdot (\psi_1 F))\mathrm{d}\mathcal{L}^d + \int_D \nabla z \cdot (\psi_1 F)\mathrm{d}\mathcal{L}^d = 0. \tag{5.20}$$

If we combine (5.18), (5.19) and (5.20), then the claim follows immediately. □

Lemma 5.1.10 (Coarea Formula on Subsets)**.** *Let $D \subset \mathbb{R}^d$ be a non-empty, open, bounded set and suppose that a Lipschitz function $v : D \to \mathbb{R}$ and a Lebesgue measurable and absolutely integrable $z : D \to \mathbb{R}$ are given. Then, it holds*

$$\int_D z \|\nabla v\|_2 \mathrm{d}\mathcal{L}^d = \int_{\mathbb{R}} \left(\int_{\{v=s\}} z \mathrm{d}\mathcal{H}^{d-1} \right) \mathrm{d}s.$$

Proof. From Kirszbraun's theorem (see [Evans and Gariepy, 2015, Theorem 3.1]), it follows that v can be extended to a Lipschitz continuous function $\tilde{v} : \mathbb{R}^d \to \mathbb{R}$, and by defining $\tilde{z}(x) := z(x)$ for $x \in D$, $\tilde{z}(x) := 0$ for $x \in \mathbb{R}^d \setminus D$, we obtain a measurable and absolutely integrable extension $\tilde{z} : \mathbb{R}^d \to \mathbb{R}$ of z. Applying the classical coarea formula (as found in [Evans and Gariepy, 2015, Theorem 3.11] or [Attouch et al., 2006, Theorem 4.2.5]) to $\tilde{v}$ and $\tilde{z}$ yields

$$\begin{aligned} \int_D z \|\nabla v\|_2 \mathrm{d}\mathcal{L}^d &= \int_{\mathbb{R}^d} \tilde{z} \|\nabla \tilde{v}\|_2 \mathrm{d}\mathcal{L}^d \\ &= \int_{\mathbb{R}} \left(\int_{\{x \in \mathbb{R}^d \,|\, \tilde{v}(x) = s\}} \tilde{z} \mathrm{d}\mathcal{H}^{d-1} \right) \mathrm{d}s = \int_{\mathbb{R}} \left(\int_{\{x \in D \,|\, v(x) = s\}} z \mathrm{d}\mathcal{H}^{d-1} \right) \mathrm{d}s. \end{aligned}$$

This proves the claim. □

We are now in the position to prove that the integrability condition in (5.4) indeed causes the function z to have a constant trace on the boundary $\partial\{\nabla w = 0\}$ (provided w is sufficiently regular):

Proposition 5.1.11. *Let $D \subset \mathbb{R}^2$ be a non-empty, open, bounded set. Suppose that a $v \in C^{0,1}(D)$, a $c \in \mathbb{R}$ with $\{v = c\} \neq \emptyset$ and an $\varepsilon > 0$ are given such that*

$$\|\nabla v\|_2 \in C(\{|v-c| < \varepsilon\}), \qquad v \in C^1(\{0 < |v-c| < \varepsilon\}),$$
$$\|\nabla v\|_2 > 0 \text{ in } \{0 < |v-c| < \varepsilon\}.$$

Assume further that two non-empty, open, bounded intervals I and J and a Lipschitz continuous C^1-map $\gamma : I \to J$ are given such that $W := I \times J$ satisfies

$$W \subset \{|v-c| < \varepsilon\} \quad \text{and} \quad W \cap \partial\{v = c\} = \{(x, \gamma(x)) \mid x \in I\} \subset \{\|\nabla v\|_2 = 0\},$$

and such that

$$\int_{\{0<|v-c|<t\}\cap W} \psi^2 \|\nabla v\|_2^3 \mathrm{d}\mathcal{L}^2 = O(t^2) \quad \text{as} \quad t \searrow 0 \tag{5.21}$$

holds for all $\psi \in C_c^1(W)$. Then, for every $z \in W^{1,1}(D)$ with

$$\int_{\{0<|v-c|<\varepsilon\}} \frac{(\nabla v^\perp \cdot \nabla z)^2}{\|\nabla v\|_2^3} \mathrm{d}\mathcal{L}^2 < \infty, \tag{5.22}$$

it holds $\mathrm{tr}(z) = \mathrm{const}$ on $W \cap \partial\{v = c\}$.

Proof. The proof is based on Lemmas 5.1.8, 5.1.9 and 5.1.10. First, we note that, in the situation under consideration, at least one of the open sets

$$W_+ := W \cap \{(x,y) \mid x \in I, \gamma(x) < y\}, \qquad W_- := W \cap \{(x,y) \mid x \in I, \gamma(x) > y\}$$

has to be contained entirely in $\{c < v < c + \varepsilon\}$ or $\{c - \varepsilon < v < c\}$. If this is not the case, then we get a contradiction with $W \cap \partial\{v = c\} = \{(x, \gamma(x)) \mid x \in I\}$. In what follows, we assume w.l.o.g. that $W_+ \subset \{c < v < c + \varepsilon\}$. The other cases are analogous. Consider now an arbitrary but fixed $\psi \in C_c^\infty(W)$ and let $z_m \in C^\infty(D) \cap W^{1,1}(D)$ be a sequence with $z_m \to z$ in $W^{1,1}(D)$ for $m \to \infty$. Then, for all $s \in (0, \varepsilon)$, it is true that

$$\begin{aligned}
&\partial\big(\{c < v < c + s\} \cap W_+\big) \cap \mathrm{supp}(\psi) \\
&\quad \subset \big(\partial\{c < v < c + s\} \cup \partial W_+\big) \cap \mathrm{supp}(\psi) \\
&\quad \subset \big(\partial\{c < v < c + s\} \cap \mathrm{supp}(\psi)\big) \cup \big(\partial\{v = c\} \cap \mathrm{supp}(\psi)\big) \\
&\quad \subset \big(\partial\{v = c + s\} \cap \mathrm{supp}(\psi)\big) \cup \big(\partial\{v = c\} \cap \mathrm{supp}(\psi)\big).
\end{aligned} \tag{5.23}$$

Note that our assumptions imply that $\partial\{v = c\} \cap \mathrm{supp}(\psi)$ is a subset of the C^1-graph $W \cap \partial\{v = c\}$ and that $\partial\{v = c + s\} \cap W = \{v = c + s\} \cap W$ is a one-dimensional C^1-submanifold. Using these properties, (5.23), a localization with a partition of unity and the integration by parts formula (5.16), we obtain

$$\begin{aligned}
&\int_{\{c<v<c+s\}\cap W_+} z_m \nabla \cdot \Big(\nabla \psi^\perp\Big) + \nabla z_m \cdot \nabla \psi^\perp \mathrm{d}\mathcal{L}^2 \\
&\quad = \int_{\{c<v<c+s\}\cap W_+} \nabla z_m \cdot \nabla \psi^\perp \mathrm{d}\mathcal{L}^2 \\
&\quad = \int_{\{v=c+s\}\cap W_+} z_m (\nu \cdot \nabla \psi^\perp) \mathrm{d}\mathcal{H}^1 + \int_{\partial\{v=c\}\cap W} z_m (\nu \cdot \nabla \psi^\perp) \mathrm{d}\mathcal{H}^1 \\
&\quad = -\int_{\{v=c+s\}\cap W_+} z_m (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1 - \int_{\partial\{v=c\}\cap W} z_m (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1,
\end{aligned} \tag{5.24}$$

where ν again denotes the outward unit normal vector. Since ν equals $\nabla v/\|\nabla v\|_2$ on $\{v = c+s\} \cap W_+$, we may use Lemma 5.1.8 to infer (for s sufficiently small)

$$\begin{aligned}
&\int_{\{c<v<c+s\}\cap W_+} \nabla z_m \cdot \nabla \psi^\perp \mathrm{d}\mathcal{L}^2 \\
&\quad = -\int_{\{v=c+s\}\cap W_+} z_m \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla \psi \right) \mathrm{d}\mathcal{H}^1 - \int_{\partial\{v=c\}\cap W} z_m (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1 \\
&\quad = \int_{\{v=c+s\}\cap W_+} \psi \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_m \right) \mathrm{d}\mathcal{H}^1 - \int_{\partial\{v=c\}\cap W} z_m (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1.
\end{aligned}$$

Integrating the above w.r.t. s over $(0,t)$, $t>0$ sufficiently small, and employing Lemma 5.1.10 yields

$$\begin{aligned}
&t \left| \int_{\partial\{v=c\}\cap W} z_m (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1 \right| \\
&\leq \left| \int_0^t \int_{\{c<v<c+s\}\cap W_+} \nabla z_m \cdot \nabla \psi^\perp \mathrm{d}\mathcal{L}^2 - \int_{\{v=c+s\}\cap W_+} \psi \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_m \right) \mathrm{d}\mathcal{H}^1 \mathrm{d}s \right| \\
&\leq t \int_{\{c<v<c+t\}\cap W_+} \left| \nabla z_m \cdot \nabla \psi^\perp \right| \mathrm{d}\mathcal{L}^2 + \int_0^t \int_{\{v=c+s\}} \left| \psi \left(\frac{\nabla v^\perp}{\|\nabla v\|_2} \cdot \nabla z_m \right) \right| \mathrm{d}\mathcal{H}^1 \mathrm{d}s \\
&\leq t \int_{\{c<v<c+t\}} \left| \nabla z_m \cdot \nabla \psi^\perp \right| \mathrm{d}\mathcal{L}^2 + \int_{\{c<v<c+t\}\cap W} |\psi| \left| \nabla v^\perp \cdot \nabla z_m \right| \mathrm{d}\mathcal{L}^2.
\end{aligned}$$

Note that, by mollification (cf. [Evans, 2010, Theorem C5.7]), we obtain that the above also holds for all $\psi \in C_c^1(W)$. We may thus relax our assumption $\psi \in C_c^\infty(W)$ accordingly. Passing to the limit $m \to \infty$, dividing by t, using the Cauchy-Schwarz inequality and taking into account our assumptions now yields

$$\begin{aligned}
&\left| \int_{\partial\{v=c\}\cap W} \mathrm{tr}(z) (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1 \right| \\
&\leq \int_{\{c<v<c+t\}} \left| \nabla z \cdot \nabla \psi^\perp \right| \mathrm{d}\mathcal{L}^2 \\
&\quad + \left(\frac{1}{t^2} \int_{\{c<v<c+t\}\cap W} \psi^2 \|\nabla v\|_2^3 \mathrm{d}\mathcal{L}^2 \right)^{1/2} \left(\int_{\{c<v<c+t\}\cap W} \frac{\left(\nabla v^\perp \cdot \nabla z\right)^2}{\|\nabla v\|_2^3} \mathrm{d}\mathcal{L}^2 \right)^{1/2} = o(1),
\end{aligned}$$

where the Landau symbol refers to the limit $t \searrow 0$. This implies

$$\int_{\partial\{v=c\}\cap W} \mathrm{tr}(z) (\nu^\perp \cdot \nabla \psi) \mathrm{d}\mathcal{H}^1 = 0 \quad \forall \psi \in C_c^1(W).$$

Consider now the function

$$\Phi : W \to \mathbb{R}^2, \quad (x,y) \mapsto (x, y - \gamma(x)).$$

Then, Φ is a diffeomorphism onto its image with the inverse

$$\Phi^{-1} : \Phi(W) \to W, \quad (x,y) \mapsto (x, y + \gamma(x)),$$

and we may deduce that

$$\begin{aligned}
&\int_I \mathrm{tr}(z)(x,\gamma(x)) \left(\begin{pmatrix} 1 \\ \gamma'(x) \end{pmatrix} \cdot (\nabla \psi)(x,\gamma(x)) \right) \mathrm{d}x \\
&\qquad = \int_I \mathrm{tr}(z)(x,\gamma(x)) \frac{\mathrm{d}}{\mathrm{d}x} \Big(\psi(x,\gamma(x)) \Big) \mathrm{d}x = 0
\end{aligned}$$

holds for all $\psi \in C_c^1(W)$. On the other hand, for every $\phi \in C_c^1(I)$, we can easily construct a function $\psi \in C_c^1(W)$ with $\psi(x, \gamma(x)) = \phi(x)$ in I (just extend ϕ suitably onto $\mathbb{R}^2$ and use the transformation Φ). We thus end up with

$$\int_I \operatorname{tr}(z)(x, \gamma(x))\phi'(x)\mathrm{d}x = 0 \quad \forall \phi \in C_c^1(I).$$

This proves that $\operatorname{tr}(z)$ is indeed constant on $W \cap \partial\{v = c\}$ and yields the claim. □

Remark 5.1.12. *The condition*

$$\int_{\{0<|v-c|<t\}\cap W} \psi^2 \|\nabla v\|_2^3 \mathrm{d}\mathcal{L}^2 = O(t^2) \quad \textit{as} \quad t \searrow 0 \qquad \forall \psi \in C_c^1(W) \tag{5.25}$$

in Proposition 5.1.11 is an assumption on how fast the function values of v and $\|\nabla v\|_2$ converge to c and zero, respectively, when one approaches the critical level set $\{v = c\}$. Note that it makes sense that such a condition is needed for Proposition 5.1.11 to hold since, for general v, it is perfectly possible that the level sets $\{v = c \pm s\}$ and the associated normal vector fields $\nabla v / \|\nabla v\|_2$ do not provide any information about the shape of the boundary $\partial\{v = c\}$. We will see in Lemma 5.1.29 that a $C^{1,1}$-solution w of (M) *with a right-hand side $f \in L^\infty(\Omega)$ always satisfies* (5.25)*.*

We would like to point out that Proposition 5.1.11 is not only a handy tool in the sensitivity analysis of Mosolov's problem but also interesting for its own sake. It yields, for example:

Corollary 5.1.13 (A Sufficient Criterion for Constant Traces)**.** *Let $D \subset \mathbb{R}^2$ be a non-empty, open, bounded set, and let $\mathcal{M} \subset D$ be a compact, one-dimensional C^2-submanifold of $\mathbb{R}^2$. Suppose that a $z \in W^{1,1}(D)$ with*

$$\int_{D\setminus\mathcal{M}} \frac{(\nabla d_{\mathcal{M}}^{\perp} \cdot \nabla z)^2}{d_{\mathcal{M}}} \mathrm{d}\mathcal{L}^2 < \infty \tag{5.26}$$

is given, where $d_{\mathcal{M}} := \operatorname{dist}(\cdot, \mathcal{M})$ denotes the Euclidean distance to $\mathcal{M}$. Then, the trace of z is constant on each connected component of $\mathcal{M}$.

Proof. Define $v := d_{\mathcal{M}}^2$. Then, v is an element of $C^{0,1}(D)$ and we obtain from [Foote, 1984, Theorem 1] and the proof of [Evans and Gariepy, 2015, Theorem 3.14] that there exists an $\varepsilon > 0$ with

$$\|\nabla v\|_2 = 2d_{\mathcal{M}} \in C^{0,1}(D), \quad v \in C^1(\{0 < v < \varepsilon\}) \quad \text{and} \quad \|\nabla v\|_2 > 0 \text{ in } \{0 < v < \varepsilon\}.$$

Let us assume now that two non-empty, open, bounded intervals I and J and a Lipschitz continuous C^1-map $\gamma : I \to J$ are given such that $W := I \times J$ satisfies

$$W \subset \{v < \varepsilon\} \quad \text{and} \quad W \cap \partial\{v = 0\} = W \cap \{v = 0\} = W \cap \mathcal{M} = \{(x, \gamma(x)) \mid x \in I\}. \tag{5.27}$$

Then, it holds

$$\begin{aligned} \{0 < v < t\} \cap W &\subset \Big\{ d_{\mathcal{M}} < t^{1/2} \Big\} \cap W \\ &\subset \Big\{ p \in I \times J \;\Big|\; \operatorname{dist}(p, \partial W \cup (\mathcal{M} \cap W)) < t^{1/2} \Big\} \\ &\subset \Big\{ p \in I \times J \;\Big|\; \operatorname{dist}(p, \partial W) < t^{1/2} \Big\} \\ &\qquad \cup \Big\{ p \in I \times J \;\Big|\; \operatorname{dist}(p, \partial W) > t^{1/2} \text{ and } \operatorname{dist}(p, \mathcal{M} \cap W) < t^{1/2} \Big\}. \end{aligned} \tag{5.28}$$

Note that for every $p = (x, y) \in I \times J$ with $\operatorname{dist}(p, \partial W) > t^{1/2}$ and $\operatorname{dist}(p, \mathcal{M} \cap W) < t^{1/2}$ we can find an $\tilde{x} \in I$ with $\|(x, y) - (\tilde{x}, \gamma(\tilde{x}))\|_2 < t^{1/2}$. For this $\tilde{x}$, we obtain

$$t^{1/2} > \frac{1}{\sqrt{2}} \Big(|x - \tilde{x}| + |y - \gamma(\tilde{x})| \Big) \geq \frac{1}{\sqrt{2}} \Big(|x - \tilde{x}| + |y - \gamma(x)| - |\gamma(x) - \gamma(\tilde{x})| \Big)$$

and

$$|y - \gamma(x)| \leq \sqrt{2}\, t^{1/2} + |\gamma(x) - \gamma(\tilde{x})| + |x - \tilde{x}| \leq C t^{1/2},$$

where $C = C(\gamma) > 0$ is some constant. Using the above in (5.28) yields

$$\begin{aligned} &\{0 < v < t\} \cap W \\ &\quad \subset \left\{ p \in I \times J \;\middle|\; \operatorname{dist}(p, \partial W) < t^{1/2} \right\} \cup \left\{ (x, y) \in I \times J \;\middle|\; |y - \gamma(x)| \leq C t^{1/2} \right\} \end{aligned}$$

and, as a consequence,

$$\int_{\{0<v<t\} \cap W} \|\nabla v\|_2^3 \mathrm{d}\mathcal{L}^2 \leq 8 t^{3/2} \mathcal{L}^2(\{0 < v < t\} \cap W) \leq \tilde{C} t^2.$$

This shows that Proposition 5.1.11 is applicable and that the trace of z is constant on $W \cap \mathcal{M}$. Since the submanifold property of $\mathcal{M}$ implies that for every $p \in \mathcal{M}$ we can find intervals I, J and a map $\gamma : I \to J$ with (5.27) and $p \in I \times J$ (at least after an orthogonal change of coordinates), the claim now follows immediately. □

Corollary 5.1.13 is remarkable because it complements the following classical result:

Theorem 5.1.14 (Reverse of Hardy's Inequality)**.** *Let $D \subset \mathbb{R}^2$ be a non-empty, open, bounded set, and let $\mathcal{M} \subset D$ be a compact, one-dimensional C^2-submanifold. Suppose that a $z \in W^{1,q}(D)$, $1 \leq q < \infty$, with*

$$\int_{D \setminus \mathcal{M}} \frac{z^q}{d_{\mathcal{M}}^q} \mathrm{d}\mathcal{L}^2 < \infty \tag{5.29}$$

is given, where $d_{\mathcal{M}}$ again denotes the Euclidean distance to $\mathcal{M}$. Then, the trace of z is zero on $\mathcal{M}$.

Proof. Since $\mathcal{M}$ is compact and since D is open, we may restrict our attention to functions $z \in W_0^{1,q}(D)$ (if this is not the case, we multiply with a cut-off function). The claim now follows immediately from [Edmunds and Evans, 1987, Theorem V.3.4, Remark 3.5] applied to $\mathbb{R}^2 \setminus \mathcal{M}$. □

Let us conclude this section with some comments on Corollary 5.1.13 and Theorem 5.1.14 and several questions that remain open regarding the integrability condition (5.22):

Remark 5.1.15.

(i) Theorem 5.1.14 is also valid in higher dimensions and for more general sets $\mathcal{M}$, cf. [Edmunds and Evans, 1987; Egert et al., 2015; Kinnunen and Martio, 1997; Maz'ya, 2011]. We work with a C^2-manifold and the assumption $d = 2$ here to facilitate the comparison with Corollary 5.1.13.

(ii) To the author's best knowledge, it is presently unknown if the assumptions of Corollary 5.1.13 can be relaxed and if a similar criterion for the constantness of the trace holds in the case $d > 2$. Note that the proof of Proposition 5.1.11 does not carry over to the higher-dimensional setting due to the use of Lemma 5.1.8 and the last step in (5.24)*.*

(iii) Instead of $d_{\mathcal{M}}$, we could also have used a function with comparable properties in Corollary 5.1.13. Proposition 5.1.11 therefore gives rise to a whole family of conditions that are sufficient for the constantness of the trace of a function z.

(iv) For sufficiently regular sets $\mathcal{M}$, the integrability condition (5.29) *is also necessary for the identity $\operatorname{tr}(z) = 0$ on $\mathcal{M}$, see, e.g., [Kinnunen and Martio, 1997]. It is unclear whether a similar result can be proved for the criterion* (5.26)*.*

5.1.4 An Interlude on Lipschitz Domains

The observation that the integrability condition (5.14) implies that the function z has a locally constant trace on the boundary $\partial\{\nabla w = 0\}$ (for sufficiently smooth w) suggests to proceed in the following steps to check the abstract differentiability criterion $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$ in Theorem 5.1.2:

(i) Consider an arbitrary but fixed $z \in \mathcal{K}$.

(ii) Construct a function $\tilde{z} \in H^1_0(\Omega)$ such that

$$\begin{aligned} \nabla\tilde{z} &= \nabla z \ \mathcal{L}^2\text{-a.e. in } \{\nabla w = 0\}, \\ \nabla\tilde{z} &= 0 \ \mathcal{L}^2\text{-a.e. in } \{\nabla w \neq 0\} \setminus E \text{ for some compact set } E \subset \{\nabla w \neq 0\}, \\ z &= \tilde{z} \ \mathcal{L}^2\text{-a.e. in } \{\nabla w = 0\}, \\ \mathrm{tr}(z) &= \mathrm{tr}(\tilde{z}) \ \mathcal{H}^1\text{-a.e. on } \partial\{\nabla w = 0\}. \end{aligned}$$

(Note that the constantness of $\mathrm{tr}(z)$ on $\partial\{\nabla w = 0\}$ is necessary for the existence of such a $\tilde{z}$ and that the above properties imply $\tilde{z} \in \mathcal{Z}$ if $w \in C^{1,1}(\Omega)$.)

(iii) Approximate the difference $z - \tilde{z} \in \mathcal{K}$ w.r.t. $\|\cdot\|_{\mathcal{K}}$ by functions which vanish in a neighborhood of the set $\{\nabla w = 0\}$.

What is surprisingly complicated in the above approach is to systematically construct the function $\tilde{z}$. We would like to point out that, if a particular solution $w \in C^{1,1}(\Omega)$ is considered, then it is typically completely obvious how $\tilde{z}$ has to be chosen for a given $z \in \mathcal{K}$. In the situation of Figure 5.2, for example, where the set $\{\nabla w = 0\}$ consists of only one connected component, $\mathcal{A}_0$ lets say, that has a smooth boundary and that is contained entirely in Ω, the function

$$\tilde{z} := \begin{cases} z|_{\mathcal{A}_0} & \text{a.e. in } \mathcal{A}_0 \\ c\psi & \text{a.e. in } \Omega \setminus \mathcal{A}_0 \end{cases} \tag{5.30}$$

has exactly the desired properties if $\psi \in C^\infty_c(\Omega)$ is a smooth bump function satisfying $\psi \equiv 1$ in a neighborhood of $\mathcal{A}_0$ and if $c \in \mathbb{R}$ is the constant with $\mathrm{tr}(z) = c$ on $\partial\mathcal{A}_0$. The main difficulty in a general proof is to handle the large variety of different topological situations that may arise when the set $\{\nabla w = 0\}$ is considered. (Note that, even in the rotationally symmetric case, the set $\{\nabla w = 0\}$ can consist of several concentric rings.) To tackle this problem, in what follows, we collect and prove several results on the properties of Lipschitz domains that are quite intuitive but hard to find in closed form in the literature. For the convenience of the reader, we begin by recalling the classical:

Definition 5.1.16 (Lipschitz Domains and Manifolds)**.**

(i) A set $\mathcal{M} \subset \mathbb{R}^d$ is called a (strong) $(d-1)$-dimensional Lipschitz submanifold of $\mathbb{R}^d$ if for every point $p \in \mathcal{M}$ there exist an orthogonal transformation $R \in \mathrm{O}(d)$, a closed ball $B_r \subset \mathbb{R}^{d-1}$ of radius $r > 0$ (not necessarily centered at the origin), an open interval J, and a Lipschitz continuous map $\gamma : B_r \to J$ such that p is contained in the interior of the set $R(B_r \times J)$ and such that

$$\mathcal{M} \cap R(B_r \times J) = R(\{(x, \gamma(x)) \mid x \in B_r\}).$$

(ii) A set $D \subset \mathbb{R}^d$ is called a (strong) Lipschitz domain if it is connected, non-empty and open, and if the boundary ∂D is a $(d-1)$-dimensional Lipschitz submanifold of $\mathbb{R}^d$ in the sense of (i).

We may now make the following observations:

Lemma 5.1.17. *If D is a non-empty, open, bounded subset of $\mathbb{R}^2$ such that D and $\mathbb{R}^2 \setminus D$ are path-connected, then D is simply connected.*

Proof. The claim follows straightforwardly from [Dudley and Norvaiša, 2010, Theorem 2.105a)c)]. □

Lemma 5.1.18. *Let $D \subset \mathbb{R}^2$ be a bounded Lipschitz domain. Then, it holds:*

(i) The boundary ∂D consists of finitely many path-connected components. All of these components are closed one-dimensional Lipschitz submanifolds of $\mathbb{R}^2$.

(ii) The set $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ consists of finitely many connected components. All of these components are Lipschitz domains.

(iii) The set $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ has one and only one unbounded connected component.

(iv) All bounded components of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ are simply connected.

Proof. Ad (i): Since D is a two-dimensional Lipschitz domain, we know that for every $p \in \partial D$ we can find an orthogonal transformation $R \in \mathrm{O}(2)$, non-empty, open, bounded intervals I, J and a Lipschitz function $\gamma : I \to J$ such that $p \in R(I \times J)$ and

$$R(I \times J) \cap D = R(\{(x, y) \in I \times J \mid \gamma(x) < y\}).$$

In what follows, we call a domain $W := R(I \times J)$ of the above type a rectification domain. Since the boundary ∂D is bounded and closed, it is compact and we can cover ∂D with finitely many rectification domains W_n, $n = 1, ..., N$. For each W_n, the set $W_n \cap \partial D$ is obviously path-connected and thus all $p \in W_n \cap \partial D$ are contained in the same path-connected component of the boundary ∂D. This yields that there can only be finitely many path-connected components of ∂D. Assume now that E is a path-connected component of the boundary ∂D and that $\{p_m\} \subset E$ is a sequence converging to some $p \in \mathbb{R}^2$. Then, p is contained in ∂D since the boundary is closed and we may find an $R \in \mathrm{O}(2)$, non-empty, open, bounded intervals I, J and a Lipschitz function $\gamma : I \to J$ such that $p \in R(I \times J)$ and $R(I \times J) \cap \partial D = R(\{(x, \gamma(x)) \mid x \in I\})$. Since $p_m \to p$, it holds $p_m \in R(I \times J)$ for all sufficiently large m. For such m, however, we can clearly find a path from p_m to p. Thus, $p \in E$ and E is closed. The submanifold property is trivial.

Ad (ii): Let W_n, $n = 1, ..., N$, be an open cover of ∂D as in (i) and denote with $\{Z_m\}$ the collection of all (open) connected components of the set $\mathbb{R}^2 \setminus \mathrm{cl}(D)$. Then, for every Z_m, it holds $\partial Z_m \subset \partial D$. This can be seen as follows:

$$\begin{aligned}\partial Z_m = \mathrm{cl}(Z_m) \setminus Z_m &= \mathrm{cl}(Z_m) \setminus (\mathbb{R}^2 \setminus \mathrm{cl}(D)) \\ &\subset \mathrm{cl}(\mathbb{R}^2 \setminus \mathrm{cl}(D)) \setminus (\mathbb{R}^2 \setminus \mathrm{cl}(D)) = \partial(\mathbb{R}^2 \setminus \mathrm{cl}(D)) = \partial(\mathrm{cl}(D)) = \partial D.\end{aligned}$$

Further, for every Z_m it holds $\partial Z_m \neq \emptyset$ (if this was not the case, $\mathrm{cl}(Z_m) = Z_m \cup \partial Z_m = Z_m$ would yield $Z_m \in \{\emptyset, \mathbb{R}^2\}$). Define

$$\tilde{Z}_n := \bigcup_{Z_m : \partial Z_m \cap W_n \neq \emptyset} Z_m, \qquad n = 1, ..., N.$$

Then each $\tilde{Z}_n$ is a connected and open subset of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$. This shows that $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ can only have finitely many connected components. The Lipschitz regularity of ∂Z_m for all components Z_m of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ follows from $\partial Z_m \subset \partial D$ and the definition of the term Lipschitz domain.

Ad (iii): Denote with $r > 0$ a number such that $D \subset B_r(0)$, where $B_r(0)$ is the closed ball of radius r around the origin. Then, all $p, \tilde{p} \in \mathbb{R}^2$ with $\|p\|_2, \|\tilde{p}\|_2 > r$ can be connected with a path in $\mathbb{R}^2 \setminus B_r(0) \subset \mathbb{R}^2 \setminus \mathrm{cl}(D)$. This shows that the set $\mathbb{R}^2 \setminus B_r(0)$ is contained in a component of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ and proves the claim.

Ad (iv): Denote the finitely many bounded connected components of the set $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ with Z_m, $m = 1, ..., M$, and the unique unbounded component of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ with Z_0. Let p be an element of the set $\mathbb{R}^2 \setminus Z_M = \mathrm{cl}(D) \cup Z_0 \cup ... \cup Z_{M-1}$. Then, the following holds true:

- If $p \in Z_m$ for some $m = 0, ..., M-1$, then there exists a continuous path connecting p with a point in ∂D (this is seen by considering a rectification domain W_n that intersects $\partial Z_m \subset \partial D$ and by using the path-connectedness of Z_m).
- If $p \in \partial D$, then there exists a continuous path connecting p with a point in D.
- If $p \in D$, then there exists a continuous path connecting p with a point in ∂Z_0.
- If $p \in \partial Z_0$, then there exists a path connecting p with the set $\mathbb{R}^2 \setminus B_r(0)$, where $B_r(0)$, $r > 0$, is a closed ball as in the proof of (iii).

The above shows that $\mathbb{R}^2 \setminus Z_M$ is path-connected, and that Z_M is simply connected by Lemma 5.1.17. Since the above argumentation works for all Z_m, $m = 1, ..., M$, the claim now follows immediately. □

Lemma 5.1.19. *If $D \subset \mathbb{R}^2$ is a bounded, simply connected Lipschitz domain, then ∂D is a closed path-connected one-dimensional Lipschitz submanifold of $\mathbb{R}^2$.*

Proof. Recall that the boundary of a two-dimensional, simply connected, bounded domain is always connected, see [Beardon, 1991, Proposition 5.1.4], but not necessarily path-connected. We know, however, that D is a Lipschitz domain and thus, by Lemma 5.1.18(i), that ∂D consists of finitely many closed path-connected components which are even one-dimensional Lipschitz submanifolds. Thus, ∂D is locally path-connected and connected. This yields that ∂D is path-connected. □

Lemma 5.1.20. *Let $D \subset \mathbb{R}^2$ be a bounded Lipschitz domain, and let Z_m, $m = 0, ..., M$, denote the finitely many disjoint Lipschitz domains that make up the set $\mathbb{R}^2 \setminus \mathrm{cl}(D)$. Then, the boundary ∂D is the disjoint union of the boundaries ∂Z_m, and every ∂Z_m is a path-connected, closed, one-dimensional Lipschitz submanifold of $\mathbb{R}^2$.*

Proof. The inclusion $\partial D \subset \bigcup \partial Z_m$ is trivial, and the inclusion $\partial D \supset \bigcup \partial Z_m$ has already been shown in the proof of Lemma 5.1.18(ii). The disjointness of the boundaries ∂Z_m is trivial (just argue by contradiction, use $\partial Z_m \subset \partial D$ and rectify the boundary). Further, we know that the Z_m are Lipschitz domains by Lemma 5.1.18(ii) and that all bounded components Z_m are simply connected by Lemma 5.1.18(iv). This yields that the boundaries ∂Z_m of all bounded components of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ are closed connected one-dimensional Lipschitz submanifolds of $\mathbb{R}^2$ by Lemma 5.1.19. It remains to prove that the boundary of the unique unbounded component of $\mathbb{R}^2 \setminus \mathrm{cl}(D)$ is a path-connected, closed, one-dimensional Lipschitz submanifold. To see the latter, it suffices to show that the complement of the closure of the unbounded component is connected, cf. Lemma 5.1.17 and Lemma 5.1.19. This, however, follows from the same arguments as in the proof of Lemma 5.1.18. □

With view on the analysis in the next section, we further recall the following classical result which yields that traces of two-dimensional H^1-functions cannot have jump-discontinuities:

Lemma 5.1.21. *Let I, J be non-empty, open, bounded intervals and let $\gamma : I \to J$ be a Lipschitz continuous function with $\mathrm{cl}(\gamma(I)) \subset J$. Let $D := \{(x, y) \in I \times J \mid \gamma(x) < y\}$ and let $H^{1/2}(\partial D)$ be the Hilbert space of all $u \in L^2(\partial D, \mathcal{H}^1)$ which satisfy*

$$\|u\|^2_{H^{1/2}(\partial D)} := \int_{\partial D} |u(p)|^2 \mathrm{d}\mathcal{H}^1(p) + \int_{\partial D} \int_{\partial D} \frac{|u(p) - u(\tilde{p})|^2}{\|p - \tilde{p}\|_2^2} \mathrm{d}\mathcal{H}^1(p) \mathrm{d}\mathcal{H}^1(\tilde{p}) < \infty.$$

Then, D is a bounded Lipschitz domain, the operator $H^1(D) \to H^{1/2}(\partial D)$, $v \mapsto \mathrm{tr}(v)$, is well-defined and continuous, and for every $u \in L^2(\partial D, \mathcal{H}^1)$ with

$$u(x, \gamma(x)) = \begin{cases} c_1 \text{ for } \mathcal{H}^1\text{-a.a. } x \in I \cap (-\infty, \alpha) \\ c_2 \text{ for } \mathcal{H}^1\text{-a.a. } x \in I \cap (\alpha, \infty) \end{cases} \tag{5.31}$$

for some $\alpha \in I$ and some constants $c_1 \neq c_2$, it holds $u \notin H^{1/2}(\partial D)$.

Proof. It is obvious that D is a strong bounded Lipschitz domain and the well-definedness and continuity of the trace operator on the boundary ∂D are classical, see, e.g., [Ding, 1996, Theorem 1]. It remains to prove that a function with (5.31) cannot be an element of $H^{1/2}(\partial D)$. To see this, we write $I = (a, b)$ and note that for every u with (5.31), we necessarily have (due to the area formula, see [Evans and Gariepy, 2015, Theorem 3.9, page 122])

$$\begin{aligned}\|u\|^2_{H^{1/2}(\partial D)} &\geq |c_1 - c_2|^2 \int_a^\alpha \int_\alpha^b \frac{\sqrt{1+\gamma'(s)^2}\sqrt{1+\gamma'(t)^2}}{(s-t)^2 + (\gamma(s)-\gamma(t))^2} \mathrm{d}s\mathrm{d}t \\ &\geq \frac{|c_1 - c_2|^2}{1 + \|\gamma\|^2_{C^{0,1}}} \int_a^\alpha \int_\alpha^b \frac{1}{(s-t)^2} \mathrm{d}s\mathrm{d}t = \infty.\end{aligned}$$

This completes the proof. □

Remark 5.1.22. *Lemma 5.1.21 implies, e.g., that the argumentation used in Section 3.4.2 cannot be employed to prove the non-polyhedricity of the set L in* (3.19) *in the borderline case $q = d = 2$ (because it yields that, for these q and d, a function v with the required discontinuity properties cannot exist).*

5.1.5 A Tangible Criterion for Directional Differentiability

We are now finally in the position to derive a condition that is sufficient for the density $\mathrm{cl}_{\|\cdot\|_\mathcal{K}}(\mathcal{Z}) = \mathcal{K}$ in Theorem 5.1.2 and that, as a consequence, ensures the directional differentiability of the solution map S to the EVI (M) and the second-order epi-differentiability of the functional j in (5.1). Henceforth, we again consider the situation in Assumption 5.1.1. Further, we use the following notation:

Definition 5.1.23. *Given a function $v \in C^1(\Omega)$, we define:*

(i) $\mathcal{I} := \{\|\nabla v\|_2 > 0\}$, $\{\mathcal{I}_i\}$ *set of all connected components of* $\mathcal{I}$,

(ii) $\mathcal{A} := \{\|\nabla v\|_2 = 0\}$, $\{\mathcal{A}_i\}$ *set of all connected components of* $\mathcal{A}$,

(iii) $\mathcal{A}^\circ := \mathrm{int}(\mathcal{A})$, $\{\mathcal{A}_i^\circ\}$ *set of all connected components of* $\mathcal{A}^\circ$,

(iv) $\mathcal{B} := \partial\mathcal{A} \cup \partial\Omega$, $\{\mathcal{B}_i\}$ *set of all connected components of* $\mathcal{B}$,

(v) $\mathcal{B}^\circ \subset \mathcal{B}$ *is the set*

$$\begin{aligned}\mathcal{B}^\circ := \Big\{ p \in \partial\mathcal{A} \;\Big|\; &\text{there exists an open neighborhood } D \subset \Omega \text{ of } p \text{ such that} \\ &D \cap \partial\mathcal{A} \text{ is a one-dimensional } C^1\text{-submanifold of } \mathbb{R}^2 \\ &\text{and such that } D \cap \partial\mathcal{A} = D \cap \partial\{v = c\} \text{ for some } c \in \mathbb{R} \Big\}.\end{aligned}$$

Note that $\mathcal{I}, \mathcal{A}, \mathcal{A}^\circ$ and $\mathcal{B}^\circ$ are subsets of Ω according to our convention

$$\{v \text{ satisfies a pointwise condition } P\} := \{x \in \Omega \mid v \text{ satisfies } P \text{ at } x\}$$

for functions $v : \Omega \to \mathbb{R}$, and that $\mathcal{B}$ always contains the boundary $\partial\Omega$, cf. Figure 5.4. In addition to Definition 5.1.23, we need:

Definition 5.1.24 (Lipschitz Connectedness)**.** *A set $Z \subset \mathrm{cl}(\Omega)$ is said to be Lipschitz connected if for all $p, \tilde{p} \in Z$ (not necessarily $p \neq \tilde{p}$) there exists a connected one-dimensional Lipschitz submanifold $\mathcal{M} \subset Z$ with $p, \tilde{p} \in \mathcal{M}$.*

Our sufficient condition for the directional differentiability of the solution operator S to the EVI (M) may now be formulated as follows:

Assumption 5.1.25. *The tuple (w, φ), $w := S(f)$, $\varphi := f + \Delta w$, is such that:*

- (Regularity) *It holds $w \in C^{1,1}(\Omega) \cap H^1_0(\Omega)$ and $\varphi \in L^\infty(\Omega)$.*
- (Structure of the Active and the Inactive Set) *The sets $\{\nabla w = 0\}$ and $\{\nabla w \neq 0\}$ satisfy:*
  - *(i) the collections $\{\mathcal{I}_i\}$, $\{\mathcal{A}_i\}$, $\{\mathcal{A}_i^\circ\}$, $\{\mathcal{B}_i\}$ are finite,*
  - *(ii) the components $\mathcal{A}_i^\circ$ and $\mathcal{I}_i$ are Lipschitz domains for all i,*
  - *(iii) the components $\mathcal{A}_i$ and $\mathcal{B}_i$ are Lipschitz connected for all i,*
  - *(iv) the set $\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$ is finite and $\mathcal{B} = \mathrm{cl}(\mathcal{B}^\circ) \cup \partial\Omega$.*
- (Well-Behavedness of the Normalized Gradient Field) *There exist a function $\omega \in C^{0,1}(\Omega)$, a constant $C > 0$, and an open set $D \subset \mathbb{R}^2$ with $\mathcal{A} \cup \partial\Omega \subset D$ and*

$$\omega = 0 \text{ on } \mathcal{A} \cup \partial\Omega, \quad \mathrm{dist}(\cdot, \mathcal{A} \cup \partial\Omega) \leq C\omega \text{ a.e. in } \mathcal{I} \cap D,$$
$$\left(\frac{\nabla w^\perp}{\|\nabla w\|_2} \cdot \frac{\nabla \omega}{\|\nabla \omega\|_2} \right)^2 \leq C \|\nabla w\|_2 \text{ a.e. in } \mathcal{I} \cap D.$$

Before we prove that the above conditions indeed imply the directional differentiability of S in f, let us briefly comment on Assumption 5.1.25:

Remark 5.1.26.

(i) As we have seen in Section 5.1.2, the assumption $w \in C^{1,1}(\Omega)$ is restrictive but not unrealistic.

(ii) For a constant right-hand side f and a sufficiently regular domain Ω, it is possible to prove that particular parts of the boundary $\partial\{\nabla w \neq 0\}$ are convex and, as a consequence, Lipschitz, see [Mosolov and Miasnikov, 1965, Section 2, Theorem 2.7]. This shows that the conditions on the sets $\{\nabla w = 0\}$ and $\{\nabla w \neq 0\}$ in Assumption 5.1.25 are not unreasonable. Compare also with the rotationally symmetric case in this context.

(iii) The third condition in Assumption 5.1.25 expresses that we can find a Lipschitz function ω that vanishes on $\{\nabla w = 0\} \cup \partial\Omega$, that does not decay too fast near this set and whose normalized gradient field approximates $\nabla w / \|\nabla w\|_2$ sufficiently well in the vicinity of the rigid zones. The author suspects that this condition is always satisfied when w and φ satisfy the first two points of Assumption 5.1.25. (The TV-seminorm (5.1) *is, after all, known for its regularizing effect on level sets, see, e.g., [Chambolle and Darbon, 2009, Section 2] and also Lemma 5.1.28 below.) We point out that the existence of the function ω is not needed to prove that functions z with* (5.14) *have a locally constant trace on the boundary $\partial\{\nabla w = 0\}$. This also indicates that the level sets of w are very regular near $\{\nabla w = 0\}$ (see Remark 5.1.12) and that the third condition in Assumption 5.1.25 is superfluous.*

(iv) We remark that conditions similar to those in Assumption 5.1.25 appear, e.g., in [De los Reyes and Meyer, 2016]. Further, it should be noted that our Assumption 5.1.25 is far more tangible than the abstract conditions that are otherwise found in the literature in the context of EVIs of the second kind, cf. [Hintermüller and Surowiec, 2017; Sokołowski, 1988].

(v) Assumption 5.1.25 is, for example, trivially satisfied in the situation of Figure 5.2.

To show that Assumption 5.1.25 ensures the equality (5.3) and the applicability of Theorem 5.1.2, we use the strategy that we have outlined at the beginning of Section 5.1.4. We begin by analyzing the properties of the function w and by proving that we may indeed employ Proposition 5.1.11 to study the traces of functions z with (5.14).

Lemma 5.1.27. *In the situation of Assumption 5.1.25, the set*

$$\mathcal{W} := \{c \in \mathbb{R} \mid \{w = c\} \cap \{\nabla w = 0\} \neq \emptyset\} \cup \{0\}$$

is finite.

Proof. From the first two points in Assumption 5.1.25, we obtain that the set $\mathcal{A} = \{\nabla w = 0\}$ consists of finitely many connected components $\mathcal{A}_i$, which are all Lipschitz connected, and that $\nabla w \in C(\Omega)$. Consider now some $p_1, p_2 \in \mathcal{A}_i$, i fixed. Then, we know that there is a connected one-dimensional Lipschitz submanifold $\mathcal{M} \subset \mathcal{A}_i \subset \Omega$ with $p_1, p_2 \in \mathcal{M}$. For each $p \in \mathcal{M}$, we obtain from $\nabla w = 0$ on $\mathcal{M} \subset \mathcal{A}_i$ by parametrization that there is an open set $D \subset \mathbb{R}^2$ with $p \in D$ and $w = \text{const}$ on $D \cap \mathcal{M}$. Consequently, w is locally constant on $\mathcal{M}$, and we obtain from the connectedness of $\mathcal{M}$ that $w(p_1) = w(p_2)$. Since p_1, p_2 were arbitrary, it follows that $w \equiv c_i$ holds on $\mathcal{A}_i$ for some constant c_i. The claim now follows immediately from the finiteness of the collection $\{\mathcal{A}_i\}$. □

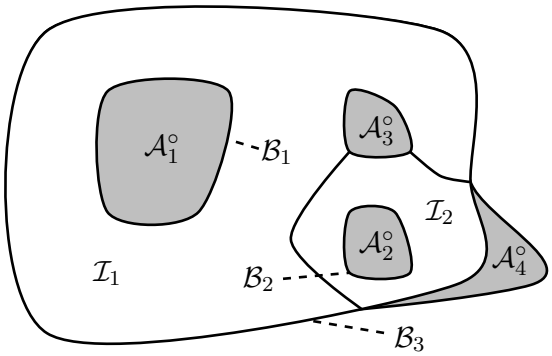

Figure 5.4: The sets $\mathcal{A}_i$, $\mathcal{A}_i^\circ$, $\mathcal{B}_i$ and $\mathcal{I}_i$ in a prototypical situation. The norm $\|\nabla w\|_2$ vanishes in the gray sets and on the black lines that are contained in Ω.

From Lemma 5.1.27, we may deduce:

Lemma 5.1.28. *In the situation of Assumption 5.1.25, the function*

$$\ell : \mathbb{R} \to [0, \infty], \quad c \mapsto \mathcal{H}^1(\{w = c\}),$$

is Lebesgue measurable and for $\mathcal{L}^1$-almost all $c \in \mathbb{R}$, it holds

$$|\ell(c)| \leq \|\varphi\|_{L^1}. \tag{5.32}$$

Proof. The Lebesgue measurability of ℓ follows from [Evans and Gariepy, 2015, Lemma 3.5]. To prove the estimate (5.32), we first note that the set $\mathcal{W}$ in Lemma 5.1.27 has one-dimensional Lebesgue measure zero. Consider now an arbitrary but fixed $c \in (0, \infty) \setminus \mathcal{W}$ with $\{w = c\} \neq \emptyset$ (else (5.32) is trivial). Then, there exists an $\varepsilon > 0$ such that $\{c - \varepsilon \leq w \leq c + \varepsilon\} \subset \{\nabla w \neq 0\}$ and it follows from the zero boundary conditions of w that the set $\{w = c\}$ is a compact one-dimensional C^1-submanifold contained in Ω, and that the set $\{w > c\}$ consists of finitely many C^1-domains whose closures are disjoint (cf. Lemma 5.1.18 and its proof). Further, we may deduce from (5.2) and the regularity assumptions on w and φ that there exists a $\lambda \in L^\infty(\Omega, \mathbb{R}^2)$ with

$$\varphi = -\nabla \cdot \lambda \in L^\infty(\Omega) \quad \text{and} \quad \lambda = \frac{\nabla w}{\|\nabla w\|_2} \text{ a.e. in } \{\nabla w \neq 0\}.$$

The above yields that λ is Lipschitz on every open set D with $\mathrm{cl}(D) \subset \{\nabla w \neq 0\}$ and that λ is an element of $H(\mathrm{div}; \Omega)$ with divergence in $L^\infty(\Omega)$. Using Lemma 5.1.9, we may now calculate

$$\begin{aligned}\mathcal{H}^1(\{w = c\}) &= \int_{\{w=c\}} \frac{\nabla w}{\|\nabla w\|_2} \cdot \frac{\nabla w}{\|\nabla w\|_2} \mathrm{d}\mathcal{H}^1 \\ &= -\int_{\partial\{w>c\}} \lambda \cdot \nu \, \mathrm{d}\mathcal{H}^1 \\ &= -\int_{\{w>c\}} \nabla \cdot \lambda \, \mathrm{d}\mathcal{L}^2 \\ &= \int_{\{w>c\}} \varphi \, \mathrm{d}\mathcal{L}^2.\end{aligned}$$

This yields (5.32) for all $c \in (0, \infty) \setminus \mathcal{W}$. The case $c \in (-\infty, 0) \setminus \mathcal{W}$ can be treated analogously. This proves the claim. □

Note that Lemma 5.1.28 expresses that the level sets $\{w = c\}$ of the solution w have finite length for $\mathcal{L}^1$-a.a. $c \in \mathbb{R}$. This has also been observed (under different assumptions) in [Mosolov and Miasnikov, 1967, Theorem 1]. Using the last two results, we can prove that the solution w indeed satisfies the growth condition (5.21) that is required in Proposition 5.1.11:

Lemma 5.1.29. *Suppose that Assumption 5.1.25 holds. Let $\mathcal{W}$ be defined as in Lemma 5.1.27, let $c \in \mathcal{W}$ be arbitrary but fixed and let $\varepsilon > 0$ be given such that $(c - \varepsilon, c) \cup (c, c + \varepsilon) \subset \mathbb{R} \setminus \mathcal{W}$. Assume that two non-empty, open, bounded intervals I and J and a Lipschitz continuous C^1-map $\gamma : I \to J$ are given such that $W := I \times J$ satisfies*

$$W \subset \{|w - c| < \varepsilon\} \quad \textit{and} \quad W \cap \partial\{w = c\} = \{(x, \gamma(x)) \mid x \in I\} \subset \{\|\nabla w\|_2 = 0\}.$$

Then, for every $\psi \in C_c^1(W)$ it holds

$$\int_{\{|w-c|<t\}} \psi^2 \|\nabla w\|_2^3 \mathrm{d}\mathcal{L}^2 = O(t^2).$$

Proof. From our assumptions, we obtain that the functions

$$\Omega \ni x \mapsto \begin{cases} \psi(x)^2 \|\nabla w(x)\|_2^2 & \text{in } W \cap \{w > c\} \\ 0 & \text{else} \end{cases}$$

and

$$\Omega \ni x \mapsto \begin{cases} \psi(x)^2 \|\nabla w(x)\|_2^2 & \text{in } W \cap \{w < c\} \\ 0 & \text{else} \end{cases}$$

are both elements of $H_0^1(\Omega) \cap W^{1,\infty}(\Omega)$ (just rectify the Lipschitz graph $W \cap \partial\{w = c\}$ and argue with an extension by zero in the half-plane). Using the latter and the fact that the functions $\max(0, -w+c+t)$ and $-\max(0, w - c + t)$ are elements of $W^{1,\infty}(\Omega)$, we obtain that the function

$$\psi_t := \begin{cases} \max(0, -w + c + t)\psi^2 \|\nabla w\|_2^2 & \text{in } W \cap \{c < w\} \\ -\max(0, w - c + t)\psi^2 \|\nabla w\|_2^2 & \text{in } W \cap \{c > w\} \\ 0 & \text{else} \end{cases}$$

is an element of $W^{1,\infty}(\Omega) \cap H_0^1(\Omega)$ for all $t > 0$ with

$$\begin{aligned}\nabla \psi_t = {} & 2\psi \, \mathrm{sgn}(w - c)(t - |w - c|)\|\nabla w\|_2^2 \nabla \psi \\ & + \mathrm{sgn}(w - c)\psi^2 (t - |w - c|) 2\nabla^T w \nabla^2 w \\ & - \psi^2 \|\nabla w\|_2^2 \nabla w\end{aligned}$$

a.e. in $\{|w-c|<t\}$ and $\nabla\psi_t = 0$ a.e. else. If we use ψ_t to test the identity $\varphi = -\nabla\cdot\lambda \in L^\infty(\Omega)$ in (5.2), then we obtain

$$\begin{aligned}
&\int_{\{|w-c|<t\}} \varphi\Big(\operatorname{sgn}(w-c)\psi^2(t-|w-c|)\|\nabla w\|_2^2\Big)\mathrm{d}\mathcal{L}^2\\
&= \int_{\{|w-c|<t\}} \lambda\cdot\Big(2\psi\operatorname{sgn}(w-c)(t-|w-c|)\|\nabla w\|_2^2\nabla\psi\Big)\mathrm{d}\mathcal{L}^2\\
&\quad + \int_{\{|w-c|<t\}} \lambda\cdot\Big(\operatorname{sgn}(w-c)\psi^2(t-|w-c|)2\nabla^T w\nabla^2 w\Big)\mathrm{d}\mathcal{L}^2\\
&\quad - \int_{\{|w-c|<t\}} \lambda\cdot\Big(\psi^2\|\nabla w\|_2^2\nabla w\Big)\mathrm{d}\mathcal{L}^2.
\end{aligned}$$

Using the properties of λ, Lemma 5.1.28, and Lemma 5.1.10 in the above yields

$$\begin{aligned}
&\int_{\{|w-c|<t\}} \psi^2\|\nabla w\|_2^3\mathrm{d}\mathcal{L}^2\\
&\le 2\|\psi\|_{W^{1,\infty}}^2\int_{W\cap\{|w-c|<t\}}(t-|w-c|)\|\nabla w\|_2^2\mathrm{d}\mathcal{L}^2\\
&\quad + 2\|\nabla^2 w\|_{L^\infty}\|\psi\|_{L^\infty}^2\int_{W\cap\{|w-c|<t\}}(t-|w-c|)\|\nabla w\|_2\mathrm{d}\mathcal{L}^2\\
&\quad + \|\varphi\|_{L^\infty}\|\psi\|_{L^\infty}^2\int_{W\cap\{|w-c|<t\}}(t-|w-c|)\|\nabla w\|_2^2\mathrm{d}\mathcal{L}^2\\
&\le C(\|\nabla w\|_{L^\infty},\|\nabla^2 w\|_{L^\infty},\|\psi\|_{W^{1,\infty}},\|\varphi\|_{L^\infty})\int_{\{|w-c|<t\}}(t-|w-c|)\|\nabla w\|_2\mathrm{d}\mathcal{L}^2\\
&= C(...)\int_{-\infty}^{\infty}\int_{\{w=s\}}(t-|w-c|)\mathbb{1}_{\{|w-c|<t\}}\mathrm{d}\mathcal{H}^1\mathrm{d}s\\
&= C(...)\int_{c-t}^{c+t}(t-|s-c|)\mathcal{H}^1(\{w=s\})\mathrm{d}s\\
&\le C(...)\|\varphi\|_{L^1}\int_{-t}^{t}(t-|s|)\mathrm{d}s\\
&\le C(...)\|\varphi\|_{L^1}t^2
\end{aligned}$$

for all $0<t<\varepsilon$, where C is some constant that may change from step to step and that depends on the mentioned quantities. This yields the claim. □

Having proved that Proposition 5.1.11 can be applied in the situation of Assumption 5.1.25, we now turn our attention to the consequences that this result has for the traces of functions z which satisfy the integrability condition (5.14). For the sake of clarity, we introduce:

Definition 5.1.30 (Constant in the Sense of Traces)**.** *Suppose that a function $z\in H_0^1(\Omega)$ is given and that $Z\subset\mathrm{cl}(\Omega)$ is some set. Then, we say that z is constant on Z in the sense of traces if there exists a constant c with the following property: If two non-empty, open, bounded intervals I, J and a Lipschitz continuous map $\gamma: I\to J$ are given such that (after possibly an orthogonal change of coordinates)*

$$\Gamma := \{(x,\gamma(x)) \mid x\in I\}\subset Z,$$

then it holds $\mathrm{tr}(z)\equiv c$ on Γ, where $\mathrm{tr}: H_0^1(\Omega)\to L^2(\Gamma,\mathcal{H}^1)$ denotes the trace operator associated with the one-dimensional Lipschitz submanifold Γ.

As a straightforward consequence of Proposition 5.1.11, Lemma 5.1.29 and our assumptions on w, we now obtain:

Lemma 5.1.31. *Suppose that Assumption 5.1.25 holds. Let $\mathcal{B}_i^\circ$ be a connected component of the set $\mathcal{B}^\circ$, and let $z \in H_0^1(\Omega)$ be a function satisfying* (5.14)*, i.e.,*

$$\int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla z)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 < \infty. \tag{5.33}$$

Then, z is constant on $\mathcal{B}_i^\circ$ in the sense of traces .

Proof. Let $p \in \mathcal{B}_i^\circ$ be arbitrary but fixed. Then, it follows from the definition of the set $\mathcal{B}^\circ$ that (after possibly an orthogonal change of coordinates) we can find non-empty, open, bounded intervals I, J and a Lipschitz continuous C^1-map $\gamma : I \to J$ such that $p \in W := I \times J \subset \Omega$ and

$$W \cap \partial\mathcal{A} = W \cap \partial\{w = c\} = \{(x, \gamma(x)) \mid x \in I\} \subset \{\|\nabla w\|_2 = 0\}$$

for some $c \in \mathbb{R}$. Note that, since the C^1-graph $\{(x, \gamma(x)) \mid x \in I\}$ is connected and since $W \cap \partial\mathcal{A} = W \cap \mathcal{B}^\circ$ (see the definition of $\mathcal{B}^\circ$), we obtain that we even have

$$W \cap \partial\mathcal{A} = W \cap \partial\{w = c\} = W \cap \mathcal{B}_i^\circ = \{(x, \gamma(x)) \mid x \in I\} \subset \{\|\nabla w\|_2 = 0\}.$$

The above implies in particular that $\mathcal{B}_i^\circ$ is a one-dimensional C^1-submanifold and that $c \in \mathcal{W}$, where $\mathcal{W}$ is defined as before. By making the sets I and J smaller, we can always achieve that, in addition to the above, $W \subset \{|w - c| < \varepsilon\}$ holds for some $\varepsilon > 0$ with $(c-\varepsilon, c) \cup (c, c+\varepsilon) \subset \mathbb{R} \setminus \mathcal{W}$, see Lemma 5.1.27. From Lemma 5.1.29, we may now deduce that

$$\int_{\{|w-c|<t\}} \psi^2 \|\nabla w\|_2^3 \mathrm{d}\mathcal{L}^2 = O(t^2)$$

holds for all $\psi \in C_c^1(W)$. This, in turn, allows us to employ Proposition 5.1.11 and to obtain that z has to be constant in the sense of traces on $W \cap \partial\{w = c\} = W \cap \mathcal{B}_i^\circ$. In summary, we have now proven that for every $p \in \mathcal{B}_i^\circ$ there exists an open set $W \subset \Omega$ such that z is constant in the sense of traces on $W \cap \mathcal{B}_i^\circ$. The connectedness of $\mathcal{B}_i^\circ$ yields that the appearing constants have to be the same. This proves the claim. □

The last result gives rise to:

Lemma 5.1.32. *Suppose that Assumption 5.1.25 is satisfied. Let $\mathcal{M}$ be a one-dimensional Lipschitz submanifold with $\mathcal{M} \subset \mathcal{B}$, and let $z \in H_0^1(\Omega)$ be a function with* (5.33)*. Then, the following holds true:*

(i) For every point $p \in \mathcal{M}$ with $p \in \mathcal{B}^\circ$ there exists an open set $D \subset \mathbb{R}^2$ with $p \in D$ such that $D \cap \mathcal{M} = D \cap \mathcal{B}^\circ = D \cap \mathcal{B}$ and such that z is constant on $D \cap \mathcal{M}$ in the sense of traces.

(ii) For every point $p \in \mathcal{M}$ with $p \in \partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)$ there exists an open set $D \subset \mathbb{R}^2$ with $p \in D$ such that $D \cap \mathcal{M} = D \cap \partial\Omega = D \cap \mathcal{B}$ and such that z is constant on $D \cap \mathcal{M}$ in the sense of traces.

(iii) For every point $p \in \mathcal{M}$ with $p \in \mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$ there exists an open set $D \subset \mathbb{R}^2$ with $p \in D$ such that z is constant on $D \cap \mathcal{M}$ in the sense of traces.

Proof. Ad (i): If $p \in \mathcal{M} \cap \mathcal{B}^\circ$, then we know from the definition of $\mathcal{B}^\circ$ that there exist (after possibly changing coordinates) non-empty, open, bounded intervals I and J and a Lipschitz continuous C^1-map $\gamma : I \to J$ such that $p \in W := I \times J \subset \Omega$ and

$$W \cap \mathcal{B} = W \cap \mathcal{B}^\circ = \{(x, \gamma(x)) \mid x \in I\}.$$

Since $\mathcal{M}$ is a Lipschitz submanifold and thus locally a Lipschitz graph, we can further find an injective continuous curve $\theta : [-1, 1] \to W$ such that

$$\theta(t) \in \mathcal{M} \subset \mathcal{B} \ \ \forall t \in [-1, 1], \quad \theta(-1) \neq \theta(1), \quad \theta(0) = p.$$

The properties of θ and W yield that $\theta(-1) = (x_1, \gamma(x_1))$ and $\theta(1) = (x_2, \gamma(x_2))$ has to hold for some $x_1 \neq x_2 \in I$. If we assume w.l.o.g. that $x_1 < x_2$ (else reparametrize), then we obtain further that $x_1 < x_p < x_2$ has to hold, where x_p denotes the x-coordinate of p (if this was not the case, there would be a contradiction with the properties of θ). Since $\theta([-1,1])$ is path-connected, we may deduce that

$$\theta([-1,1]) = \{(x, \gamma(x)) \mid x \in [x_1, x_2]\}.$$

Defining $\tilde{I} := (x_1, x_2)$ and $\tilde{W} := \tilde{I} \times J$, we now obtain an open set $\tilde{W} \subset W$ with $p \in \tilde{W}$ and

$$\tilde{W} \cap \mathcal{B} = \tilde{W} \cap \mathcal{B}^\circ = \tilde{W} \cap \mathcal{M} = \{(x, \gamma(x)) \mid x \in \tilde{I}\}.$$

Note that z is trivially constant in the sense of traces on $\mathcal{M} \cap \tilde{W}$ since $\mathcal{M} \cap \tilde{W}$ is a connected subset of $\mathcal{B}^\circ$ and since Lemma 5.1.31 yields that z is constant in the sense of traces on every connected component of the set $\mathcal{B}^\circ$. This proves (i).

Ad (ii): The proof is along the lines of that of (i). If a $p \in \mathcal{M} \cap \partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)$ is given, then we obtain from the fact that Ω is a Lipschitz domain that there exist (after possibly changing coordinates) non-empty, open, bounded intervals I, J and a Lipschitz continuous map $\gamma : I \to J$ such that $p \in W := I \times J$ and

$$W \cap \partial\Omega = \{(x, \gamma(x)) \mid x \in I\}.$$

Since $\mathrm{cl}(\mathcal{B}^\circ)$ is closed, by making the sets I, J smaller, we can always obtain that $W \cap \mathrm{cl}(\mathcal{B}^\circ) = \emptyset$. If the latter is the case, then $\mathcal{B} = \mathrm{cl}(\mathcal{B}^\circ) \cup \partial\Omega$ yields

$$W \cap \mathcal{B} = W \cap \partial\Omega = \{(x, \gamma(x)) \mid x \in I\}.$$

Using exactly the same argumentation as in (i), we can replace I with a smaller interval $\tilde{I}$ so that

$$\tilde{W} \cap \mathcal{B} = \tilde{W} \cap \partial\Omega = \tilde{W} \cap \mathcal{M} = \{(x, \gamma(x)) \mid x \in \tilde{I}\}$$

holds with $\tilde{W} := \tilde{I} \times J$. Note that z is trivially constant in the sense of traces on $\mathcal{M} \cap \tilde{W}$ due to the zero boundary conditions. This proves (ii)

Ad (iii): Since the set $\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$ is finite and since $\mathcal{M}$ is a Lipschitz manifold, we can again find (at least after changing coordinates) non-empty, open, bounded intervals I, J and a Lipschitz continuous map $\gamma : I \to J$ such that $p \in W := I \times J$ and

$$W \cap \mathcal{M} = \{(x, \gamma(x)) \mid x \in I\}, \qquad W \cap \mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ = \{p\}.$$

Let $x_p \in I$ be the x-coordinate of p, i.e., $p = (x_p, \gamma(x_p))$, and define

$$\Gamma_- := \{(x, \gamma(x)) \mid x \in I, x < x_p\},$$
$$\Gamma_+ := \{(x, \gamma(x)) \mid x \in I, x > x_p\}.$$

Then, it follows from $W \cap \mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ = \{p\}$ that the sets Γ_- and Γ_+ each have to be contained entirely in either $\mathcal{B}^\circ$ or $\partial\Omega$ (if this was not the case, there would be an element of $\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$ different from p in W). The latter yields that z is constant in the sense of traces on Γ_- and Γ_+. From Lemma 5.1.21 we obtain that the constants associated with Γ_- and Γ_+ are the same. This proves the claim in case (iii). □

If we combine Lemmas 5.1.31 and 5.1.32, then we arrive at the following result which yields that functions z with (5.33) have a constant trace on each connected component of the set $\partial\{\nabla w = 0\} \cup \partial\Omega$.

Lemma 5.1.33. *Suppose that Assumption 5.1.25 holds, and let $z \in H_0^1(\Omega)$ be a function with* (5.33)*. Then, z is constant in the sense of traces on each component $\mathcal{B}_i$ of the set $\mathcal{B}$. Moreover, the appearing constant is zero for the unique component $\mathcal{B}_i$ with $\mathcal{B}_i \cap \partial\Omega \neq \emptyset$.*

Proof. In what follows, we only consider the unique component $\mathcal{B}_i$ of $\mathcal{B}$ that contains the boundary $\partial\Omega$ (recall that Ω is simply connected and thus has a path-connected boundary by Lemma 5.1.19). The proof for the other components of $\mathcal{B}$ is along the same lines but simpler. From the second point in Assumption 5.1.25 and Lemma 5.1.18, we obtain that

$$\mathcal{B}_i = \Big(\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)\Big) \cup \bigcup_{m\in Z_1} \mathcal{B}_m^\circ \cup \bigcup_{n\in Z_2} \{q_n\},$$

where the unions on the right-hand side are disjoint, where Z_1, Z_2 are suitable index sets, where $\{\mathcal{B}_m^\circ\}$ is the (possibly infinite) collection of all connected components of $\mathcal{B}^\circ$, and where $\{q_n\}$ is the (necessarily finite) collection of all elements of $\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$. In what follows, we restrict our analysis to the case where the sets Z_1, Z_2 are both non-empty. The cases where one of the index sets is empty can be treated analogously. Recall that z is constant in the sense of traces on each $\mathcal{B}_m^\circ$ with a constant $c_m \in \mathbb{R}$ by Lemma 5.1.31, and that z is trivially constant zero in the sense of traces on $\partial\Omega$. Consider now two arbitrary but fixed $\mathcal{B}_{m_1}^\circ, \mathcal{B}_{m_2}^\circ$, $m_1, m_2 \in Z_1$, and some $p_1 \in \mathcal{B}_{m_1}^\circ$, $p_2 \in \mathcal{B}_{m_2}^\circ$. Then, we know from our assumptions that we can find a connected one-dimensional Lipschitz submanifold $\mathcal{M} \subset \mathcal{B}_i$ with $p_1, p_2 \in \mathcal{M}$. From Lemma 5.1.32 we may deduce further that for every $p \in \mathcal{M}$ there exists an open set $D \subset \mathbb{R}^2$ with $p \in D$ such that z is constant in the sense of traces on $D \cap \mathcal{M}$, and that there exist open sets $D_1, D_2 \subset \mathbb{R}^2$ with $p_1 \in D_1$, $p_2 \in D_2$ and $D_s \cap \mathcal{M} = D_s \cap \mathcal{B} = D_s \cap \mathcal{B}_{m_s}^\circ$, $s = 1, 2$. The latter yields in combination with the connectedness of $\mathcal{M}$ that z is constant in the sense of traces along $\mathcal{M}$ and, consequently, that $c_{m_1} = c_{m_2}$ has to hold. Since m_1, m_2 were arbitrary, we obtain that z has to be constant in the sense of traces on $\bigcup_{m\in Z_1} \mathcal{B}_m^\circ$ with some constant c. If we consider now an arbitrary but fixed $\mathcal{B}_m^\circ$, $m \in Z_1$, and the set $\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ) \neq \emptyset$, then we can use exactly the same argumentation as above to obtain that the constant c has to be zero. Suppose now that two non-empty, open, bounded intervals I and J and a Lipschitz continuous map $\gamma : I \to J$ are given such that (after possibly rotating the coordinate system)

$$\Gamma := \{(x, \gamma(x)) \mid x \in I\} \subset \mathcal{B}_i.$$

Then, the set Γ is obviously a one-dimensional connected Lipschitz submanifold, and we may employ exactly the same argumentation as for $\mathcal{M}$ above to deduce that the function z is constant in the sense of traces on Γ. On the other hand, the Lipschitz graph Γ has to intersect the set $\mathcal{B}^\circ \cup (\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ))$ since $\mathcal{B} = \mathcal{B}^\circ \cup (\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)) \cup (\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ)$ and since the set $\mathrm{cl}(\mathcal{B}^\circ) \setminus \mathcal{B}^\circ$ is finite. For every point p in $\Gamma \cap (\mathcal{B}^\circ \cup (\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)))$, however, we obtain from Lemma 5.1.32 that there exists an open set $D \subset \mathbb{R}^2$ such that $p \in D$ and $\Gamma \cap D = D \cap (\mathcal{B}^\circ \cup (\partial\Omega \setminus \mathrm{cl}(\mathcal{B}^\circ)))$, and we know that z is zero in the sense of traces on the latter set. This shows that z is zero in the sense of traces on Γ and proves the claim of the lemma. □

We are now in the position to construct the function $\tilde{z}$ that appears in step two of the strategy that we have outlined at the beginning of Section 5.1.4. In what follows, the basic idea of our analysis is to exploit that the constantness of the trace of z on $\partial\{\nabla w = 0\} \cup \partial\Omega$ allows us to "isolate" the parts $z|_{\mathcal{A}}$ and $z|_{\Omega\setminus\mathcal{A}}$ from each other without losing the H_0^1-regularity, cf. the example in (5.30). We proceed in two steps starting with:

Lemma 5.1.34. *Suppose that Assumption 5.1.25 holds. Let $z \in H_0^1(\Omega)$ be a function with* (5.33)*. Then there exists a $\tilde{z} \in H_0^1(\Omega)$ satisfying*

$$\nabla\tilde{z} = \begin{cases} \nabla z & \text{a.e. in } \mathcal{A}^\circ \\ 0 & \text{a.e. else} \end{cases}.$$

Proof. Consider an arbitrary but fixed connected component $\mathcal{A}_i^\circ$ of the set $\mathcal{A}^\circ$. Then, it follows from our assumptions that $\mathcal{A}_i^\circ$ is a bounded Lipschitz domain and we obtain from Lemmas 5.1.18 and 5.1.20 that the following is true:

(i) The set $\mathbb{R}^2 \setminus \mathrm{cl}(\mathcal{A}_i^\circ)$ consists of finitely many connected components Z_m, $m = 0, ..., M$, and all of these components are Lipschitz domains. Moreover, there is one and only one component, w.l.o.g. Z_0, that is unbounded, and the bounded components $Z_1, ..., Z_M$ are all simply connected.

(ii) The boundary $\partial\mathcal{A}_i^\circ$ is the disjoint union of the boundaries ∂Z_m, $m = 0, ..., M$, and every ∂Z_m is a path-connected, closed, one-dimensional Lipschitz submanifold of $\mathbb{R}^2$.

Note that each boundary $\partial Z_m \subset \partial\Omega \cup \partial\mathcal{A} = \mathcal{B}$ has to be contained in a connected component of the set $\mathcal{B}$. We may thus employ Lemma 5.1.33 to obtain that there exist numbers $c_m \in \mathbb{R}$ such that z is constant c_m in the sense of traces on ∂Z_m for all $m = 0, ..., M$. Consider now the unique function

$$z_{\mathcal{A}_i^\circ} = \begin{cases} 0 & \text{in } Z_0 \\ z - c_0 & \text{in } \mathcal{A}_i^\circ \\ c_m - c_0 & \text{in } Z_m \ \ \forall m = 1, ..., M. \end{cases}$$

Then, $z_{\mathcal{A}_i^\circ} \in H^1(\mathbb{R}^2)$ (since the traces are compatible on all ∂Z_m) and

$$\nabla z_{\mathcal{A}_i^\circ} = \begin{cases} \nabla z & \text{a.e. in } \mathcal{A}_i^\circ \\ 0 & \text{a.e. else} \end{cases}.$$

Recall that, since Ω is a simply connected Lipschitz domain, $\partial\Omega$ is path-connected by Lemma 5.1.19. This implies in combination with Lemma 5.1.20 that the set $\mathbb{R}^2 \setminus \mathrm{cl}(\Omega)$ can only have one connected component, i.e., $\mathbb{R}^2 \setminus \mathrm{cl}(\Omega)$ is connected itself. Since $\mathbb{R}^2 \setminus \mathrm{cl}(\Omega) \subset \mathbb{R}^2 \setminus \mathrm{cl}(\mathcal{A}_i^\circ)$ and since Z_0 is the only unbounded connected component of $\mathbb{R}^2 \setminus \mathrm{cl}(\mathcal{A}_i^\circ)$, it follows that $\mathbb{R}^2 \setminus \mathrm{cl}(\Omega) \subset Z_0$ and, consequently, $\partial\Omega \subset \mathbb{R}^2 \setminus \Omega \subset \mathrm{cl}(Z_0)$. The latter implies that $z_{\mathcal{A}_i^\circ}$ is an element of $H_0^1(\Omega)$. Defining

$$\tilde{z} := \sum_i z_{\mathcal{A}_i^\circ},$$

we now obtain a function with the desired properties. □

By subtracting smooth bump functions, we can refine the last lemma as follows:

Lemma 5.1.35. *Suppose that Assumption 5.1.25 holds and let $z \in H_0^1(\Omega)$ be a function with* (5.33)*. Then, there exists a function $\tilde{z} \in H_0^1(\Omega)$ such that*

$$\nabla\tilde{z} = \begin{cases} \nabla z & \textit{a.e. in } \mathcal{A}^\circ \\ 0 & \textit{a.e. in } \Omega \setminus (E \cup \mathcal{A}^\circ) \\ \textit{sth.} & \textit{a.e. in } E \end{cases}$$

holds for a compact set $E \subset \mathcal{I}$ and such that $z - \tilde{z} = 0$ holds on $\mathcal{B}$ in the sense of traces and on $\mathcal{A}^\circ$ almost everywhere.

Proof. From Lemma 5.1.34, we know that there exists a function $\tilde{z}_1 \in H_0^1(\Omega)$ such that the difference $\tilde{z}_2 := z - \tilde{z}_1$ satisfies

$$\nabla\tilde{z}_2 = \begin{cases} 0 & \text{a.e. in } \mathcal{A}^\circ \\ \nabla z & \text{a.e. else}, \end{cases}$$

i.e., $\tilde{z}_2$ is constant almost everywhere on every connected component $\mathcal{A}_i^\circ$ of $\mathcal{A}^\circ$. Note that for $\tilde{z}_2$ it still holds

$$\int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla\tilde{z}_2)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 < \infty.$$

We may thus invoke Lemma 5.1.33 to deduce that $\tilde{z}_2$ is constant on each $\mathcal{B}_i$ in the sense of traces and zero on the unique $\mathcal{B}_i$ that intersects the boundary $\partial\Omega$. Consider now an arbitrary but fixed connected component $\mathcal{A}_i \subset \Omega$ of the active set $\mathcal{A} = \{\nabla w = 0\} \subset \Omega$. Then, two cases are possible:

(i) If $\mathrm{int}(\mathcal{A}_i) = \emptyset$, then $\mathcal{A}_i \subset \partial\mathcal{A} \subset \mathcal{B}$ and it follows from the connectedness of $\mathcal{A}_i$ that there exists a component $\mathcal{B}_i$ of $\mathcal{B}$ with $\mathcal{A}_i \subset \mathcal{B}_i$. In this case, $\tilde{z}_2$ is obviously constant in the sense of traces on $\mathcal{A}_i$, and if $\partial\Omega \cap \mathrm{cl}(\mathcal{A}_i) \neq \emptyset$, then $\tilde{z}_2$ has to vanish on $\mathcal{A}_i$ by Lemma 5.1.33.

(ii) If $\mathrm{int}(\mathcal{A}_i) \neq \emptyset$, then it follows from the fact that $\mathcal{A}_i$ is a connected component of $\mathcal{A}$, that there exists a finite index set Z with

$$\mathrm{int}(\mathcal{A}_i) = \bigcup_{m \in Z} \mathcal{A}_m^\circ.$$

In this case, we obtain from $\nabla \tilde{z}_2 = 0$ a.e. in $\mathcal{A}^\circ$ that $\tilde{z}_2$ is constant on every $\mathcal{A}_m^\circ$, $m \in Z$. Note that $\partial\mathcal{A}_m^\circ \subset \mathcal{B}$ for all m and that the $\mathcal{A}_m^\circ$ are all Lipschitz domains by our assumptions. This implies that $\partial\mathcal{A}_m^\circ$ consists of finitely many one-dimensional path-connected Lipschitz submanifolds for all $m \in Z$, see Lemma 5.1.20. Using the latter and Lemma 5.1.33, we may deduce that $\tilde{z}_2$ is constant in each of the sets

$$\tilde{\mathcal{A}}_m^i := \mathcal{A}_m^\circ \cup \bigcup_{\mathcal{B}_n \,:\, \mathcal{B}_n \cap \mathrm{cl}(\mathcal{A}_m^\circ) \neq \emptyset} \mathcal{B}_n, \quad m \in Z \tag{5.34}$$

(in the sense of traces on the $\mathcal{B}_n$ and classically on $\mathcal{A}_m^\circ$). Define

$$\tilde{\mathcal{A}}_i := \bigcup_{m \in Z} \tilde{\mathcal{A}}_m^i.$$

Then, $\tilde{\mathcal{A}}_i$ is closed and it holds $\mathcal{A}_i \subset \tilde{\mathcal{A}}_i$. The latter can be seen as follows: If $p \in \mathrm{int}(\mathcal{A}_i)$, then $p \in \tilde{\mathcal{A}}_i$ is obvious, so we only have to prove $\mathcal{A}_i \setminus \mathrm{int}(\mathcal{A}_i) \subset \tilde{\mathcal{A}}_i$. We argue by contradiction. If there is a $p \in \mathcal{A}_i \setminus \mathrm{int}(\mathcal{A}_i)$ with $p \notin \tilde{\mathcal{A}}_i$, then $\mathcal{A}_i \setminus \mathrm{int}(\mathcal{A}_i) \subset \partial\mathcal{A}_i \subset \mathcal{B}$ yields

$$p \in \tilde{\mathcal{B}} := \bigcup_{\mathcal{B}_n : \mathcal{B}_n \cap \mathrm{cl}(\mathrm{int}(\mathcal{A}_i)) = \emptyset} \mathcal{B}_n.$$

Since $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{A}}_i$ are closed and disjoint (recall that the collection $\{\mathcal{B}_n\}$ is finite and that the $\mathcal{B}_n$ are the connected components of the closed set $\mathcal{B}$ and thus closed themselves), we know that $\mathrm{dist}(\tilde{\mathcal{B}}, \tilde{\mathcal{A}}_i) > 0$. Further, our construction yields $\tilde{\mathcal{B}} \cup \tilde{\mathcal{A}}_i = \mathrm{int}(\mathcal{A}_i) \cup \mathcal{B}$, and from the Lipschitz connectedness of $\mathcal{A}_i$, we obtain that there exists a one-dimensional connected Lipschitz submanifold $\mathcal{M}$ satisfying $p \in \mathcal{M}$, $\mathcal{M} \subset \mathcal{A}_i \subset \mathrm{int}(\mathcal{A}_i) \cup \mathcal{B}$ and $\mathcal{M} \cap \tilde{\mathcal{A}}_i \neq \emptyset$. This is a contradiction with the disconnectedness of the set $\tilde{\mathcal{B}} \cup \tilde{\mathcal{A}}_i$, so we indeed have $\mathcal{A}_i \subset \tilde{\mathcal{A}}_i$. Note that the connectedness of the sets $\tilde{\mathcal{A}}_m^i$ in (5.34), the Lipschitz connectedness of $\mathcal{A}_i$, and the fact that the sets $\tilde{\mathcal{A}}_m^i$ all intersect $\mathcal{A}_i$ imply that the set $\tilde{\mathcal{A}}_i$ is connected. Consider now an arbitrary $m_1 \in Z$. Then, we obtain from the connectedness of $\tilde{\mathcal{A}}_i$ that there exists some $\mathcal{B}_n$ such that $\mathcal{B}_n \subset \tilde{\mathcal{A}}_{m_1}^i$ and $\mathcal{B}_n \subset \tilde{\mathcal{A}}_{m_2}^i$ for some $m_2 \in Z \setminus \{m_1\}$ (if this was not the case, we would get a contradiction analogous to that for $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{A}}_i$ above). In particular, $\tilde{z}_2$ is constant on $\tilde{\mathcal{A}}_{m_1}^i \cup \tilde{\mathcal{A}}_{m_2}^i$. Repeating the latter argumentation with $\tilde{\mathcal{A}}_{m_1}^i \cup \tilde{\mathcal{A}}_{m_2}^i$, we obtain that there is an $m_3 \in Z \setminus \{m_1, m_2\}$ such that $\tilde{\mathcal{A}}_{m_1}^i \cup \tilde{\mathcal{A}}_{m_2}^i$ and $\tilde{\mathcal{A}}_{m_3}^i$ share a connected component of $\mathcal{B}$. Iterating, we may now deduce that $\tilde{z}_2$ has to be constant on $\tilde{\mathcal{A}}_i$ and thus also on $\mathcal{A}_i$. Note that, if $\partial\Omega \cap \mathrm{cl}(\mathcal{A}_i) \neq \emptyset$, then $\tilde{\mathcal{A}}_i$ has to contain the unique component of $\mathcal{B}$ that intersects the boundary $\partial\Omega$ and we may deduce that $\tilde{z}_2$ vanishes on $\tilde{\mathcal{A}}_i$ and $\mathcal{A}_i$.

In summary, the above yields that $\tilde{z}_2$ is constant on each connected component $\mathcal{A}_i$ of $\mathcal{A}$ and that $\tilde{z}_2$ vanishes on every $\mathcal{A}_i$ with $\partial\Omega \cap \mathrm{cl}(\mathcal{A}_i) \neq \emptyset$. Consider now the components $\mathcal{A}_i$ of $\mathcal{A}$ with $\partial\Omega \cap \mathrm{cl}(\mathcal{A}_i) = \emptyset$ and denote the corresponding (necessarily finite) index set with Z°. Then, all $\mathcal{A}_i$ with $i \in Z^\circ$ are closed subsets of Ω since ∇w is continuous in Ω, and for all $\mathcal{A}_i$, $i \in Z^\circ$, we can find open sets $D_i, \tilde{D}_i \subset \Omega$ such that $\mathcal{A}_i \subset D_i \subset \mathrm{cl}(D_i) \subset \tilde{D}_i \subset \mathrm{cl}(\tilde{D}_i) \subset \Omega$ and $\mathrm{cl}(\tilde{D}_i) \cap \mathcal{A} \setminus \mathcal{A}_i = \emptyset$. This follows simply from the fact that the sets

$$\mathcal{A}_i \quad \text{and} \quad \partial\Omega \cup \bigcup_{m \neq i} \mathcal{A}_m$$

are compact for every $i \in Z^\circ$. For each $i \in Z^\circ$ we choose a bump function $\psi_i \in C_c^\infty(\Omega)$ such that

$$\operatorname{supp}(\psi_i) \subset \tilde{D}_i, \quad \psi_i \in [0,1] \text{ a.e. in } \Omega, \quad \psi_i \equiv 1 \text{ on } D_i$$

and define

$$\tilde{z} := \tilde{z}_1 + \sum_{i \in Z^\circ} c_i \psi_i,$$

where c_i are the constants in "$\tilde{z}_2$ is constant c_i on $\mathcal{A}_i$". Then, it holds $\tilde{z} \in H_0^1(\Omega)$ and

$$\nabla \tilde{z} = \begin{cases} \nabla \tilde{z}_1 = \nabla z & \text{in } \mathcal{A}^\circ \\ 0 & \text{a.e. in } \Omega \setminus (E \cup \mathcal{A}^\circ) \\ \text{sth.} & \text{a.e. in } E \end{cases}$$

with

$$E := \bigcup_{i \in Z^\circ} \operatorname{cl}\left(\tilde{D}_i \setminus D_i\right) \subset \Omega \setminus \mathcal{A}$$

and $z - \tilde{z} = \tilde{z}_2 - \sum_{i \in Z^\circ} c_i \psi_i = 0$ on all $\mathcal{A}_i^\circ$ and all $\mathcal{B}_i$. This proves the claim. □

Note that the function $\tilde{z}$ in Lemma 5.1.35 has exactly the properties that we need in the second step of the strategy described in Section 5.1.4. By means of an approximation argument, we may now finally prove that Assumption 5.1.25 is sufficient for the density (5.3) in Theorem 5.1.2:

Proposition 5.1.36. *Suppose that (w, φ) is a tuple as in Assumption 5.1.25 and that $\lambda \in L^2(\Omega, \mathbb{R}^2)$ is a multiplier for φ as in* (5.2)*. Let $\mathcal{Z}$, $\mathcal{K}$ and $\|\cdot\|_{\mathcal{K}}$ be defined as in Theorem 5.1.2. Then, it holds*

$$\operatorname{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}.$$

Proof. Let $z \in \mathcal{K}$ be arbitrary but fixed. Then, Lemma 5.1.35 implies that there exists a $\tilde{z} \in H_0^1(\Omega)$ such that

$$\nabla \tilde{z} = \begin{cases} \nabla z & \text{a.e. in } \mathcal{A}^\circ \\ 0 & \text{a.e. in } \Omega \setminus (E \cup \mathcal{A}^\circ) \\ \text{sth.} & \text{a.e. in } E \end{cases}$$

holds for some compact set $E \subset \mathcal{I} = \{\nabla w \neq 0\}$ and such that $z - \tilde{z} = 0$ holds in the sense of traces on $\mathcal{B}$ and almost everywhere on $\mathcal{A}^\circ$. Consider now a function ω and a set D as in Assumption 5.1.25, i.e., suppose that $\omega \in C^{0,1}(\Omega)$, that $D \subset \mathbb{R}^2$ is open, and that

$$\mathcal{A} \cup \partial\Omega \subset D, \qquad \omega = 0 \text{ on } \mathcal{A} \cup \partial\Omega, \qquad \operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega) \leq C\omega \text{ a.e. in } \mathcal{I} \cap D,$$
$$\left(\frac{\nabla w^\perp}{\|\nabla w\|_2} \cdot \frac{\nabla \omega}{\|\nabla \omega\|_2}\right)^2 \leq C\|\nabla w\|_2 \text{ a.e. in } \mathcal{I} \cap D$$

holds for some $C > 0$. Then $\Omega \setminus D$ is a compact subset of $\mathcal{I}$ and we may find a smooth bump function $\psi \in C_c^\infty(\mathbb{R}^2)$ with $\psi \equiv 1$ in a neighborhood of $\Omega \setminus D$, $0 \leq \psi \leq 1$ everywhere, and $\operatorname{supp}(\psi) \subset \mathcal{I}$. Define

$$z_\varepsilon := \tilde{z} + \psi(z - \tilde{z}) + (1 - \psi)(z - \tilde{z}) \min\left(1, \max\left(0, \frac{\omega}{\varepsilon} - \varepsilon\right)\right), \qquad \varepsilon \in (0,1).$$

Then, it holds $z_\varepsilon \in \mathcal{Z}$ for all $\varepsilon \in (0,1)$. This can be seen as follows: Since ω (or its extension on $\operatorname{cl}(\Omega)$, respectively) is Lipschitz continuous, and since $\omega = 0$ on $\mathcal{A} \cup \partial\Omega$, there exist open sets $D_\varepsilon \subset \mathbb{R}^2$ with $\mathcal{A} \cup \partial\Omega \subset D_\varepsilon$ such that $\min(1, \max(0, \omega/\varepsilon - \varepsilon))$ vanishes in $D_\varepsilon \cap \Omega$ for all $\varepsilon \in (0,1)$, i.e., such that

$$\min\left(1, \max\left(0, \frac{\omega}{\varepsilon} - \varepsilon\right)\right) = \begin{cases} 0 & \text{a.e. in } \Omega \setminus E_\varepsilon \\ \text{sth.} & \text{a.e. in } E_\varepsilon \end{cases}$$

holds with the compact set $E_\varepsilon := \Omega \setminus D_\varepsilon \subset \mathcal{I}$. This yields that

$$\nabla z_\varepsilon = \begin{cases} \nabla \tilde{z} = \nabla z & \text{a.e. in } \mathcal{A}^\circ \\ \nabla \tilde{z} = 0 & \text{a.e. in } \Omega \setminus (\operatorname{supp}(\psi) \cup E \cup E_\varepsilon \cup \mathcal{A}^\circ) \\ \text{sth.} & \text{a.e. in } \operatorname{supp}(\psi) \cup E \cup E_\varepsilon \end{cases}.$$

Note that the boundary $\partial\mathcal{A}$ has two-dimensional Lebesgue measure zero by our assumptions so that

$$\nabla z_\varepsilon \cdot \lambda = \|\nabla z_\varepsilon\|_2 \, \mathcal{L}^2\text{-a.e. in } \mathcal{A} \quad \iff \quad \nabla z_\varepsilon \cdot \lambda = \|\nabla z_\varepsilon\|_2 \, \mathcal{L}^2\text{-a.e. in } \mathcal{A}^\circ.$$

Further, we obtain from the continuity of ∇w and the compactness of $\operatorname{supp}(\psi) \cup E \cup E_\varepsilon \subset \mathcal{I}$ that there exist $m_\varepsilon > 0$ with $\|\nabla w\|_2 > m_\varepsilon$ in $\operatorname{supp}(\psi) \cup E \cup E_\varepsilon$ for all $\varepsilon \in (0,1)$. Combining all of this and using the properties of z_ε, $\tilde{z}$ and z, we obtain that we indeed have $z_\varepsilon \in \mathcal{Z}$ for all $\varepsilon \in (0,1)$. (Note that z_ε is even contained in the left-hand side of (5.15)). Consider now the function

$$r_\varepsilon := (1-\psi)(z-\tilde{z}) \left(1 - \min\left(1, \max\left(0, \frac{\omega}{\varepsilon} - \varepsilon\right)\right)\right)$$

whose weak gradient takes the form

$$\begin{aligned} \nabla r_\varepsilon = &- (z-\tilde{z}) \left(1 - \min\left(1, \max\left(0, \frac{\omega}{\varepsilon} - \varepsilon\right)\right)\right) \nabla \psi \\ &+ (1-\psi) \left(1 - \min\left(1, \max\left(0, \frac{\omega}{\varepsilon} - \varepsilon\right)\right)\right) \nabla (z - \tilde{z}) \\ &+ \frac{1}{\varepsilon} (1-\psi)(z-\tilde{z})(-\nabla\omega) \mathbb{1}_{\{\varepsilon^2 < \omega < \varepsilon + \varepsilon^2\}}. \end{aligned}$$

Then, we obtain from the dominated convergence theorem and the properties of ψ, D, ω etc. that

$$\begin{aligned} &\int_{\operatorname{supp}(\psi)} \|\nabla r_\varepsilon\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\operatorname{supp}(\psi)} \frac{(\nabla w^\perp \cdot \nabla r_\varepsilon)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \\ &= \int_{\operatorname{supp}(\psi) \cap D} \|\nabla r_\varepsilon\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\operatorname{supp}(\psi) \cap D} \frac{(\nabla w^\perp \cdot \nabla r_\varepsilon)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \to 0 \end{aligned} \tag{5.35}$$

for $\varepsilon \searrow 0$ (since $\|\nabla w\|_2 > m$ in $\operatorname{supp}(\psi) \subset \mathcal{I}$ and since the set $\operatorname{supp}(\psi) \cap \{\varepsilon^2 < \omega < \varepsilon + \varepsilon^2\} \cap D$ is empty for all sufficiently small $\varepsilon > 0$ due to $\operatorname{dist}(\operatorname{supp}(\psi), \partial\Omega \cup \mathcal{A}) > 0$ and $\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega) \leq C\omega$ a.e. in $\mathcal{I} \cap D$). On the other hand, we have (again by the dominated convergence theorem, Young's inequality, straightforward estimates, $\mathcal{I} \setminus D \subset \operatorname{supp}(\psi)$, and the properties of ω)

$$\begin{aligned} &\int_{\mathcal{I} \setminus \operatorname{supp}(\psi)} \|\nabla r_\varepsilon\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\mathcal{I} \setminus \operatorname{supp}(\psi)} \frac{(\nabla w^\perp \cdot \nabla r_\varepsilon)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \\ &\leq 2 \int_{\mathcal{I} \cap D \cap \{\omega < 2\varepsilon\}} \frac{(z-\tilde{z})^2}{\varepsilon^2} \left(\|\nabla\omega\|_2^2 + \frac{(\nabla w^\perp \cdot \nabla\omega)^2}{\|\nabla w\|_2^3} \right) \mathrm{d}\mathcal{L}^2 + o(1) \\ &\leq C \int_{\mathcal{I} \cap \{\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega) < 2C\varepsilon\}} \frac{(z-\tilde{z})^2}{\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega)^2} \mathrm{d}\mathcal{L}^2 + o(1) \\ &= C \sum_{\mathcal{I}_i} \int_{\mathcal{I}_i \cap \{\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega) < 2C\varepsilon\}} \frac{(z-\tilde{z})^2}{\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega)^2} \mathrm{d}\mathcal{L}^2 + o(1) \\ &= C \sum_{\mathcal{I}_i} \int_{\mathcal{I}_i \cap \{\operatorname{dist}(\cdot, \partial\mathcal{I}_i) < 2C\varepsilon\}} \frac{(z-\tilde{z})^2}{\operatorname{dist}(\cdot, \partial\mathcal{I}_i)^2} \mathrm{d}\mathcal{L}^2 + o(1) \end{aligned} \tag{5.36}$$

for $\varepsilon \searrow 0$ with some positive constant $C = C(\omega)$. Here, we have used that $\partial\mathcal{I}_i \subset \mathcal{A} \cup \partial\Omega$ entails $\operatorname{dist}(\cdot, \mathcal{A} \cup \partial\Omega) = \operatorname{dist}(\cdot, \partial\mathcal{I}_i)$ on $\mathcal{I}_i$. Since the sets $\mathcal{I}_i$ are Lipschitz domains by Assumption 5.1.25,

since $\partial\mathcal{I}_i \subset \partial\mathcal{B}$ and since $z - \tilde{z}$ vanishes in the sense of traces on $\mathcal{B}$, i.e., $z - \tilde{z} \in H_0^1(\mathcal{I}_i)$ for all i, we obtain from Hardy's inequality, [Kinnunen and Martio, 1997, Corollary 3.11], that $(z - \tilde{z})/\mathrm{dist}(\cdot, \partial\mathcal{I}_i)$ is in $L^2(\mathcal{I}_i)$. Consequently, we may again use the dominated convergence theorem to deduce

$$\int_{\mathcal{I}} \|\nabla r_\varepsilon\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\mathcal{I}} \frac{(\nabla w^\perp \cdot \nabla r_\varepsilon)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \to 0 \quad \text{as } \varepsilon \searrow 0$$

from (5.35) and (5.36). The latter yields

$$\begin{aligned}\int_\Omega \|\nabla(z - z_\varepsilon)\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla(z - z_\varepsilon))^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \\ = \int_{\mathcal{I}} \|\nabla r_\varepsilon\|_2^2 \mathrm{d}\mathcal{L}^2 + \int_{\mathcal{I}} \frac{(\nabla w^\perp \cdot \nabla r_\varepsilon)^2}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \to 0\end{aligned}$$

for $\varepsilon \searrow 0$. This shows that our arbitrary but fixed function $z \in \mathcal{K}$ can be approximated in the norm $\|\cdot\|_{\mathcal{K}}$ by functions $z_\varepsilon \in \mathcal{Z}$ and that we indeed have $\mathrm{cl}_{\|\cdot\|_{\mathcal{K}}}(\mathcal{Z}) = \mathcal{K}$. □

As a straightforward consequence of Theorem 5.1.2 and Proposition 5.1.36, we obtain the following differentiability result for the EVI (M) that uses only regularity information about the tuple (w, φ):

Theorem 5.1.37 (Tangible Criterion for Directional Differentiability)**.** *Suppose that Assumption 5.1.1 holds, that the functional j is defined as in* (5.1)*, and that $f \in H^{-1}(\Omega)$ is a right-hand side such that the solution $w := S(f)$ to* (M) *and the associated subgradient $\varphi := \Delta w + f \in \partial j(w)$ satisfy the conditions in Assumption 5.1.25. Then, the solution operator $S : H^{-1}(\Omega) \to H_0^1(\Omega)$ associated with* (M) *is Hadamard directionally differentiable in f and the directional derivatives $\delta := S'(f; g)$, $g \in H^{-1}(\Omega)$, in f are uniquely characterized by the EVI*

$$\delta \in \mathcal{K}, \qquad \int_\Omega \nabla\delta \cdot \nabla(z - \delta)\mathrm{d}\mathcal{L}^2 + \int_{\{\nabla w \neq 0\}} \frac{(\nabla w^\perp \cdot \nabla\delta)(\nabla w^\perp \cdot \nabla(z - \delta))}{\|\nabla w\|_2^3} \mathrm{d}\mathcal{L}^2 \geq \langle g, z - \delta\rangle \qquad \forall z \in \mathcal{K},$$

where $\mathcal{K}$ is defined as in Theorem 5.1.2.

Due to Theorems 1.4.1 and 4.3.16, the last result can also be stated in the following form:

Theorem 5.1.38 (Tangible Criterion for Second-Order Epi-Differentiability)**.** *Suppose that Assumption 5.1.1 holds, that the functional j is defined as in* (5.1)*, and that a tuple $(v, \varphi) \in \mathrm{graph}(\partial j)$ is given such that v and φ satisfy the conditions in Assumption 5.1.25. Assume further that $\lambda \in L^2(\Omega, \mathbb{R}^2)$ is a multiplier for φ as in* (5.2)*. Then, j is twice epi-differentiable in v for φ with*

$$\mathcal{K}_j^{red}(v, \varphi) = \left\{ z \in H_0^1(\Omega) \;\middle|\; \int_{\{\nabla v \neq 0\}} \frac{(\nabla v^\perp \cdot \nabla z)^2}{\|\nabla v\|_2^3} \mathrm{d}\mathcal{L}^2 < \infty,\ \|\nabla z\|_2 = \lambda \cdot \nabla z \text{ a.e. in } \{\nabla v = 0\} \right\}$$

and

$$Q_j^{v,\varphi}(z) = \int_{\{\nabla v \neq 0\}} \frac{(\nabla v^\perp \cdot \nabla z)^2}{\|\nabla v\|_2^3} \mathrm{d}\mathcal{L}^2 \qquad \forall z \in \mathcal{K}_j^{red}(v, \varphi).$$

Let us conclude this section with some remarks on Theorems 5.1.37 and 5.1.38:

Remark 5.1.39.

(i) Theorem 5.1.37 and Lemma 5.1.33 show that the directional derivatives $S'(f; g)$ have constant traces on the components of the set $\partial\{\nabla w = 0\} \cup \partial\Omega$ in the situation of Assumption 5.1.25. This means that, in a first-order model $T(h) := S(f) + S'(f; h - f)$ of the solution operator S, the boundary components of the rigid zones can only move "as a whole", i.e., it is not possible that two points of the same component of the set $\partial\{\nabla w = 0\} \cup \Omega$ begin to flow with different velocities.

(ii) As we have already mentioned, the conditions in Assumption 5.1.25 are, e.g., satisfied in the case of the rotationally symmetric flow depicted in Figure 5.2. This is thus an example of a situation where Theorem 5.1.37 is applicable. Note that the rotational symmetry is only important for the validation of Assumption 5.1.25 in this context. In particular, we may consider all directions g in the derivative $S'(f;g)$ and not only those that are themselves rotationally symmetric.

(iii) Note that Theorems 5.1.37 and 5.1.38 give precisely the same result that Theorem 4.3.3 gives for a function of the form (4.20) *with a surjective inner map G. Compare also with Corollary 4.3.5 in this context. The reader should be warned that this is just a coincidence and that functionals j of the type* (4.20) *can also behave differently. See, e.g., Section 5.2.2 for details.*

5.1.6 Remarks on the Discretized Mosolov Problem

Before we move on to our second example, we would like to mention that the differential sensitivity analysis of Mosolov's problem becomes much simpler when the function space $H_0^1(\Omega)$ in (M) is replaced with an appropriately chosen subspace. To be more precise, we have the following:

Theorem 5.1.40. *Suppose that $\Omega \subset \mathbb{R}^d$, $d \geq 1$, is an open, bounded and non-empty set and assume that V_h is a closed subspace of $H_0^1(\Omega)$ such that the function $\|\nabla v_h\|_2 \in L^2(\Omega)$ is essentially finitely-valued for all $v_h \in V_h$. Then, the solution operator $S_h : V_h^* \to V_h$, $f_h \mapsto w_h$, associated with the EVI*

$$w_h \in V_h, \quad \int_\Omega \nabla w_h \cdot \nabla(v_h - w_h)\mathrm{d}\mathcal{L}^d + \int_\Omega \|\nabla v_h\|_2 \mathrm{d}\mathcal{L}^d - \int_\Omega \|\nabla w_h\|_2 \mathrm{d}\mathcal{L}^d \geq \langle f_h, v_h - w_h\rangle \quad v_h \in V_h \tag{5.37}$$

is well-defined, globally Lipschitz continuous and Hadamard directionally differentiable and the directional derivatives $\delta_h := S_h'(f_h; g_h)$, $g_h \in V_h^*$, in a point $f_h \in V_h^*$ with associated solution $w_h := S_h(f_h)$ are uniquely characterized by the EVI

$$\begin{aligned} &\delta_h \in \mathcal{K}_h, \\ &\int_\Omega \nabla \delta_h \cdot \nabla(z_h - \delta_h)\mathrm{d}\mathcal{L}^d + \int_{\{\nabla w_h \neq 0\}} \frac{(\nabla w_h^\perp \cdot \nabla \delta_h)(\nabla w_h^\perp \cdot \nabla(z_h - \delta_h))}{\|\nabla w_h\|_2^3}\mathrm{d}\mathcal{L}^d \geq \langle g_h, z_h - \delta_h\rangle \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall z_h \in \mathcal{K}_h, \end{aligned}$$

where

$$\mathcal{K}_h := \left\{ z_h \in V_h \,\middle|\, \|\nabla z_h\|_2 = \lambda_h \cdot \nabla z_h \text{ a.e. in } \{\nabla w_h = 0\} \right\}$$

and where $\lambda_h \in L^2(\Omega, \mathbb{R}^d)$ is an arbitrary but fixed multiplier for $\varphi_h := \Delta w_h + f_h \in V_h^*$ as in (4.22).

Proof. The well-definedness and the Lipschitz continuity of the solution map S_h are trivial consequences of Theorem 1.2.2. Consider now an arbitrary but fixed $v_h \in V_h$. Then, the fact that $\|\nabla v_h\|_2$ is essentially finitely-valued implies that $\|\nabla v_h\|_2 > \varepsilon$ holds a.e. in $\{\nabla v_h \neq 0\}$ for some $\varepsilon > 0$, and we may deduce from Theorem 4.3.16 that the functional $j : V_h \to \mathbb{R}$, $v_h \mapsto \int_\Omega \|\nabla v_h\|_2 \mathrm{d}\mathcal{L}^d$, is twice epi-differentiable in v_h for all $\varphi_h \in \partial j(v_h)$ (since the sets $\mathcal{Z}$ and $\mathcal{K}$ in (4.47) and (4.48) are identical). The claim now follows immediately from Theorem 1.4.1 and the explicit formulas in Theorem 4.3.16. □

Note that, if $\mathfrak{T}_h := \{T_h\}$ is a triangulation of Ω (in the sense of [Christof, 2017, Definition 2]), then the standard FE-space $\{v \in C(\mathrm{cl}(\Omega)) \cap H_0^1(\Omega) \mid v|_{T_h} \text{ is affine for all } T_h \in \mathfrak{T}_h\}$ trivially satisfies the conditions in Theorem 5.1.40. The directional differentiability of the solution operator to (M) is thus guaranteed without any additional structural assumptions after a discretization with piecewise linear finite elements. It is further noteworthy that the regularization effect explored in Sections 4.2.2 and 4.3.3 can also be exploited in the context of the EVI (5.37). If we replace the functional $j(v_h) := \int_\Omega \|\nabla v_h\|_2 \mathrm{d}\mathcal{L}^d$ in (5.37), e.g., with $\tilde{j}(v_h) := \int_\Omega \|\nabla v_h\|_2 + \|\nabla v_h\|_2^{3/2} \mathrm{d}\mathcal{L}^d$, then it follows completely analogously to the proof of Theorem 5.1.40 that the solution map of the resulting EVI is Gâteaux differentiable everywhere. We expect that this effect is in particular useful in the study of Casson fluids, cf. [Huilgol and You, 2005].

5.2 The L^1-Norm on $\mathrm{H}^1_0(\Omega)$

As a second example of a "proper" H^1_0-elliptic variational inequality of the second kind, we consider the model problem

$$w \in H^1_0(\Omega), \quad \int_\Omega \nabla w \cdot \nabla(v-w)\mathrm{d}\mathcal{L}^d + \int_\Omega |v|\mathrm{d}\mathcal{L}^d - \int_\Omega |w|\mathrm{d}\mathcal{L}^d \geq \langle f, v-w\rangle \quad \forall v \in H^1_0(\Omega) \quad \text{(L)}$$

that may also be found, e.g., in [Christof and Meyer, 2016; Christof and Wachsmuth, 2017c; De los Reyes and Meyer, 2016; Hintermüller and Surowiec, 2017]. Let us again begin by stating our standing assumptions:

Assumption 5.2.1 (Standing Assumptions for Section 5.2)**.**

- $\Omega \subset \mathbb{R}^d$, $d \geq 2$, *is a bounded Lipschitz domain,*
- $H^1_0(\Omega)$, $H^{-1}(\Omega)$ *and* $\mathcal{L}^d$ *are defined as before,*
- $f \in H^{-1}(\Omega)$ *is a given datum.*

Note that the functional

$$j : H^1_0(\Omega) \to \mathbb{R}, \qquad v \mapsto \int_\Omega |v|\mathrm{d}\mathcal{L}^d, \tag{5.38}$$

in (L) has precisely the form (4.20) and that Proposition 4.3.2(i) yields

$$\partial j(v) = \left\{ \varphi \in H^{-1}(\Omega) \;\middle|\; \exists \lambda \in L^2(\Omega) \text{ s.t. } \lambda \in \partial|v| \text{ a.e. in } \Omega,\ \langle \varphi, z\rangle = \int_\Omega \lambda z \mathrm{d}\mathcal{L}^d \ \ \forall z \in H^1_0(\Omega) \right\} \quad \forall v \in H^1_0(\Omega). \tag{5.39}$$

We may thus try to apply the results of Section 4.3 to prove that j is twice epi-differentiable and that the solution operator $S : H^{-1}(\Omega) \to H^1_0(\Omega)$, $f \mapsto w$, to (L) is directionally differentiable. Unfortunately, proceeding in this way turns out to be not very helpful as the following subsection shows.

5.2.1 Failure of the Chain Rule and the Density Criterion

To see that it is not sensible/possible to use the results of Section 4.3 for the study of the EVI (L), we first note that, in (5.38), the operator G in (4.20) is precisely the embedding $H^1_0(\Omega) \hookrightarrow L^2(\Omega)$. This implies that the surjectivity condition in Theorem 4.3.3 is violated and that the analysis of Section 4.3.1 and the chain rule in Theorem 2.4.8 are inapplicable when we consider the problem (L). Note that the situation is even worse here than in the case of the EVI (M) since the embedding $H^1_0(\Omega) \hookrightarrow L^2(\Omega)$ is not only non-surjective but even compact.

It remains to show that working with the density criterion in Theorem 4.3.16 is not useful, either, when we try to prove the directional differentiability of the solution operator S to (L). To this end, we observe that every solution $w = S(f)$ of (L) with associated subgradient $\varphi = f + \Delta w \in \partial j(w)$ and multiplier $\lambda \in L^\infty(\Omega)$ satisfies $-\Delta w = f - \lambda$ in $H^{-1}(\Omega)$. The latter implies in combination with the classical $W^{2,q}$-regularity theory for the Poisson equation (as found, e.g., in [Gilbarg and Trudinger, 2001; Grisvard, 1985]) that $w \in C^1(\Omega) \cap H^1_0(\Omega)$ holds for all sufficiently regular f and Ω. Consider now the situation in Theorem 4.3.16. Then, for our particular choice of j, the density condition $\mathrm{cl}_{\|\cdot\|_\mathcal{K}}(\mathcal{Z}) = \mathcal{K}$ takes the form

$$\mathrm{cl}_{H^1}\left(\left\{ z \in H^1_0(\Omega) \;\middle|\; \int_{\{v\neq 0\}} \frac{|z|^2}{|v|}\mathrm{d}\mathcal{L}^d < \infty,\ |z| = \lambda z \text{ a.e. in } \{v=0\} \right\} \right) = \left\{ z \in H^1_0(\Omega) \;\middle|\; |z| = \lambda z \text{ a.e. in } \{v=0\} \right\}$$

and we may deduce that for all $v \in H_0^1(\Omega) \cap C(\Omega)$ with $\mathcal{L}^d(\{v = 0\}) = 0$ (and thus also possibly for our solution w) we have to check the equality

$$\mathrm{cl}_{H^1}\left(\left\{ z \in H_0^1(\Omega) \;\middle|\; \int_{\{v\neq 0\}} \frac{|z|^2}{|v|}\mathrm{d}\mathcal{L}^d < \infty \right\}\right) = H_0^1(\Omega) \tag{5.40}$$

to prove the second-order epi-differentiability of j in v for all $\varphi \in \partial j(v)$. Let us assume now that there exists a compact $(d-1)$-dimensional Lipschitz submanifold $\mathcal{N} \subset \{v = 0\} \subset \Omega$ such that

$$|v(x)| \leq C \operatorname{dist}(x, \mathcal{N})^2 \qquad \forall x \in D$$

holds for some constant $C > 0$ and some open neighborhood D of $\mathcal{N}$. Then, we clearly have

$$\int_{D\setminus\mathcal{N}} \frac{|z|^2}{\operatorname{dist}(\cdot,\mathcal{N})^2}\mathrm{d}\mathcal{L}^d \leq C \int_{\{v\neq 0\}} \frac{|z|^2}{|v|}\mathrm{d}\mathcal{L}^d \qquad \forall z \in H_0^1(\Omega)$$

and it follows from [Edmunds and Evans, 1987, Theorem V.3.4] and the continuity of the trace operator that all functions z in the set on the left-hand side of (5.40) have a vanishing trace on $\mathcal{N}$. This shows that (5.40) cannot be satisfied and that Theorem 4.3.16 can indeed be expected to fail when we consider the EVI (L). (Later on we will see that the density criterion (4.49) has to fail for every $v \in C^1(\Omega) \cap H_0^1(\Omega)$ that has a sufficiently regular non-empty zero level set of measure zero, cf. Remark 5.2.16(v).)

5.2.2 Second-Order Epi-Differentiability of the L^1-Norm

Since the functional j in (5.38) is beyond the scope of Theorems 4.3.3 and 4.3.16, we have to go back to the analysis of Chapter 1 and the epi-differentiability criterion in Lemma 1.3.13 to prove the directional differentiability of the solution operator S to (L). To this end, we first note that, for our particular choice of j, Definition 1.3.1 and a simple distinction of cases yield

$$\begin{aligned}
&Q_j^{v,\varphi}(z)\\
&= \inf\left\{ \liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n} - \langle\varphi, z_n\rangle\right) \;\middle|\; \begin{array}{l}\{t_n\}\subset\mathbb{R}^+, \{z_n\}\subset H_0^1(\Omega),\\ t_n \searrow 0,\ z_n \rightharpoonup z\end{array}\right\}\\
&= \inf\left\{ \liminf_{n\to\infty} \left(\int_{\{v>0\}} 4\frac{(-v-t_nz_n)^+}{t_n^2}\mathrm{d}\mathcal{L}^d + \int_{\{v<0\}} 4\frac{(v+t_nz_n)^+}{t_n^2}\mathrm{d}\mathcal{L}^d \right.\right.\\
&\qquad\qquad \left.\left. + \int_{\{v=0\}} 2\frac{|z_n| - \lambda z_n}{t_n}\mathrm{d}\mathcal{L}^d \right) \;\middle|\; \begin{array}{l}\{t_n\}\subset\mathbb{R}^+, \{z_n\}\subset H_0^1(\Omega),\\ t_n \searrow 0,\ z_n \rightharpoonup z\end{array}\right\}
\end{aligned} \tag{5.41}$$

for all $(v, \varphi) \in \operatorname{graph}(\partial j)$, where $\lambda \in L^\infty(\Omega)$ is the (in this case necessarily unique) multiplier for φ as in (5.39) and where $(\cdot)^+$ is short for $\max(0, \cdot)$. Due to the non-negativity of the three integrals in the infimum on the right-hand side of (5.41), the above implies:

Lemma 5.2.2. *For every tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ with associated multiplier $\lambda \in L^\infty(\Omega)$, it holds*

$$\mathcal{K}_j^{red}(v,\varphi) \subset \left\{ z \in H_0^1(\Omega) \;\middle|\; |z| = \lambda z \text{ a.e. in } \{v = 0\}\right\}.$$

Proof. From the non-negativity of the integrals on the right-hand side of (5.41), the compactness of the embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ and a straightforward estimate, we obtain that, for every $z \in \mathcal{K}_j^{red}(v, \varphi)$, we can find sequences $\{z_n\} \subset H_0^1(\Omega)$, $\{t_n\} \subset \mathbb{R}^+$ with $z_n \rightharpoonup z$, $t_n \searrow 0$, and

$$0 = \liminf_{n\to\infty} \int_{\{v=0\}} |z_n| - \lambda z_n \mathrm{d}\mathcal{L}^d = \int_{\{v=0\}} |z| - \lambda z \,\mathrm{d}\mathcal{L}^d.$$

The claim now follows immediately from the properties of λ. □

As Lemma 5.2.2 shows, the third integral on the right-hand side of (5.41) enforces that the elements of the reduced critical cone $\mathcal{K}_j^{red}(v,\varphi)$ satisfy $Q_{|\cdot|}^{v,\lambda}(z)=0$ pointwise a.e. in Ω, where $Q_{|\cdot|}^{x,\eta}$ is the second subderivative of the absolute value function (as appearing in (4.15)). Note that this is precisely what one would expect in view of the analysis of Section 2.5. To get an idea of how the remaining terms in (5.39) behave, we prove the following prototypical result that has, along with the subsequent analysis, already been published in [Christof and Meyer, 2016]:

Proposition 5.2.3. *Let $B_r \subset \mathbb{R}^{d-1}$ be a closed ball of radius $r>0$ and let $a>0$. Define $B_r^\circ := \operatorname{int}(B_r)$ and suppose that $v,\psi \in C(B_r\times[0,a])$ are functions satisfying*

$$v=0 \text{ on } B_r\times\{0\}, \quad v>0 \text{ in } B_r\times(0,a], \quad \text{and} \quad \psi \geq 0 \text{ in } B_r\times[0,a].$$

Assume further that sequences $\{t_n\}\subset\mathbb{R}^+$ and $\{z_n\}\subset H^1(B_r^\circ\times(0,a))$ are given such that $t_n \searrow 0$ and $z_n \rightharpoonup z$ in $H^1(B_r^\circ\times(0,a))$ holds for some function z. Then, the following is true:

(i) If $v\in C^{0,1}(B_r^\circ\times(0,a))$, $\psi>0$ in $B_r\times\{0\}$, and

$$\lim_{t\to 0}\left(\|\nabla v\|_{L^\infty(B_r^\circ\times(0,t))}\right)=0,$$

and if $\operatorname{tr}(z)^-$ is not identical zero on $B_r^\circ\times\{0\}$ (where $(\cdot)^-:=\min(0,\cdot)$), then

$$\liminf_{n\to\infty}\left(\int_{B_r^\circ\times(0,a)}\psi\frac{(-v-t_nz_n)^+}{t_n^2}\mathrm{d}\mathcal{L}^d\right)=\infty.$$

(ii) If $v\in C^1(B_r\times[0,a])$ and $\|\nabla v\|_2\geq\varepsilon>0$ on $B_r\times\{0\}$, and if the sequence $\{z_n\}$ is bounded in $L^\infty(B_r^\circ\times(0,a))$, then it holds

$$\int_{B_r^\circ\times(0,a)}\psi\frac{(-v-t_nz_n)^+}{t_n^2}\mathrm{d}\mathcal{L}^d \to \frac{1}{2}\int_{B_r^\circ\times\{0\}}\psi\frac{(\operatorname{tr}(z)^-)^2}{(\partial_d v)}\mathrm{d}\mathcal{H}^{d-1}.$$

Here, $\operatorname{tr}(z)\in L^2(B_r^\circ\times\{0\},\mathcal{H}^{d-1})$ is the trace of the function z on $B_r^\circ\times\{0\}$.

Proof. We assume w.l.o.g. that $z_n\in C(B_r\times[0,a])$ holds for all $n\in\mathbb{N}$. (If this is not the case, then we replace z_n with a sequence $\tilde{z}_n\in C(B_r\times[0,a])$ that is sufficiently close to z_n in the relevant norms.) In what follows, we denote the first $d-1$ coordinates of the Euclidean space with $x\in\mathbb{R}^{d-1}$ and the d-th coordinate with $y\in\mathbb{R}$. Further, we write $\mathrm{d}x$ and $\mathrm{d}y$ for $\mathrm{d}\mathcal{L}^{d-1}(x)$ and $\mathrm{d}\mathcal{L}^1(y)$.

Ad (i): Suppose that an arbitrary but fixed $C>0$ is given. Then, for all sufficiently large n, it holds $t_nC<a$ and (since $v\geq 0$ in $B_r\times[0,a]$)

$$\int_{B_r^\circ\times(0,a)}\psi\frac{(-v-t_nz_n)^+}{t_n^2}\mathrm{d}\mathcal{L}^d \geq \int_{B_r^\circ}\int_0^{t_nC}\psi\frac{(-z_n)^+}{t_n}-\psi\frac{v}{t_n^2}\mathrm{d}y\mathrm{d}x. \tag{5.42}$$

From $v=0$ on $B_r\times\{0\}$ and $v\in C^{0,1}(B_r^\circ\times(0,a))$, we obtain further

$$\begin{aligned}\left|\int_{B_r^\circ}\int_0^{t_nC}\psi(x,y)\frac{v(x,y)}{t_n^2}\mathrm{d}y\mathrm{d}x\right| &\leq \|\psi\|_{L^\infty}\int_{B_r^\circ}\int_0^{t_nC}\frac{1}{t_n^2}\int_0^y|(\partial_d v)(x,s)|\mathrm{d}s\mathrm{d}y\mathrm{d}x\\ &\leq \frac{1}{2}\|\psi\|_{L^\infty}\|\nabla v\|_{L^\infty(B_r^\circ\times(0,t_nC))}\mathcal{L}^{d-1}(B_r)C^2.\end{aligned} \tag{5.43}$$

Similarly, we may calculate that

$$\begin{aligned}&\int_{B_r^\circ}\int_0^{t_nC}\psi(x,y)\frac{(-z_n(x,y))^+}{t_n}\mathrm{d}y\mathrm{d}x\\ &\quad=\int_{B_r^\circ}\int_0^{t_nC}\psi(x,y)\frac{(-z_n(x,0))^+}{t_n}\mathrm{d}y\mathrm{d}x+R_n\\ &\quad=\int_{B_r^\circ}(-z_n(x,0))^+\int_0^C\psi(x,t_ny)\mathrm{d}y\mathrm{d}x+R_n\end{aligned} \tag{5.44}$$

with

$$\begin{aligned}|R_n| &= \left| \int_{B_r^\circ} \int_0^{t_n C} \psi(x,y) \frac{(-z_n(x,y))^+ - (-z_n(x,0))^+}{t_n} \mathrm{d}y\mathrm{d}x \right| \\ &\leq \|\psi\|_{L^\infty} \int_{B_r^\circ} \int_0^{t_n C} \frac{1}{t_n} \int_0^y |\partial_d z_n(x,s)| \mathrm{d}s\mathrm{d}y\mathrm{d}x \\ &\leq \|\psi\|_{L^\infty} \int_{B_r^\circ} \int_0^{t_n C} \frac{1}{t_n} y^{1/2} \left(\int_0^a |\partial_d z_n(x,s)|^2 \mathrm{d}s \right)^{1/2} \mathrm{d}y\mathrm{d}x \\ &\leq \frac{2}{3} \|\psi\|_{L^\infty} \|z_n\|_{H^1} \mathcal{L}^{d-1}(B_r)^{1/2} C^{3/2} t_n^{1/2}.\end{aligned} \tag{5.45}$$

Using (5.43), (5.44), (5.45), the boundedness of z_n in $H^1(B_r^\circ \times (0,a))$, our assumptions on v, and the compactness of the trace operator $\mathrm{tr} : H^1(B_r^\circ \times (0,a)) \to L^2(B_r^\circ, \mathcal{L}^{d-1}) \cong L^2(B_r^\circ \times \{0\}, \mathcal{H}^{d-1})$ (cf. [Nečas, 2012, Chapter 2, Theorem 6.2]), we can pass to the limit $n \to \infty$ in (5.42) to obtain

$$\liminf_{n\to\infty} \left(\int_{B_r^\circ \times (0,a)} \psi \frac{(-v - t_n z_n)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \right) \geq C \int_{B_r^\circ \times \{0\}} \psi |\operatorname{tr}(z)^-| \mathrm{d}\mathcal{H}^{d-1}. \tag{5.46}$$

Since $C > 0$ was arbitrarily large and since $\psi|\operatorname{tr}(z)^-|$ is non-negative and not identical zero on $B_r^\circ \times \{0\}$ by our assumptions, it follows that the limes inferior in (5.46) has to be infinite. This proves (i).

Ad (ii): The claim in (ii) is obtained similarly to that in (i): Due to the C^1-regularity of v, the function

$$\rho(x,y) := \frac{v(x,y)}{y} = \int_0^1 \partial_d v(x,sy) \mathrm{d}s$$

is continuous, and from $v > 0$ in $B_r \times (0,a]$ and $\|\nabla v\|_2 = \partial_d v = \rho \geq \varepsilon > 0$ on $B_r \times \{0\}$ it follows that ρ is positive everywhere in $B_r \times [0,a]$. Thus, there exists an $\tilde{\varepsilon} > 0$ such that $\rho \geq \tilde{\varepsilon}$ holds everywhere in $B_r \times [0,a]$. On the other hand, the integrand in the integral under consideration can only be non-zero if

$$0 \leq -v(x,y) - t_n z_n(x,y) = -y\rho(x,y) - t_n z_n(x,y),$$

i.e., if it holds

$$0 \leq y \leq t_n \frac{\|z_n\|_{L^\infty}}{\tilde{\varepsilon}} \leq C t_n$$

with a constant C independent of n. This implies that, for all large enough n, we have

$$\begin{aligned}\int_{B_r^\circ \times (0,a)} \psi \frac{(-v - t_n z_n)^+}{t_n^2} \mathrm{d}\mathcal{L}^d &= \int_{B_r^\circ} \int_0^{Ct_n} \psi(x,y) \frac{(-y\rho(x,y) - t_n z_n(x,y))^+}{t_n^2} \mathrm{d}y\mathrm{d}x \\ &= \int_{B_r^\circ} \int_0^C \psi(x,t_n y) \Big(- y\rho(x,t_n y) - z_n(x,t_n y) \Big)^+ \mathrm{d}y\mathrm{d}x \\ &= \int_{B_r^\circ} \int_0^C \psi(x,t_n y) \Big(- y\rho(x,0) - z_n(x,0) \Big)^+ \mathrm{d}y\mathrm{d}x + R_n.\end{aligned} \tag{5.47}$$

Thanks to the continuity of ρ and the boundedness of z_n in $H^1(B_r^\circ \times (0,a))$, we can estimate the remainder R_n by

$$\begin{aligned}|R_n| &\leq \|\psi\|_{L^\infty} \int_{B_r^\circ} \int_0^C |y\rho(x,t_n y) - y\rho(x,0)| + |z_n(x,t_n y) - z_n(x,0)| \mathrm{d}y\mathrm{d}x \\ &\leq \|\psi\|_{L^\infty} \int_{B_r^\circ} \int_0^C \int_0^{Ct_n} |\partial_d z_n(x,s)| \mathrm{d}s\mathrm{d}y\mathrm{d}x + o(1) \\ &\leq \|\psi\|_{L^\infty} \int_{B_r^\circ} \int_0^C (Ct_n)^{1/2} \left(\int_0^a |\partial_d z_n(x,s)|^2 \mathrm{d}s \right)^{1/2} \mathrm{d}y\mathrm{d}x + o(1) \\ &\leq \|\psi\|_{L^\infty} \|z_n\|_{H^1} C^{3/2} \mathcal{L}^{d-1}(B_r)^{1/2}\, t_n^{1/2} + o(1) = o(1).\end{aligned} \tag{5.48}$$

From (5.47), (5.48), and the compactness of the trace operator, it follows

$$\int_{B_r^\circ\times(0,a)} \psi \frac{(-v - t_n z_n)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \to \int_{B_r^\circ} \psi(x,0) \int_0^C \Big(- y\partial_d v(x,0) - \mathrm{tr}(z)(x)\Big)^+ \mathrm{d}y\mathrm{d}x. \tag{5.49}$$

Using $\|\nabla v\|_2 = \partial_d v \geq \varepsilon > 0$ on $B_r \times \{0\}$, we can compute the inner integral on the right-hand side of (5.49). This yields

$$\int_{B_r^\circ\times(0,a)} \psi \frac{(-v - t_n z_n)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \to \frac{1}{2} \int_{B_r^\circ} \psi(x,0) \frac{(\mathrm{tr}(z)(x)^-)^2}{(\partial_d v)(x,0)} \mathrm{d}\mathcal{L}^{d-1}(x)$$

and completes the proof. □

The above result suggests that the first two integrals in the infimum on the right-hand side of (5.41) contribute singular terms to the second subderivative $Q_j^{v,\varphi}(z)$ that take the form of surface integrals over the boundary $\partial\{v = 0\}$. We will see in the following that this is indeed the case provided the function v under consideration is sufficiently regular. To be more precise, we require:

Assumption 5.2.4.

- (Regularity) *It holds* $v \in C^1(\Omega) \cap H_0^1(\Omega)$.
- (Structure of the Active Set) *There exists a set* $\mathcal{C} \subset \partial\{v \neq 0\} \cup \partial\Omega$ *such that*
  - *(i)* $\mathcal{C}$ *is closed and has* $H^1(\mathbb{R}^d)$*-capacity zero, i.e., (cf. [Attouch et al., 2006, Chapter 5.8.2])*

$$0 = \mathrm{cap}_2(\mathcal{C}, \mathbb{R}^d) = \inf\Big\{ \|\phi\|_{H^1} \,\Big|\, \phi \in C_c(\mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d), \\ 0 \leq \phi \leq 1,\ \phi \equiv 1 \text{ in a nbhd. of } \mathcal{C} \Big\},$$

  - *(ii)* $(\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$ *is a (strong)* $(d-1)$*-dimensional Lipschitz submanifold of* $\mathbb{R}^d$*,*
  - *(iii) the sets*

$$\mathcal{N}_+ := \{\nabla v = 0\} \cap \partial\{v > 0\} \setminus \mathcal{C}, \\ \mathcal{N}_- := \{\nabla v = 0\} \cap \partial\{v < 0\} \setminus \mathcal{C}$$

  are relatively open in $(\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$.

Here and in the remainder of this section, we again use the convention that sets of the form $\{v \neq 0\}$, $\{\nabla v = 0\}$ etc. are always defined w.r.t. the continuous representative (when available). Some remarks are in order regarding the conditions in Assumption 5.2.4:

Remark 5.2.5.

(i) As already mentioned, for solutions of the EVI (L)*,* C^1*-regularity can be ensured a priori if the right-hand side* f *and the domain* Ω *are sufficiently regular. The first condition in Assumption 5.2.4 is thus not very restrictive.*

(ii) Since v *is assumed to be a* C^1*-function, the implicit function theorem yields that the set*

$$\mathcal{N}_0 := \partial\{v \neq 0\} \cap \{\nabla v \neq 0\} = \{v = 0\} \cap \{\nabla v \neq 0\}$$

is a $(d-1)$*-dimensional* C^1*-submanifold of* $\mathbb{R}^d$*.*

(iii) Recall that, according to our conventions, sets of the form $\{v = 0\}$*,* $\{\nabla v \neq 0\}$ *etc. are always subsets of* Ω*. This implies in particular that* $\mathcal{N}_0$*,* $\mathcal{N}_+$*, and* $\mathcal{N}_-$ *are contained in* Ω *and do not intersect the boundary* $\partial\Omega$*.*

(iv) Since $\mathcal{N}_+$ and $\mathcal{N}_-$ are relatively open subsets of $(\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$, they are themselves strong $(d-1)$-dimensional Lipschitz submanifolds of $\mathbb{R}^d$.

(v) Since $\mathcal{N}_0$, $\mathcal{N}_+$, and $\mathcal{N}_-$ are strong $(d-1)$-dimensional Lipschitz submanifolds, traces on these sets are well-defined (cf. [Adams, 1975; Nečas, 2012] and Definition 5.1.16).

(vi) Assumption 5.2.4 substantially weakens the assumptions on the active set in [De los Reyes and Meyer, 2016], where the zero level set is not allowed to have $(d-1)$-dimensional components and where the sets $\{v > 0\}$ and $\{v < 0\}$ have to have positive distance from each other. The geometric setting depicted in Figure 5.5, for example, satisfies the conditions in Assumption 5.2.4 but violates the assumptions imposed in [De los Reyes and Meyer, 2016].

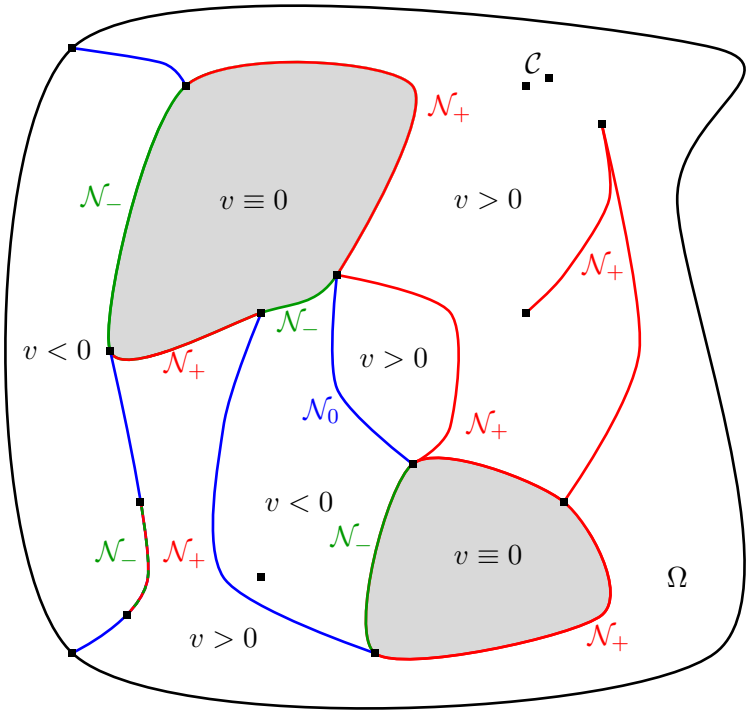

Figure 5.5: Geometric situation in Assumption 5.2.4. All visible lines are part of the boundary $\partial\{v \neq 0\}$. Points contained in $\mathcal{C}$ are marked by black squares. The sets $\mathcal{N}_0$, $\mathcal{N}_-$, and $\mathcal{N}_+$ are depicted in blue, green, and red, respectively. We point out that $\mathcal{N}_+$ and $\mathcal{N}_-$ do not necessarily have to be disjoint and that there is always a change of sign along $\mathcal{N}_0$.

Using the conditions in Assumption 5.2.4, we can prove, for example, the following global version of Proposition 5.2.3(i):

Proposition 5.2.6. *Suppose that a tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ is given such that the function v satisfies Assumption 5.2.4. Then, for every $z \in \mathcal{K}_j^{red}(v, \varphi)$, it holds*

$$\operatorname{tr}(z)^- = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+ \quad \text{and} \quad \operatorname{tr}(z)^+ = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_-.$$

Proof. From the definition of the reduced critical cone and (5.41), we obtain that, for every arbitrary but fixed $z \in \mathcal{K}_j^{red}(v, \varphi)$, we can find sequences $\{t_n\} \subset \mathbb{R}^+$, $\{z_n\} \subset H_0^1(\Omega)$ and a constant $C > 0$ such that $t_n \searrow 0$, $z_n \rightharpoonup z$ and

$$C \geq \int_{\{v>0\}} \frac{(-v - t_n z_n)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \qquad \forall n \in \mathbb{N}. \tag{5.50}$$

Consider now an arbitrary but fixed $p \in \mathcal{N}_+ \subset (\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$. Then, Assumption 5.2.4 implies that (after possibly an orthogonal change of coordinates) we can find a closed ball $B_r \subset \mathbb{R}^{d-1}$ of radius $r > 0$, an open interval J, and a Lipschitz map $\gamma : B_r \to J$ such that

$$p \in \operatorname{int}(B_r \times J) \quad \text{and} \quad (\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C} \cap (B_r \times J) = \{(x, \gamma(x)) \mid x \in B_r\}. \tag{5.51}$$

Note that, since $\mathcal{C}$ is closed, since $\mathcal{N}_+ \subset \Omega$ and since $\mathcal{N}_+$ is relatively open in $(\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$, by making the sets J and B_r smaller, we can always obtain that, in addition to (5.51), the following holds true for some $\varepsilon > 0$:

$$\begin{gathered}\operatorname{cl}(B_r \times J) \subset \Omega \setminus \mathcal{C},\\ \mathcal{N}_+ \cap \operatorname{cl}(B_r \times J) = \partial\{v \neq 0\} \cap \operatorname{cl}(B_r \times J) = \{(x, \gamma(x)) \mid x \in B_r\},\\ \{(x, y) \mid x \in B_r \text{ and } |y - \gamma(x)| < \varepsilon\}) \subset B_r \times J.\end{gathered}$$

In the above situation, the inclusion $\mathcal{N}_+ \subset \partial\{v > 0\}$ yields that v is positive in at least one of the sets

$$D_1 := \{(x, y) \in \operatorname{cl}(B_r \times J) \mid y > \gamma(x)\}, \quad D_2 := \{(x, y) \in \operatorname{cl}(B_r \times J) \mid y < \gamma(x)\}.$$

Let us assume that this is true for D_1 (the other case is analogous). Then, (5.50) and the area formula (cf. [Evans and Gariepy, 2015, Theorem 3.9]) imply

$$C \geq \int_{\operatorname{int}(B_r)} \int_0^{\varepsilon} \left(\frac{(-v - t_n z_n)^+}{t_n^2} \right)\Big|_{(x, y+\gamma(x))} \mathrm{d}\mathcal{L}^1(y) \mathrm{d}\mathcal{L}^{d-1}(x). \tag{5.52}$$

If we define

$$\tilde{v}(x, y) := v(x, y + \gamma(x)), \qquad \tilde{\psi}(x, y) := 1, \qquad \tilde{z}_n(x, y) := z_n(x, y + \gamma(x)),$$

then the right-hand side of (5.52) has exactly the same form as the integral studied in Proposition 5.2.3. To see that Proposition 5.2.3(i) is indeed applicable here, we note that the chain rule for Lipschitz functions (cf. [Ziemer, 1989, Theorem 2.2.2]) implies

$$\tilde{v} \in C^{0,1}(\operatorname{int}(B_r) \times (0, \varepsilon)) \quad \text{and} \quad \|(\nabla \tilde{v})(x, y)\|_2 \leq C \|(\nabla v)(x, y + \gamma(x))\|_2$$

with some constant $C = C(\gamma) > 0$. Since the definition of $\mathcal{N}_+$ yields $(\nabla v)(x, \gamma(x)) = 0$ for all $x \in B_r$, the above together with the continuity of ∇v gives $\lim_{t \to 0} \|\nabla \tilde{v}\|_{L^\infty(\operatorname{int}(B_r) \times (0,t))} = 0$. Using that the mapping $\operatorname{int}(B_r) \times (0, \varepsilon) \ni (x, y) \mapsto (x, y + \gamma(x))$ is bi-Lipschitz and again [Ziemer, 1989, Theorem 2.2.2], we obtain further that $\{\tilde{z}_n\}$ is a bounded sequence in $H^1(\operatorname{int}(B_r) \times (0, \varepsilon))$. This implies in combination with the strong convergence $z_n \to z$ in $L^2(\Omega)$ that $\tilde{z}_n$ converges weakly in $H^1(\operatorname{int}(B_r) \times (0, \varepsilon))$ to $z(x, y + \gamma(x))$. Furthermore, the regularity properties of v and γ immediately give $\tilde{v}, \tilde{\psi} \in C(B_r \times [0, \varepsilon])$. Finally, since v is positive in D_1 and since the definition of $\mathcal{N}_+$ implies $v(x, \gamma(x)) = 0$, we also obtain the remaining conditions needed in Proposition 5.2.3(i), i.e.,

$$\tilde{v} = 0 \text{ on } B_r \times \{0\}, \quad \tilde{v} > 0 \text{ in } B_r \times (0, \varepsilon], \quad \tilde{\psi} > 0 \text{ in } B_r \times [0, \varepsilon].$$

From Proposition 5.2.3 and (5.52) it now follows straightforwardly by contradiction that $\operatorname{tr}(z)^- = 0$ $\mathcal{H}^{d-1}$-a.e. on $\mathcal{N}_+ \cap \operatorname{int}(B_r \times J)$. This, together with the arbitrariness of the point $p \in \mathcal{N}_+$, proves the claim for the negative part of the trace. The result for $\operatorname{tr}(z)^+$ is obtained completely analogously. □

Similarly to the above, we can also generalize part (ii) of Proposition 5.2.3. This leads to:

Proposition 5.2.7. *Suppose that functions v and ψ are given such that*

$$v \in C^1(\Omega), \quad 0 \leq \psi \in C_c(\Omega) \quad \textit{and} \quad \operatorname{supp}(\psi) \cap \partial\{v > 0\} \cap \{\nabla v = 0\} = \emptyset. \tag{5.53}$$

Assume further that $\{t_n\} \subset \mathbb{R}^+$ *and* $\{z_n\} \subset H^1(\Omega)$ *are sequences satisfying*

$$t_n \searrow 0, \quad \|z_n\|_{L^\infty} \leq C \quad \textit{and} \quad z_n \rightharpoonup z \textit{ in } H^1(\Omega)$$

for some constant C *independent of* n *and some* $z \in H^1(\Omega)$*. Then, it holds*

$$\lim_{n\to\infty} \int_\Omega \frac{\psi}{t_n} \left(\frac{|v + t_n z_n^-| - |v|}{t_n} - |\cdot|'(v; z_n^-) \right) \mathrm{d}\mathcal{L}^d = \int_{\{v=0\}\cap\{\nabla v \neq 0\}} \psi \frac{(\mathrm{tr}(z)^-)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}. \tag{5.54}$$

Proof. A simple distinction of cases shows

$$\int_\Omega \frac{\psi}{t_n} \left(\frac{|v + t_n z_n^-| - |v|}{t_n} - |\cdot|'(v; z_n^-) \right) \mathrm{d}\mathcal{L}^d = \int_{\{v>0\}} 2\psi \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \quad \forall n \in \mathbb{N}. \tag{5.55}$$

We thus again end up with an expression similar to that in Proposition 5.2.3. To be able to apply this proposition, we define $E := \mathrm{supp}(\psi)$ and note that (5.53) yields $E \cap \partial\{v > 0\} \subset \{\nabla v \neq 0\}$. The latter implies, together with our assumption $v \in C^1(\Omega)$ and the compactness of the set $E \cap \partial\{v > 0\}$, that there exists a constant $c > 0$ with $\|\nabla v\|_2 \geq c$ on $E \cap \partial\{v > 0\}$. Let us again write $\mathcal{N}_0 := \{v = 0\} \cap \{\nabla v \neq 0\}$. Then, the implicit function theorem yields that for every point $p \in E \cap \partial\{v > 0\} \subset \mathcal{N}_0$ we can find an orthogonal transformation $R_p \in \mathrm{O}(d)$, a closed non-empty ball $B_p \subset \mathbb{R}^{d-1}$, an open interval J_p, and a C^1-function $\gamma_p : B_p \to J_p$ such that

$$\begin{gathered} p \in \mathrm{int}(R_p(B_p \times J_p)), \qquad R_p(B_p \times J_p) \subset \Omega, \\ \mathcal{N}_0 \cap R_p(B_p \times J_p) = \{v = 0\} \cap R_p(B_p \times J_p) = R_p(\{(x, \gamma_p(x)) \mid x \in B_p\}). \end{gathered} \tag{5.56}$$

Note that, since ∇v and γ_p are continuous and since we can make the sets B_p and J_p arbitrarily small, we may assume w.l.o.g. that, in addition to (5.56), we have

$$\begin{aligned} &\mathrm{cl}(R_p(B_p \times J_p)) \subset \Omega, \qquad \|\nabla v\|_2 \geq c/2 \text{ in } R_p(B_p \times J_p), \\ &\quad v \neq 0 \text{ in } \mathrm{cl}(R_p(B_p \times J_p)) \setminus R_p(\{(x, \gamma_p(x)) \mid x \in B_p\}), \\ \text{and} \quad &R_p(\{(x, y) \mid x \in \mathrm{int}(B_p),\, |y - \gamma_p(x)| < \varepsilon_p\}) \subset R_p(B_p \times J_p) \end{aligned} \tag{5.57}$$

for some $\varepsilon_p > 0$. Let us denote the ε_p-tube on the left-hand side of (5.57) with W_p. Then, the collection $\{W_p\}$ defines an open cover of $E \cap \partial\{v > 0\}$ and it follows from the compactness of $E \cap \partial\{v > 0\}$ that there exist points $p_m \in E \cap \partial\{v > 0\}$, $m = 1, ..., M$, $M \in \mathbb{N}$, with

$$E \cap \partial\{v > 0\} \subset \bigcup_{m=1}^M W_{p_m}.$$

Since the set $\bigcup_{m=1}^M W_{p_m}$ is open, we can find an open set $D \subset \mathbb{R}^d$ such that $\bigcup_{m=1}^M W_{p_m} \cup D = \mathbb{R}^d$ and $\mathrm{cl}(D) \cap E \cap \partial\{v > 0\} = \emptyset$. Consider now a partition of unity of the Euclidean space subordinate to the cover $W_{p_1}, ..., W_{p_M}, D$ (cf. [Warner, 1983, Theorem 1.11]), i.e., a collection of smooth functions $\phi_m : \mathbb{R}^d \to [0, 1]$, $m = 1, ..., M + 1$, such that

$$\mathrm{supp}(\phi_m) \subset W_{p_m}, \quad m = 1, ..., M, \quad \mathrm{supp}(\phi_{M+1}) \subset D, \quad \text{and} \quad \sum_{m=1}^{M+1} \phi_m \equiv 1.$$

Then, we obtain

$$\begin{aligned} \int_{\{v>0\}} 2\psi \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d = \sum_{m=1}^M \int_{W_{p_m}\cap\{v>0\}} 2\phi_m \psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \\ + \int_{D\cap E\cap\{v>0\}} 2\phi_{M+1} \psi \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d. \end{aligned} \tag{5.58}$$

Note that $\mathrm{cl}(D) \cap E \cap \partial\{v > 0\} = \emptyset$ implies $\mathrm{cl}(D) \cap E \cap \{v > 0\} = \mathrm{cl}(D) \cap E \cap \mathrm{cl}(\{v > 0\})$. The set $\mathrm{cl}(D) \cap E \cap \{v > 0\}$ is thus a compact subset of $\{v > 0\}$ and the continuity of v gives the existence of an $\varepsilon > 0$ such that $v \geq \varepsilon > 0$ in $\mathrm{cl}(D) \cap E \cap \{v > 0\}$. This, together with $\|z_n\|_{L^\infty} \leq C$ for all n and $t_n \searrow 0$, yields that the integral associated with ϕ_{M+1} in (5.58) is identical zero for all sufficiently large n. It remains to analyze the first M integrals on the right-hand side of (5.58), i.e., the contributions to (5.55) that come from the immediate vicinity of the boundary $\partial\{v > 0\}$. To this end, we consider one of the first M integrals in (5.58), drop the index m, and assume w.l.o.g. that $R_p = \mathrm{Id}$. In this prototypical situation, the integral in question satisfies (cf. (5.52))

$$\begin{aligned}&\int_{W_p \cap \{v>0\}} 2\phi\psi \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \\ &\quad = \int_{\mathrm{int}(B_p)} \int_{-\varepsilon_p}^{\varepsilon_p} \left[\mathbb{1}_{\{v>0\}} 2\phi\psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \right] \Big|_{(x, y+\gamma_p(x))} \mathrm{d}\mathcal{L}^1(y) \mathrm{d}\mathcal{L}^{d-1}(x).\end{aligned}$$

Note that, since $\{v = 0\} \cap W_p \subset \{(x, \gamma_p(x)) \mid x \in B_p\} \subset \{v = 0\} \cap \{\nabla v \neq 0\}$ and due to $v \neq 0$ in $\mathrm{cl}(B_p \times J_p) \setminus \{(x, \gamma_p(x)) \mid x \in B_p\}$, it has to hold either

$$\begin{aligned}&v(x, y + \gamma_p(x)) > 0 \text{ in } B_p \times (0, \varepsilon_p], \quad v(x, y + \gamma_p(x)) < 0 \text{ in } B_p \times [-\varepsilon_p, 0) \\ \text{or} \quad &v(x, y + \gamma_p(x)) < 0 \text{ in } B_p \times (0, \varepsilon_p], \quad v(x, y + \gamma_p(x)) > 0 \text{ in } B_p \times [-\varepsilon_p, 0).\end{aligned}$$

If the first case is true (the second one is analogous), then it holds

$$\begin{aligned}&\int_{W_p \cap \{v>0\}} 2\phi\psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \\ &\quad = \int_{\mathrm{int}(B_p)} \int_0^{\varepsilon_p} \left[2\phi\psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \right] \Big|_{(x, y+\gamma_p(x))} \mathrm{d}\mathcal{L}^1(y) \mathrm{d}\mathcal{L}^{d-1}(x).\end{aligned} \tag{5.59}$$

The integral on the right-hand side of (5.59) has exactly the form of that studied in Proposition 5.2.3(ii). Moreover, the functions $\tilde{v}(x, y) := v(x, y + \gamma_p(x))$, $\tilde{\psi}(x, y) := 2\phi(x, y + \gamma_p(x))\psi(x, y + \gamma_p(x))$, and $\tilde{z}_n(x, y) := z_n^-(x, y + \gamma_p(x))$ satisfy all assumptions of Proposition 5.2.3(ii) (as one can easily check, cf. the proof of Proposition 5.2.6). Recall in this context that the map $H^1(\Omega) \ni u \mapsto u^- \in H^1(\Omega)$ is weakly continuous so that $\tilde{z}_n \rightharpoonup \tilde{z}$ in $H^1(\mathrm{int}(B_p) \times (0, \varepsilon_p))$ with $\tilde{z}(x, y) := z^-(x, y + \gamma_p(x))$. We may thus deduce

$$\int_{W_p \cap \{v>0\}} 2\phi\psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \to \int_{\mathrm{int}(B_p)} \left[\phi\psi \frac{(\mathrm{tr}(z)^-)^2}{\partial_d v} \right] \Big|_{(x, \gamma_p(x))} \mathrm{d}\mathcal{L}^{d-1}(x), \tag{5.60}$$

where $\mathrm{tr}(z)^- = \mathrm{tr}(z^-)$ denotes the trace of z^- on $\mathcal{N}_0$. Using the identity

$$\|(\nabla v)(x, \gamma_p(x))\|_2 = (\partial_d v)(x, \gamma_p(x)) \sqrt{1 + \|\nabla \gamma_p(x)\|_2^2} \quad \forall x \in B_p,$$

which follows from the implicit function theorem and the change of variables formula (cf. [Evans and Gariepy, 2015, Theorem 3.9]), we can rewrite (5.60) as follows:

$$\int_{W_p \cap \{v>0\}} 2\phi\psi \, \frac{(-v - t_n z_n^-)^+}{t_n^2} \mathrm{d}\mathcal{L}^d \to \int_{\mathcal{N}_0} \phi\psi \frac{(\mathrm{tr}(z)^-)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}.$$

Here, we used that $\{(x, \gamma_p(x)) \mid x \in \mathrm{int}(B_p)\} = W_p \cap \mathcal{N}_0$ holds according to (5.56) and (5.57). Since W_{p_m}, $m = 1, ..., M$, covers $E \cap \partial\{v > 0\}$ and since $\sum_{m=1}^M \phi_m \equiv 1$ on $E \cap \partial\{v > 0\} = \mathrm{supp}(\psi) \cap \mathcal{N}_0$, by summation, we finally arrive at

$$\int_\Omega \frac{\psi}{t_n} \left(\frac{|v + t_n z_n^-| - |v|}{t_n} - |\cdot|'(v; z_n^-) \right) \mathrm{d}\mathcal{L}^d \to \int_{\mathcal{N}_0} \psi \frac{(\mathrm{tr}(z)^-)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}.$$

This proves the claim. □

Note that, by exchanging signs, we immediately obtain the following from Proposition 5.2.7:

Corollary 5.2.8. *Suppose that functions v and ψ are given such that*

$$v \in C^1(\Omega), \quad 0 \leq \psi \in C_c(\Omega) \quad \textit{and} \quad \operatorname{supp}(\psi) \cap \partial\{v < 0\} \cap \{\nabla v = 0\} = \emptyset.$$

Assume further that $\{t_n\} \subset \mathbb{R}^+$ and $\{z_n\} \subset H^1(\Omega)$ are sequences satisfying

$$t_n \searrow 0, \quad \|z_n\|_{L^\infty} \leq C \quad \textit{and} \quad z_n \rightharpoonup z \textit{ in } H^1(\Omega)$$

for some constant C independent of n and some $z \in H^1(\Omega)$. Then, it holds

$$\lim_{n\to\infty} \int_\Omega \frac{\psi}{t_n} \left(\frac{|v + t_n z_n^+| - |v|}{t_n} - |\cdot|'(v; z_n^+) \right) \mathrm{d}\mathcal{L}^d = \int_{\{v=0\}\cap\{\nabla v \neq 0\}} \psi \frac{(\operatorname{tr}(z)^+)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}. \tag{5.61}$$

Proof. Write $z_n^+ = -(-z_n)^-$ and invoke (5.54). □

Before we turn our attention to the consequences that the last two results have for the functional $Q_j^{v,\varphi}$ in (5.41), we would like to point out that Proposition 5.2.7 and Corollary 5.2.8 are also interesting on their own. They yield, for example, the following non-standard Taylor expansion for the L^1-norm:

Corollary 5.2.9. *Suppose that $v \in C^1(\Omega)$ is a function with $\{v = 0\} \cap \{\nabla v = 0\} = \emptyset$. Then, for all $z \in C_c(\Omega) \cap H^1(\Omega)$ and all $t > 0$, it holds*

$$\int_\Omega |v + tz| \mathrm{d}\mathcal{L}^d = \int_\Omega |v| \mathrm{d}\mathcal{L}^d + t \int_\Omega \operatorname{sgn}(v) z \mathrm{d}\mathcal{L}^d + t^2 \int_{\{v=0\}} \frac{z^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} + o(t^2), \tag{5.62}$$

where the Landau symbol refers to the limit $t \searrow 0$.

Proof. Choose an arbitrary but fixed non-negative $\psi \in C_c(\Omega)$ with $\psi \equiv 1$ in $\operatorname{supp}(z)$. Then, it follows from (5.54) and (5.61) that, for every sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$, we have

$$\begin{aligned}
&\int_\Omega \frac{1}{t_n} \left(\frac{|v + t_n z| - |v|}{t_n} - \operatorname{sgn}(v) z \right) \mathrm{d}\mathcal{L}^d \\
&\quad = \int_\Omega \frac{\psi}{t_n} \left(\frac{|v + t_n z^-| - |v|}{t_n} - \operatorname{sgn}(v) z^- + \frac{|v + t_n z^+| - |v|}{t_n} - \operatorname{sgn}(v) z^+ \right) \mathrm{d}\mathcal{L}^d \\
&\quad \to \int_{\{v=0\}\cap\{\nabla v\neq 0\}} \psi \frac{(z^-)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} + \int_{\{v=0\}\cap\{\nabla v\neq 0\}} \psi \frac{(z^+)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}.
\end{aligned}$$

If we rewrite the above, then the claim follows immediately. □

We remark that expansions similar to (5.62) also appear, e.g., in the study of integrals with highly oscillatory integrands. Details on this topic may be found in [Huybrechs and Olver, 2009; Iserles and Nørsett, 2004; Iserles et al., 2006]. As shown in [Christof and Wachsmuth, 2017b], Corollary 5.2.9 further allows to prove no-gap second-order conditions for bang-bang optimal control problems.

For the second subderivative of our functional j in (5.38), we may now deduce the following:

Proposition 5.2.10. *Suppose that a tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ is given such that the function v satisfies Assumption 5.2.4. Then, for every $z \in \mathcal{K}_j^{red}(v, \varphi)$, it holds*

$$2 \int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} \leq Q_j^{v,\varphi}(z) < \infty.$$

Here, we again use the notation $\mathcal{N}_0 := \{v = 0\} \cap \{\nabla v \neq 0\}$.

Proof. Let $z \in \mathcal{K}_j^{red}(v,\varphi)$ be arbitrary but fixed, let $\{t_n\} \subset \mathbb{R}^+$ and $\{z_n\} \subset H_0^1(\Omega)$ be sequences with $t_n \searrow 0$, $z_n \rightharpoonup z$ and

$$\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z_n)-j(v)}{t_n} - \langle \varphi, z_n\rangle\right) < \infty,$$

and let $\psi_1^m, \psi_2^m \in C_c(\Omega)$ be bump functions satisfying $0 \le \psi_1^m, \psi_2^m \le 1$ in Ω and

$$\begin{aligned}
\psi_1^m &\equiv 1 \text{ in } \{x \in \Omega \mid \operatorname{dist}(x, \mathcal{C} \cup \mathcal{N}_+ \cup \partial\Omega) \ge 1/m\},\\
\psi_2^m &\equiv 1 \text{ in } \{x \in \Omega \mid \operatorname{dist}(x, \mathcal{C} \cup \mathcal{N}_- \cup \partial\Omega) \ge 1/m\},\\
\operatorname{dist}(\operatorname{supp}(\psi_1^m), \mathcal{C} \cup \mathcal{N}_+) > 0, &\quad \operatorname{dist}(\operatorname{supp}(\psi_2^m), \mathcal{C} \cup \mathcal{N}_-) > 0 \quad \forall\, m \in \mathbb{N}.
\end{aligned}$$

Then, we may use exactly the same arguments as in (5.41) and (5.55) to obtain

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z_n)-j(v)}{t_n} - \langle \varphi, z_n\rangle\right)\\
&\ge \liminf_{n\to\infty} \left(\int_{\{v>0\}} 4\frac{(-v-t_n z_n^-)^+}{t_n^2}\mathrm{d}\mathcal{L}^d + \int_{\{v<0\}} 4\frac{(v+t_n z_n^+)^+}{t_n^2}\mathrm{d}\mathcal{L}^d\right)\\
&\ge \liminf_{n\to\infty} \left(\int_{\{v>0\}} 4\psi_1^m\frac{(-v-t_n \max(-m, z_n^-))^+}{t_n^2}\mathrm{d}\mathcal{L}^d\right.\\
&\qquad\qquad \left. + \int_{\{v<0\}} 4\psi_2^m\frac{(v+t_n \min(m, z_n^+))^+}{t_n^2}\mathrm{d}\mathcal{L}^d\right)\\
&\ge \liminf_{n\to\infty} \int_\Omega \psi_1^m \frac{2}{t_n}\left(\frac{|v+t_n\max(-m,z_n^-)|-|v|}{t_n} - |\cdot|'(v;\max(-m,z_n^-))\right)\mathrm{d}\mathcal{L}^d\\
&\qquad + \liminf_{n\to\infty} \int_\Omega \psi_2^m \frac{2}{t_n}\left(\frac{|v+t_n\min(m,z_n^+)|-|v|}{t_n} - |\cdot|'(v;\min(m,z_n^+))\right)\mathrm{d}\mathcal{L}^d.
\end{aligned}$$

Note that Stampacchia's lemma (see [Kinderlehrer and Stampacchia, 2000, Chapter II, Theorem A.1]) implies that $\max(-m, z_n^-) \rightharpoonup \max(-m, z^-)$ and $\min(m, z_n^+) \rightharpoonup \min(m, z^+)$ holds in $H_0^1(\Omega)$ and that our choice of the sequences ψ_1^m, ψ_2^m yields

$$\operatorname{supp}(\psi_1^m) \cap \partial\{v>0\} \cap \{\nabla v = 0\} = \emptyset \quad \text{and} \quad \operatorname{supp}(\psi_2^m) \cap \partial\{v<0\} \cap \{\nabla v = 0\} = \emptyset$$

for all $m \in \mathbb{N}$. We may thus invoke Proposition 5.2.7 and Corollary 5.2.8 to infer

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z_n)-j(v)}{t_n} - \langle \varphi, z_n\rangle\right)\\
&\ge 2\int_{\{v=0\}\cap\{\nabla v\neq 0\}} \psi_1^m \frac{\max(-m, \operatorname{tr}(z)^-)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1} + 2\int_{\{v=0\}\cap\{\nabla v\neq 0\}} \psi_2^m \frac{\min(m, \operatorname{tr}(z)^+)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1}.
\end{aligned}$$

Taking the limes inferior for $m \to \infty$ in the above and using the lemma of Fatou yields

$$\liminf_{n\to\infty} \frac{2}{t_n}\left(\frac{j(v+t_n z_n)-j(v)}{t_n} - \langle \varphi, z_n\rangle\right) \ge 2\int_{\{v=0\}\cap\{\nabla v\neq 0\}} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1}.$$

Since $\{z_n\}$ and $\{t_n\}$ were arbitrary, the claim now follows immediately. □

In summary (see Lemma 5.2.2, Proposition 5.2.6 and Proposition 5.2.10), we have now proved that, for every $v \in H_0^1(\Omega)$ satisfying Assumption 5.2.4 and for every $\varphi \in \partial j(v)$ with associated multiplier $\lambda \in L^\infty(\Omega)$, we have

$$\begin{aligned}
\mathcal{K}_j^{red}(v,\varphi) \subset \Bigg\{ z \in H_0^1(\Omega) \;\Bigg|\; &\operatorname{tr}(z)^- = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+,\ \operatorname{tr}(z)^+ = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_-,\\
&|z| = \lambda z \ \mathcal{L}^d\text{-a.e. in } \{v=0\},\ \int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1} < \infty \Bigg\} =: \mathcal{K}
\end{aligned} \tag{5.63}$$

and

$$Q_j^{v,\varphi}(z) \geq 2\int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1} =: Q(z) \qquad \forall z \in \mathcal{K}. \tag{5.64}$$

In what follows, our aim is to invoke Lemma 1.3.13 with the above $\mathcal{K}$ and Q. Recall that, to be able to do so, we have to construct a set $\mathcal{Z} \subset \mathcal{K}_j^{red}(v,\varphi) \subset \mathcal{K}$ such that

(i) for each $z \in \mathcal{Z}$ and each $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ there exists a sequence $\{z_n\} \subset V$ satisfying

$$z_n \to z \quad \text{and} \quad \frac{2}{t_n}\left(\frac{j(v+t_nz_n)-j(v)}{t_n} - \langle \varphi, z_n\rangle\right) \to Q(z) \qquad \text{as } n \to \infty,$$

(ii) for each $z \in \mathcal{K}$ there exists a sequence $\{z_n\} \subset \mathcal{Z}$ with $z_n \to z$ and

$$Q(z) \geq \liminf_{n\to\infty} Q(z_n).$$

With this in mind, we prove:

Proposition 5.2.11. *Suppose that a tuple $(v,\varphi) \in \operatorname{graph}(\partial j)$ is given such that the function v satisfies Assumption 5.2.4. Let $\lambda \in L^\infty(\Omega)$ be the multiplier associated with φ as in* (5.39) *and define*

$$\begin{aligned}\mathcal{Z} := \Bigg\{ z \in L^\infty(\Omega)\cap H_0^1(\Omega) \,\Bigg|\, & z^- = 0 \ \mathcal{L}^d\text{-a.e. in a nbhd. of } \partial\Omega \cup \mathcal{N}_+ \cup \mathcal{C},\\ & z^+ = 0 \ \mathcal{L}^d\text{-a.e. in a nbhd. of } \partial\Omega \cup \mathcal{N}_- \cup \mathcal{C},\\ & |z| = \lambda z \ \mathcal{L}^d\text{-a.e. in } \{v=0\}, \ \int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1} < \infty \Bigg\},\end{aligned} \tag{5.65}$$

where $\mathcal{N}_0$ again denotes the set $\{v=0\}\cap\{\nabla v \neq 0\}$. Then, for every $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$ and for every $z \in \mathcal{Z}$, we have

$$\frac{2}{t_n}\left(\frac{j(v+t_nz)-j(v)}{t_n} - \langle \varphi, z\rangle\right) \to 2\int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1}.$$

Proof. Consider an arbitrary but fixed function $z \in \mathcal{Z}$ and a sequence $\{t_n\} \subset \mathbb{R}^+$ with $t_n \searrow 0$, and let $D_1, D_2 \subset \mathbb{R}^d$ be open sets with $\partial\Omega \cup \mathcal{N}_+ \cup \mathcal{C} \subset D_1$, $\partial\Omega \cup \mathcal{N}_- \cup \mathcal{C} \subset D_2$, $z^- = 0$ a.e. in $D_1 \cap \Omega$ and $z^+ = 0$ a.e. in $D_2 \cap \Omega$. Then, we can find bump functions $\psi_1, \psi_2 \in C_c(\Omega)$ such that

$$\begin{gathered}0 \leq \psi_1, \psi_2 \leq 1 \text{ in } \Omega, \quad \psi_1 \equiv 1 \text{ in } \Omega \setminus D_1, \quad \psi_2 \equiv 1 \text{ in } \Omega \setminus D_2,\\ \operatorname{dist}(\operatorname{supp}(\psi_1), \mathcal{C}\cup\mathcal{N}_+) > 0, \quad \text{and} \quad \operatorname{dist}(\operatorname{supp}(\psi_2), \mathcal{C}\cup\mathcal{N}_-) > 0.\end{gathered}$$

Recall that the definitions of the sets $\mathcal{N}_+$ and $\mathcal{N}_-$ yield that $\partial\{v>0\}\cap\{\nabla v = 0\} \subset \mathcal{C}\cup\mathcal{N}_+$ and $\partial\{v<0\}\cap\{\nabla v = 0\} \subset \mathcal{C}\cup\mathcal{N}_-$. Using this, the properties of z, λ, ψ_1 and ψ_2, and Proposition 5.2.7 and Corollary 5.2.8, we obtain

$$\begin{aligned}&\frac{2}{t_n}\left(\frac{j(v+t_nz)-j(v)}{t_n} - \langle \varphi, z\rangle\right)\\ &= \int_\Omega \frac{2}{t_n}\left(\frac{|v+t_nz|-|v|}{t_n} - |\cdot|'(v;z)\right)\mathrm{d}\mathcal{L}^d\\ &= \int_\Omega \frac{2\psi_1}{t_n}\left(\frac{|v+t_nz^-|-|v|}{t_n} - |\cdot|'(v;z^-)\right)\mathrm{d}\mathcal{L}^d + \int_\Omega \frac{2\psi_2}{t_n}\left(\frac{|v+t_nz^+|-|v|}{t_n} - |\cdot|'(v;z^+)\right)\mathrm{d}\mathcal{L}^d\\ &\to 2\int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2}\mathrm{d}\mathcal{H}^{d-1}.\end{aligned}$$

This proves the claim. □

To see that the set $\mathcal{Z}$ in (5.65) also has the approximation property in point (iv) of Lemma 1.3.13, we note the following:

Lemma 5.2.12. *Suppose that a function $v \in C^1(\Omega) \cap H^1_0(\Omega)$ satisfying Assumption 5.2.4 is given. Then, for every $z \in H^1_0(\Omega) \cap L^\infty(\Omega)$ with*

$$\mathrm{tr}(z^-) = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+ \quad \text{and} \quad \mathrm{tr}(z^+) = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_- \tag{5.66}$$

there exists a sequence $\{z_n\} \subset H^1_0(\Omega) \cap L^\infty(\Omega)$ such that $z_n \to z$ holds in $H^1(\Omega)$ for $n \to \infty$ and such that

$$\begin{aligned} z_n^- &= 0 \ \mathcal{L}^d\text{-a.e. in a neighborhood of the set } \partial\Omega \cup \mathcal{N}_+ \cup \mathcal{C}, \\ z_n^+ &= 0 \ \mathcal{L}^d\text{-a.e. in a neighborhood of the set } \partial\Omega \cup \mathcal{N}_- \cup \mathcal{C} \end{aligned}$$

for all $n \in \mathbb{N}$.

Proof. Consider an arbitrary but fixed $z \in H^1_0(\Omega) \cap L^\infty(\Omega)$ satisfying (5.66). Then, it follows from our assumptions on $\mathcal{C}$ that there exists a sequence $\{\phi_m\} \subset C_c(\mathbb{R}^d) \cap W^{1,\infty}(\mathbb{R}^d)$ with

$$0 \leq \phi_m \leq 1, \quad \phi_m \equiv 1 \text{ in a neighborhood of } \mathcal{C} \quad \text{and} \quad \|\phi_m\|_{H^1} \to 0 \text{ as } m \to \infty.$$

Define $z_m := (1 - \phi_m)z \in H^1_0(\Omega)$. Then, each z_m vanishes in a neighborhood of $\mathcal{C}$, and we may compute (using the dominated convergence theorem and Friedrichs' inequality)

$$\|z - z_m\|_{H^1} \leq C\|\nabla z - \nabla z_m\|_{L^2} \leq C\|z\|_{L^\infty} \|\phi_m\|_{H^1} + C\|\phi_m \nabla z\|_{L^2} \to 0.$$

The above implies that we may assume w.l.o.g. that $z = 0$ holds $\mathcal{L}^d$-a.e. in a set of the form $\Omega \cap D$, where $D \subset \mathbb{R}^d$ is an open neighborhood of $\mathcal{C}$. Due to Stampacchia's lemma, see [Kinderlehrer and Stampacchia, 2000, Chapter II, Theorem A.1], we further have $z = z^+ + z^-$ with $z^+, z^- \in H^1_0(\Omega)$. This allows us to approximate the positive and the negative part of the function z separately. Let us focus on the positive part z^+ (the argumentation for z^- is along the same lines). Then, the properties of z yield that $\mathrm{tr}(z)^+ = \mathrm{tr}(z^+) = 0$ holds $\mathcal{H}^{d-1}$-a.e. on $(\partial\Omega \cup \mathcal{N}_-) \setminus D$ and that $z^+ = 0$ holds $\mathcal{L}^d$-a.e. in $\Omega \cap D$. Note that Assumption 5.2.4 implies $\mathrm{cl}(\mathcal{N}_-) \setminus \mathcal{N}_- \subset \mathcal{C}$ and that, as a consequence, $(\partial\Omega \cup \mathcal{N}_-) \setminus D$ is a compact subset of the Lipschitz manifold $(\partial\{v \neq 0\} \cup \partial\Omega) \setminus \mathcal{C}$. We may thus cover the set $(\partial\Omega \cup \mathcal{N}_-) \setminus D$ with a finite number of rectification neighborhoods $\mathrm{int}(R(B_r \times J))$ as appearing in Definition 5.1.16 (cf. also with the proof of Proposition 5.2.6). Using this cover, a localization with a partition of unity and classical rectification arguments (as found, e.g., in [Evans, 2010, Theorem 5.5-2], [Nečas, 2012, Theorem 4.10]), it is straightforward to show that z^+ can be approximated in $H^1(\Omega)$ by continuous functions z_n with $\mathrm{dist}(\mathrm{supp}(z_n), \partial\Omega \cup \mathcal{N}_- \cup \mathcal{C}) > 0$. Since the map $H^1_0(\Omega) \ni u \mapsto u^+ \in H^1_0(\Omega)$ is continuous, we may assume w.l.o.g. that the approximating sequence $\{z_n\}$ is non-negative. If we combine all of the above, then we obtain that z^+ can be approximated in $H^1(\Omega)$ by non-negative functions z_n with the desired properties. Using the same argumentation for the negative part z^- and adding the approximating sequences for z^+ and z^- now yields the claim. □

From Lemma 5.2.12, we may deduce:

Proposition 5.2.13. *Suppose that v, φ, λ, $\mathcal{Z}$ and $\mathcal{N}_0$ are as in Proposition 5.2.11, and let $\mathcal{K}$ be defined as in* (5.63)*. Then, for every $z \in \mathcal{K}$ there exists a sequence $\{z_n\} \subset \mathcal{Z}$ with $z_n \to z$ in $H^1(\Omega)$ and*

$$\int_{\mathcal{N}_0} \frac{\mathrm{tr}(z_n)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} \to \int_{\mathcal{N}_0} \frac{\mathrm{tr}(z)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}$$

for $n \to \infty$.

Proof. Note that the properties of λ imply

$$|z| = \lambda z \ \mathcal{L}^d\text{-a.e. in } \{v = 0\} \quad \Longleftrightarrow \quad \begin{aligned} z^- &= 0 \, \mathcal{L}^d\text{-a.e. in } \{v = 0\} \cap \{\lambda \in (-1, 1]\} \\ z^+ &= 0 \, \mathcal{L}^d\text{-a.e. in } \{v = 0\} \cap \{\lambda \in [-1, 1)\} \end{aligned}. \tag{5.67}$$

Consider now an arbitrary but fixed $z \in \mathcal{K}$ and define $\zeta_n := \min(z^+, n) + \max(z^-, -n)$. Then, the definition of $\mathcal{K}$, (5.67) and the continuity of the map $H_0^1(\Omega) \ni u \mapsto u^+ \in H_0^1(\Omega)$ yield that $\{\zeta_n\} \subset \mathcal{K} \cap L^\infty(\Omega)$ and $\zeta_n \to z$ in $H^1(\Omega)$ for $n \to \infty$. Since ζ_n satisfies $\operatorname{tr}(\zeta_n^-) = 0$ $\mathcal{H}^{d-1}$-a.e. on $\mathcal{N}_+$, $\operatorname{tr}(\zeta_n^+) = 0$ $\mathcal{H}^{d-1}$-a.e. on $\mathcal{N}_-$ and $\{\zeta_n\} \subset H_0^1(\Omega) \cap L^\infty(\Omega)$, we may use Lemma 5.2.12 to deduce that there exist functions $\zeta_{n,m} \in H_0^1(\Omega) \cap L^\infty(\Omega)$ with $\zeta_{n,m} \to \zeta_n$ for $m \to \infty$ in $H^1(\Omega)$, $\zeta_{n,m}^- = 0$ a.e. in a neighborhood of $\partial\Omega \cup \mathcal{N}_+ \cup \mathcal{C}$ for all n, m and $\zeta_{n,m}^+ = 0$ a.e. in a neighborhood of $\partial\Omega \cup \mathcal{N}_- \cup \mathcal{C}$ for all n, m. Define

$$z_{n,m} := \min(\zeta_{n,m}^+, \min(z^+, n)) + \max(\zeta_{n,m}^-, \max(z^-, -n)) \in H_0^1(\Omega) \cap L^\infty(\Omega).$$

Then, it holds

$$\begin{aligned} z_{n,m}^- &= 0 \ \mathcal{L}^d\text{-a.e. in a neighborhood of } \partial\Omega \cup \mathcal{N}_+ \cup \mathcal{C}, \\ z_{n,m}^+ &= 0 \ \mathcal{L}^d\text{-a.e. in a neighborhood of } \partial\Omega \cup \mathcal{N}_- \cup \mathcal{C}, \\ z_{n,m}^- &= 0 \ \mathcal{L}^d\text{-a.e. in } \{v = 0\} \cap \{\lambda \in (-1, 1]\}, \\ z_{n,m}^+ &= 0 \ \mathcal{L}^d\text{-a.e. in } \{v = 0\} \cap \{\lambda \in [-1, 1)\}, \\ \frac{\operatorname{tr}(z_{n,m})^2}{\|\nabla v\|_2} \leq \frac{\operatorname{tr}(\zeta_n)^2}{\|\nabla v\|_2} &\leq \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2} \in L^1(\mathcal{N}_0, \mathcal{H}^{d-1}) \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_0 \end{aligned}$$

for all n, m and $z_{n,m} \to \zeta_n$ in $H^1(\Omega)$ for all n as $m \to \infty$. The above implies in combination with the dominated convergence theorem, (5.67) and the definition of the set $\mathcal{Z}$ that

$$\begin{gathered} \mathcal{Z} \ni z_{n,m} \xrightarrow{m\to\infty} \zeta_n \xrightarrow{n\to\infty} z \text{ in } H^1(\Omega), \\ \int_{\mathcal{N}_0} \frac{\operatorname{tr}(z_{n,m})^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} \xrightarrow{m\to\infty} \int_{\mathcal{N}_0} \frac{\operatorname{tr}(\zeta_n)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} \xrightarrow{n\to\infty} \int_{\mathcal{N}_0} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1}. \end{gathered}$$

Choosing a suitable subsequence $\{m_n\}$ and defining $z_n := z_{n,m_n}$ now yields the claim. □

As Proposition 5.2.13 and the previous results show, the sets $\mathcal{K}$ and $\mathcal{Z}$ in (5.63) and (5.65) and the functional Q in (5.64) indeed have the properties that are needed for the application of Lemma 1.3.13. By invoking this result, we arrive at the following main theorem:

Theorem 5.2.14. *Consider the situation in Assumption 5.2.1, let $j : H_0^1(\Omega) \to \mathbb{R}$ be defined as in* (5.38)*, and suppose that a tuple $(v, \varphi) \in \operatorname{graph}(\partial j)$ is given such that the function v satisfies the conditions in Assumption 5.2.4. Let $\lambda \in L^\infty(\Omega)$ be the multiplier associated with φ as in* (5.39)*. Then, the function j is twice epi-differentiable in v for φ with*

$$\mathcal{K}_j^{red}(v, \varphi) = \left\{ z \in H_0^1(\Omega) \;\middle|\; \begin{gathered} \operatorname{tr}(z)^- = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+, \ \operatorname{tr}(z)^+ = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_-, \\ |z| = \lambda z \ \mathcal{L}^d\text{-a.e. in } \{v = 0\}, \ \int_{\{v=0\}\cap\{\nabla v\neq 0\}} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} < \infty \end{gathered} \right\} \tag{5.68}$$

and

$$Q_j^{v,\varphi}(z) = 2 \int_{\{v=0\}\cap\{\nabla v\neq 0\}} \frac{\operatorname{tr}(z)^2}{\|\nabla v\|_2} \mathrm{d}\mathcal{H}^{d-1} \qquad \forall z \in \mathcal{K}_j^{red}(v, \varphi). \tag{5.69}$$

Due to Theorem 1.4.1, the above implies the following for our model problem (L):

Theorem 5.2.15. *Suppose that Assumption 5.2.1 holds, that the functional j is defined as in* (5.38)*, and that $f \in H^{-1}(\Omega)$ is such that the solution $w := S(f)$ to* (L) *satisfies the conditions in Assumption 5.2.4. Let $\lambda \in L^\infty(\Omega)$ be the multiplier associated with the subgradient $\varphi := \Delta w + f \in \partial j(w)$ (defined as in* (5.39)*). Then, the solution operator $S : H^{-1}(\Omega) \to H^1_0(\Omega)$ associated with* (L) *is Hadamard directionally differentiable in f and the directional derivatives $\delta := S'(f; g)$, $g \in H^{-1}(\Omega)$, in f are uniquely characterized by the EVI*

$$\delta \in \mathcal{K}, \quad \int_\Omega \nabla\delta \cdot \nabla(z-\delta)\mathrm{d}\mathcal{L}^d + 2\int_{\{w=0\}\cap\{\nabla w\neq 0\}} \frac{\operatorname{tr}(\delta)\operatorname{tr}(z-\delta)}{\|\nabla w\|_2}\mathrm{d}\mathcal{H}^{d-1} \geq \langle g, z-\delta\rangle \quad \forall z \in \mathcal{K}$$

with

$$\mathcal{K} := \left\{ z \in H^1_0(\Omega) \;\middle|\; \operatorname{tr}(z)^- = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+,\ \operatorname{tr}(z)^+ = 0 \ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_-, \right.$$
$$\left. |z| = \lambda z \ \mathcal{L}^d\text{-a.e. in } \{w=0\},\ \int_{\{w=0\}\cap\{\nabla w\neq 0\}} \frac{\operatorname{tr}(z)^2}{\|\nabla w\|_2}\mathrm{d}\mathcal{H}^{d-1} < \infty \right\}.$$

The last two results are interesting for several reasons:

Remark 5.2.16.

(i) The expression on the right-hand side of (5.69) *is closely related to the pull-back $v \star |\cdot|''$ of the second distributional derivative of the absolute value function $|\cdot|$ by v in the sense of Hörmander, see [Hörmander, 1990, Chapter VI]. To be more precise, we formally have*

$$Q_j^{v,\varphi}(z) = \left\langle (v \star |\cdot|''), z^2 \right\rangle,$$

where the brackets $\langle ., . \rangle$ denote the distributional pairing (cf. [Hörmander, 1990, Example 6.1.5] and [Wong, 2001, Section V]). Recall that, in the classical theory, the pull-back of a distribution by a function $v \in C^1(\Omega)$ is defined by extending the composition map $C_c^\infty(\mathbb{R}) \ni \phi \mapsto \phi(v) \in C^1(\Omega)$ continuously to the space $\mathcal{D}'(\mathbb{R})$ of distributions and that this extension is only possible if the gradient ∇v vanishes nowhere in the domain Ω (see, e.g., [Hörmander, 1990, Theorem 6.1.2]). What can be observed in Theorem 5.2.14 is that, in the second subderivative of the L^1-norm, the terms known from the distributional pull-back $v \star |\cdot|''$ appear everywhere where they make sense and that on the set $\{v = 0\} \cap \{\nabla v = 0\}$, where the classical definition fails, the pull-back terms are replaced with the conditions $\operatorname{tr}(z)^- = 0$ $\mathcal{H}^{d-1}$-a.e. on $\mathcal{N}_+$, $\operatorname{tr}(z)^+ = 0$ $\mathcal{H}^{d-1}$-a.e. on $\mathcal{N}_-$ and $|z| = \lambda z$ $\mathcal{L}^d$-a.e. in $\{v = 0\}$, respectively. That the quantity $v \star |\cdot|''$ emerges in this way when the second-order epi-differentiability of the L^1-norm and the differential stability of the EVI (L) *are studied is remarkable and has, at least to the author's best knowledge, not been observed so far.*

(ii) It should be noted that the surface integrals and the trace conditions in (5.68) *and* (5.69) *do not appear in the results of [De los Reyes and Meyer, 2016] and [Hintermüller and Surowiec, 2017]. The reason for this is that, under the assumptions of these papers, the zero level set $\{v = 0\}$ cannot have $(d-1)$-dimensional components. The significant change that appears in the structure of the second subderivative $Q_j^{v,\varphi}$ when, e.g., [De los Reyes and Meyer, 2016, Assumption 3.2] is relaxed, is rather surprising and shows that the approach in [De los Reyes and Meyer, 2016] and [Hintermüller and Surowiec, 2017] cannot be extended easily to more general situations.*

(iii) We would like to point out that distributional curvature effects similar to those in Theorem 5.2.14 are still present when the space $H^1_0(\Omega)$ in (L) *is replaced with a finite element space. In particular, it is not possible to exploit an effect comparable to that in Section 5.1.6 when problems of the type* (L) *are considered.*

(iv) Note that the variational inequality for the directional derivatives $S'(f;g)$ in Theorem 5.2.15 is again in general a problem in a weighted Sobolev space since the function $1/\|\nabla w\|_2$ typically blows up near the relative boundary of the set $\{w = 0\} \cap \{\nabla w \neq 0\}$. In contrast to the results in, e.g., Theorem 5.1.37 and Corollaries 4.3.5 and 4.3.18, this time, however, the additional integrability condition is a condition on the trace $\mathrm{tr}(z) \in L^0(\{w = 0\} \cap \{\nabla w \neq 0\}, \mathcal{H}^{d-1})$.

(v) Theorem 5.2.14 demonstrates that the density criterion in Theorem 4.3.16 has to be inapplicable when the function j in (5.38) *and, e.g., a $v \in C^1(\Omega) \cap H^1_0(\Omega)$ satisfying $\emptyset \neq \{v = 0\} \subset \{\nabla v \neq 0\}$ are considered (because* (4.50) *can never reproduce the surface integral in* (5.69)*).*

5.3 Comments and Interpretation

Let us conclude this chapter with some general remarks on the results in the previous two sections and the sensitivity analysis of elliptic variational inequalities in Sobolev spaces as a whole.

First of all, we would like to point out that, using Taylor expansions similar to (2.13), it is possible to study the second-order epi-differentiability of functionals of the form $H^1_0(\Omega) \ni v \mapsto \int_\Omega k(\|\nabla v\|_2)\mathrm{d}\mathcal{L}^2$ and $H^1_0(\Omega) \ni v \mapsto \int_\Omega k(|v|)\mathrm{d}\mathcal{L}^d$ with exactly the same instruments that we have used in Sections 5.1 and 5.2. Details on this topic can be found, for example, in [Christof and Meyer, 2016] where the analysis of Section 5.2 has been carried out in a more general setting. Further, it should be noted that Theorems 5.1.38 and 5.2.14 immediately yield differentiability results for all EVIs of the type (M) and (L), respectively, in which the Laplacian is replaced with an arbitrary operator $A : H^1_0(\Omega) \to H^{-1}(\Omega)$ satisfying the conditions in Assumption 1.2.1, cf. Theorem 1.4.1. A more interesting question that arises in this context is whether the rather inconvenient assumptions on the regularity of the involved functions and level sets in Assumptions 5.1.25 and 5.2.4 can be relaxed or dropped entirely. The structure of the second subderivative in (5.69) suggests that, for the functional j in (5.38), this is not the case because the appearing expressions are only sensible when the gradient of the function v under consideration can be evaluated pointwise. Further, it is easy to see that the proof of Proposition 5.2.6 requires a distinct normal direction on the set $\partial\{v \neq 0\} \cap \{\nabla v = 0\}$ so that, at least for the analysis in Section 5.2.2, structural assumptions on the zero level set seem to be unavoidable. This is, of course, different in Theorem 5.1.38, where the formula for the second subderivative is sensible for all $v \in H^1_0(\Omega)$.

Regardless of whether the conditions in Assumptions 5.1.25 and 5.2.4 are optimal or not, we can conclude that the sensitivity analysis of EVIs in Sobolev spaces is rather counterintuitive and offers numerous pitfalls. Note, for example, that Theorem 5.1.38 yields precisely what one would expect from the chain rule in Theorem 2.4.8 while Theorem 5.2.14 gives a formula for the second subderivative that is entirely unanticipated in view of the analysis in Chapter 4. Compare also with Corollary 4.3.6 and Theorem 5.2.15 in this regard and observe that Theorem 5.2.14 implies in particular that the epigraph of the L^1-norm on $H^1_0(\Omega)$ is not polyhedric (which is quite surprising). The main problem that manifests itself here is that, to properly analyze the second subderivatives of non-smooth functionals on, e.g., the space $H^1_0(\Omega)$, one has to work in a fairly unexplored region between capacity theory and distributional analysis. Mathematical tools that allow to study the large variety of effects that may occur in this border area in its entirety seem to be unavailable so far. We remark that, e.g., the analysis of the L^1-norm becomes even more involved when Sobolev spaces with higher derivatives are considered.

Lastly, we would like to point out that Theorem 5.2.15 is directly related to the counterexample (3.17) in Section 3.4.1. As shown in [Christof and Wachsmuth, 2017c], if we dualize the problem (L), then we arrive precisely at an H^{-1}-elliptic variational inequality of the first kind whose admissible set takes the form $\{v \in L^\infty(\Omega) \mid -1 \leq v \leq 1 \ \mathcal{L}^d\text{-a.e. in } \Omega\} \subset H^{-1}(\Omega)$ and is thus exactly of the type (3.17). Using this relation, it is possible to prove that sets with upper and lower bounds in $H^{-1}(\Omega)$ are not only non-polyhedric but even possess positive curvature. The latter implies in particular that the dualization technique in [Sokołowski, 1988; Sokołowski and Zolésio, 1988, 1992] fails in the case of the EVI (L). For more details on this topic, we refer to [Christof and Wachsmuth, 2017c].

6 Applications in Optimal Control

In what follows, we briefly discuss possible applications of the differentiability results that we have proved in the previous chapters. Section 6.1 is first concerned with stationarity conditions for optimal control problems that are governed by elliptic variational inequalities of the first and the second kind. Here, we demonstrate that the EVI for the directional derivatives in point (ii) of Theorem 1.4.1 can be used to derive a strong stationarity system that covers a large variety of different situations. See Theorem 6.1.7, Corollary 6.1.9 and Corollary 6.1.10 for the main results and Section 6.1.1 for some tangible examples. In Section 6.2, we then give some further remarks on applications in the context of necessary and sufficient second-order optimality conditions and optimization algorithms.

6.1 Strong Stationarity Conditions in a General Setting

The prime example of an application area for Theorem 1.4.1 and its corollaries is, of course, the field of optimal control of elliptic variational inequalities of the first and the second kind. To illustrate this, in what follows, we demonstrate that the sensitivity analysis of Chapter 1 can be used to derive strong stationarity conditions in a very general setting. Recall that the property that is nowadays referred to as strong stationarity essentially goes back to [Mignot, 1976] and [Mignot and Puel, 1984] where it was used to study optimal control problems governed by the classical obstacle problem (cf. Corollary 3.4.3). Since these first contributions, the concept of strong stationarity has been employed for the analysis of various EVIs of the first kind, see, e.g., [Herzog et al., 2013; Outrata et al., 2011; Wachsmuth, 2014, 2017], and, more recently, also for the analysis of non-smooth PDEs and EVIs of the second kind, see [Christof et al., 2017; De los Reyes and Meyer, 2016; Meyer and Susu, 2017]. In the present section, we will see that the majority of the results in the aforementioned papers falls under the scope of a more general stationarity system that is an immediate consequence of the EVI in Theorem 1.4.1(ii). Let us consider the following situation:

Assumption 6.1.1 (Standing Assumptions for Section 6.1)**.** *We are given an optimal control problem of the form*

$$\begin{aligned} &\min\ \mathcal{J}(w,f) \\ &\text{s.t. } w = S(f), \quad f \in \mathcal{F}_{ad}, \end{aligned} \tag{O}$$

such that the following is true:

- *V, H are Hilbert spaces such that $V \subset H$ holds and such that the inclusion map is a continuous embedding, i.e., such that there exists a constant $C > 0$ with*

$$\|v\|_H \leq C\|v\|_V \quad \forall v \in V,$$

- *$\mathcal{F}_{ad} \subset H$ is a non-empty, H-closed and convex set (the set of admissible controls),*
- *$\mathcal{J} : V \times H \to \mathbb{R}$ is Fréchet differentiable (the objective functional),*
- *$S : \mathcal{F}_{ad} \to V$, $f \mapsto w$, is continuous (the control-to-state map).*

Note that, for a directionally differentiable control-to-state mapping S, we trivially have the following necessary optimality condition:

Proposition 6.1.2 (Bouligand Stationarity)**.** *Suppose that $\bar{f} \in \mathcal{F}_{ad}$ is locally optimal for the problem* (O) *and that the map $S : \mathcal{F}_{ad} \to V$ is directionally differentiable in $\bar{f}$ in all directions $g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$. Then, it holds*

$$\big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), S'(\bar{f}; g) \big\rangle_V + \big\langle \partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big\rangle_H \geq 0 \qquad \forall g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}). \tag{6.1}$$

Here, $\bar{w} := S(\bar{f})$ is the state associated with the control $\bar{f}$ and $\partial_w \mathcal{J}(\bar{w}, \bar{f}) \in V^*$ and $\partial_f \mathcal{J}(\bar{w}, \bar{f}) \in H^*$ are the partial Fréchet derivatives of the objective function $\mathcal{J}$ in $(\bar{w}, \bar{f})$.

Proof. The map $\mathcal{F}_{ad} \ni f \mapsto (S(f), f) \in V \times H$ is directionally differentiable in $\bar{f}$, and $\mathcal{J}$ is Fréchet and thus Hadamard. Consequently, we may apply the chain rule, see [Bonnans and Shapiro, 2000, Proposition 2.47], to obtain

$$\frac{\mathcal{J}(S(\bar{f} + tg), \bar{f} + tg) - \mathcal{J}(S(\bar{f}), \bar{f})}{t} \to \big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), S'(\bar{f}; g) \big\rangle_V + \big\langle \partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big\rangle_H$$

for all $g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ and $t \searrow 0$. The local optimality now yields the claim. □

The problem with the Bouligand stationarity condition (6.1) (which is sometimes also referred to as first-order optimality condition in primal formulation) is that it is not very tangible and, e.g., not amenable to numerical solution procedures. To obtain a more usable system of equations and inequalities, we introduce the following set of assumptions:

Assumption 6.1.3.

- $\bar{f} \in \mathcal{F}_{ad} \subset H$,
- $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ *is H-dense in H,*
- *S is directionally differentiable in $\bar{f}$ in all directions $g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$, and there exist a Hilbert space $U \subset V$, a convex, U-lsc., proper functional $Q : U \to [0, \infty]$ that is positively homogeneous of degree two, and an $A \in L(U, U^*)$ that is strongly monotone in $\operatorname{dom}(Q)$ (everything possibly dependent on the element $\bar{f}$) such that $\operatorname{cl}_H(U) = H$, such that the inclusion of U into V is continuous and such that for all $g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ the directional derivative $\delta := S'(\bar{f}; g)$ is characterized by the variational inequality*

$$\delta \in U, \quad \langle A\delta, z - \delta \rangle_U + \frac{1}{2} Q(z) - \frac{1}{2} Q(\delta) \geq (g, z - \delta)_H \quad \forall z \in U. \tag{6.2}$$

Here and in the remainder of this section, we always identify H^* with H and we interpret H as a subset of V^* and V^* as a subset of U^* via the adjoint embeddings, i.e.,

$$\begin{aligned} \langle h, v \rangle_V &= (h, v)_H \quad \forall h \in H \ \forall v \in V, \\ \langle v^*, u \rangle_U &= \langle v^*, u \rangle_V \quad \forall v^* \in V^* \ \forall u \in U. \end{aligned}$$

Note that the above conventions imply $\partial_w \mathcal{J}(S(f), f) \in V^* \subset U^*$ and $\partial_f \mathcal{J}(S(f), f) \in H$ for all $f \in \mathcal{F}_{ad}$ so that the condition (6.1) in Proposition 6.1.2 can be rewritten as

$$\big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), S'(\bar{f}; g) \big\rangle_U + \big(\partial_f \mathcal{J}(\bar{w}, \bar{f}), g\big)_H \geq 0 \qquad \forall g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}). \tag{6.3}$$

Observe further that the inequality (6.2) has precisely the form (1.49). Assumption 6.1.3 thus fits exactly to the results that we have obtained in Chapters 1 to 5. Compare also with Corollaries 3.4.3, 3.4.4, 4.3.5, 4.3.6 and 4.3.18, Theorems 5.1.37 and 5.2.15, and the examples in Section 6.1.1 in this context. For the reformulation of (6.3), we need the following three lemmas:

Lemma 6.1.4. *In the situation of Assumption 6.1.3, it holds* $\mathrm{cl}_{U^*}(H) = U^*$.

Proof. Let us assume that H is not a dense subset of U^*. Then, there exists a $u^* \in U^* \setminus \mathrm{cl}_{U^*}(H)$ and we can apply the (strict) separation theorem to construct an $l \in U^{**}$ such that

$$l(h) = 0 < l(u^*) \quad \forall h \in H.$$

Since U is reflexive, the canonical embedding $\iota : U \to U^{**}$, $u \mapsto \langle \cdot, u \rangle_U$, is surjective and we can find a $\tilde{u} \in U \setminus \{0\}$ such that $\iota(\tilde{u}) = l$, i.e., such that

$$\langle h, \tilde{u} \rangle_U = 0 < \langle u^*, \tilde{u} \rangle_U \quad \forall h \in H.$$

Using how H is identified with a subset of U^*, we now obtain

$$\langle h, \tilde{u} \rangle_U = \langle h, \tilde{u} \rangle_V = (h, \tilde{u})_H = 0 \quad \forall h \in H.$$

Since $\tilde{u} \in U \subset H$, the above implies $\tilde{u} = 0$. This is a contradiction. □

Lemma 6.1.5. *Suppose that an $\bar{f} \in \mathcal{F}_{ad}$ satisfying Assumption 6.1.3 is given and define $\bar{w} := S(\bar{f})$. Then, the map $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}) \ni g \mapsto S'(\bar{f}; g) \in U$ can be extended to a Lipschitz continuous and positively homogeneous function $T : U^* \to U$ with $T(U^*) = \mathrm{dom}(\partial Q)$. This extension is unique.*

Proof. If we define $T(g)$ for an arbitrary $g \in U^*$ via

$$T(g) \in U, \quad \langle AT(g), z - T(g) \rangle_U + \frac{1}{2} Q(z) - \frac{1}{2} Q(T(g)) \geq \langle g, z - T(g) \rangle_U \quad \forall z \in U, \tag{6.4}$$

then we get a Lipschitz continuous extension of the map $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}) \ni g \mapsto S'(\bar{f}; g) \in U$ to U^* by Theorem 1.2.2. This proves the existence of a Lipschitz continuous extension. Further, we know that $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ is H-dense in H and that H is U^*-dense in U^*. Thus, there can only be one continuous extension T and this extension is necessarily positively homogeneous (due to the positive homogeneity of the derivative $S'(\bar{f}; \cdot)$). The identity $T(U^*) = \mathrm{dom}(\partial Q)$ is trivial. This completes the proof. □

Lemma 6.1.6. *Suppose that an $\bar{f} \in \mathcal{F}_{ad}$ satisfying Assumption 6.1.3 and condition* (6.3) *is given. Define $\bar{w} := S(\bar{f})$ and $\bar{p} := -\partial_f \mathcal{J}(\bar{w}, \bar{f})$. Then, $\bar{p}$ is an element of U and it holds*

$$\big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), T(g) \big\rangle_U - \langle g, \bar{p} \rangle_U \geq 0 \qquad \forall g \in U^*. \tag{6.5}$$

Proof. We know that

$$\big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), T(g) \big\rangle_U + \big(\partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big)_H \geq 0 \qquad \forall g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}), \tag{6.6}$$

where $T : U^* \to U$ is the extension of $S'(\bar{f}; \cdot)$ in Lemma 6.1.5. Further, the Lipschitz continuity of T implies that there exists a constant $C > 0$ with

$$\|T(g)\|_U = \|T(g) - S'(\bar{f}; 0)\|_U = \|T(g) - T(0)\|_U \leq C \|g\|_{U^*} \quad \forall g \in U^*.$$

Thus,

$$\big(-\partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big)_H \leq \big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), T(g) \big\rangle_U \leq \tilde{C} \|g\|_{U^*} \quad \forall g \in \mathbb{R}^+(\mathcal{F}_{ad} - \bar{f}).$$

Since $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ is H-dense (and thus also U^*-dense) in H, it follows

$$\big| \big(-\partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big)_H \big| \leq \tilde{C} \|g\|_{U^*} \quad \forall g \in H.$$

The theorem of Hahn-Banach and the reflexivity of U now yield that there exists a $\bar{p} \in U$ with

$$\big(-\partial_f \mathcal{J}(\bar{w}, \bar{f}), g \big)_H = \langle g, \bar{p} \rangle_U = (g, \bar{p})_H = (\bar{p}, g)_H \quad \forall g \in H.$$

This proves $\bar{p} = -\partial_f \mathcal{J}(\bar{w}, \bar{f}) \in U$. The inequality (6.5) follows immediately from (6.6), the H-density of $\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})$ in H and the U^*-density of H in U^*. □

We are now in the position to prove:

Theorem 6.1.7 (Strong Stationarity in Prototypical Form)**.** *In the situation of Assumption 6.1.1, the following holds true:*

(i) For every $\bar{f} \in \mathcal{F}_{ad}$ that satisfies Assumption 6.1.3 and the Bouligand stationarity condition (6.3)*, there exist a unique adjoint state $\bar{p} \in U$ and a unique multiplier $\bar{\eta} \in U^*$ such that*

$$\begin{aligned} A^*\bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\ \langle \bar{\eta}, u \rangle_U \geq \frac{Q(u) - Q(u - t\bar{p})}{2t} \qquad & \forall u \in \operatorname{dom}(\partial Q) \quad \forall t \in (0, \infty). \end{aligned} \tag{6.7}$$

Here, $\bar{w} := S(\bar{f})$ again denotes the state associated with $\bar{f}$.

(ii) If $\bar{f} \in \mathcal{F}_{ad}$ satisfies Assumption 6.1.3 and the system (6.7) *with an adjoint state $\bar{p} \in \operatorname{dom}(\partial Q)$ and a multiplier $\bar{\eta} \in U^*$, and if the set $\operatorname{dom}(\partial Q)$ is convex, then it also holds*

$$\begin{aligned} A^*\bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\ \langle \bar{\eta}, u \rangle_U \geq \frac{1}{2} Q'(u; \bar{p}) \qquad & \forall u \in \operatorname{dom}(\partial Q). \end{aligned} \tag{6.8}$$

(iii) If $\bar{f} \in \mathcal{F}_{ad}$ satisfies Assumption 6.1.3 and the system (6.8) *with a $\bar{p} \in \operatorname{dom}(Q)$ and an $\bar{\eta} \in U^*$, then $\bar{f}$ is Bouligand stationary in the sense of* (6.3)*.*

Proof. Ad (i): Suppose that an $\bar{f} \in \mathcal{F}_{ad}$ satisfying the assumptions in (i) is given. Then, Lemma 6.1.6 implies that $\bar{p} := -\partial_f \mathcal{J}(\bar{w}, \bar{f})$ is an element of U and we may define $\bar{\eta} := \partial_w \mathcal{J}(\bar{w}, \bar{f}) - A^*\bar{p} \in U^*$. From (6.4) with $z = u - t\bar{p} \in U$, $t > 0$, $u = T(g)$, $g \in U^*$, T as in Lemma 6.1.5, and (6.5), we obtain further that

$$\langle Au, -t\bar{p} \rangle_U + \frac{1}{2} Q(u - t\bar{p}) - \frac{1}{2} Q(u) \geq \langle g, -t\bar{p} \rangle_U \geq -t \left\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), u \right\rangle_U .$$

Due to the definition of $\bar{\eta}$, the above can be rewritten as

$$\langle \bar{\eta}, u \rangle_U \geq \frac{Q(u) - Q(u - t\bar{p})}{2t} \qquad \forall u \in T(U^*) = \operatorname{dom}(\partial Q) \quad \forall t \in (0, \infty).$$

The system (6.7) now follows immediately. Note that $\bar{p}$ and $\bar{\eta}$ are trivially unique (since (6.7) does not leave any choice). This proves the first part of the theorem.

Ad (ii): Consider an arbitrary but fixed $\bar{f} \in \mathcal{F}_{ad}$ that satisfies the assumptions in part (ii) of the theorem with some $\bar{p} \in \operatorname{dom}(\partial Q)$ and an $\bar{\eta} \in U^*$, and let T again denote the extension in Lemma 6.1.5. Then, the convexity of $\operatorname{dom}(\partial Q)$, the identity $T(U^*) = \operatorname{dom}(\partial Q)$ and the positive homogeneity of T imply that the set $\operatorname{dom}(\partial Q)$ is a convex cone, and we may deduce that for every $u \in \operatorname{dom}(\partial Q)$ and every $t > 0$, we have

$$u + t\bar{p} = (t+1) \left(\left(1 - \frac{t}{t+1} \right) u + \frac{t}{t+1} \bar{p} \right) \in \operatorname{dom}(\partial Q). \tag{6.9}$$

From (6.7), it now follows

$$\langle \bar{\eta}, u + t\bar{p} \rangle_U \geq \frac{Q(u + t\bar{p}) - Q(u)}{2t} \quad \forall u \in \operatorname{dom}(\partial Q) \quad \forall t > 0.$$

Letting $t \searrow 0$ in the above yields

$$\langle \bar{\eta}, u \rangle_U \geq \frac{1}{2} Q'(u; \bar{p}) \qquad \forall u \in \operatorname{dom}(\partial Q).$$

The claim now follows immediately.

Ad (iii): To prove the third assertion, we note that (6.4), (6.8), our assumption $\bar{p} \in \operatorname{dom}(Q)$ and the properties of Q yield that for every $g \in U^*$ with associated $u := T(g) \in \operatorname{dom}(\partial Q) = T(U^*) \subset \operatorname{dom}(Q)$ and every $t > 0$, we have $u + t\bar{p} \in \operatorname{dom}(Q)$ and

$$\begin{aligned}
\left\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), u \right\rangle_U &= \langle A^* \bar{p} + \bar{\eta}, u \rangle_U \\
&= \frac{1}{t} \left(\langle Au, u + t\bar{p} - u \rangle_U + \frac{1}{2} Q(u + t\bar{p}) - \frac{1}{2} Q(u) - \langle g, t\bar{p} \rangle_U \right) \\
&\qquad + \langle g, \bar{p} \rangle_U + \langle \bar{\eta}, u \rangle_U + \frac{Q(u) - Q(u + t\bar{p})}{2t} \\
&\geq \langle g, \bar{p} \rangle_U + \frac{1}{2} Q'(u; \bar{p}) + \frac{Q(u) - Q(u + t\bar{p})}{2t}.
\end{aligned}$$

Letting $t \searrow 0$ in the above and using that $\bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) = 0$, we arrive at (6.5) and, consequently, at the inequality (6.3). This completes the proof. □

Remark 6.1.8.

(i) Note that the directional derivative $Q'(u; \bar{p})$ in Theorem 6.1.7 always exists in the classical sense (due to the convexity and the positive homogeneity of Q and the assumptions $\bar{p} \in \operatorname{dom}(\partial Q)$ and $\bar{p} \in \operatorname{dom}(Q)$, respectively, cf. (6.9)*).*

(ii) We would like to point out that the domain of the subdifferential of a convex proper and lower semicontinuous function does not necessarily have to be convex. See, e.g., [Borwein and Zhu, 2005, Exercise 4.2.6] for a counterexample. The convexity of the set $\operatorname{dom}(\partial Q)$ *in Theorem 6.1.7(ii) is thus an actual assumption (although a very weak one).*

Before we demonstrate that the systems (6.7) and (6.8) indeed generalize the classical notion of strong stationarity, we state two corollaries of Theorem 6.1.7:

Corollary 6.1.9. *Consider the situation in Assumption 6.1.1 and suppose that an $\bar{f} \in \mathcal{F}_{ad}$ is given such that the conditions in Assumption 6.1.3 are satisfied and such that* $\operatorname{dom}(\partial Q) = U$*. Then, $\bar{f}$ is Bouligand stationary in the sense of* (6.3) *if and only if there exist a $\bar{p} \in U$ and an $\bar{\eta} \in U^*$ with*

$$\begin{aligned}
A^* \bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\
\bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\
\langle \bar{\eta}, u \rangle_U \geq -\frac{1}{2} Q'(u; -\bar{p}) \quad \textit{and} \quad \langle \bar{\eta}, u \rangle_U &\geq \frac{1}{2} Q'(u; \bar{p}) \qquad \forall u \in U.
\end{aligned} \tag{6.10}$$

Proof. If $\bar{f}$ is Bouligand stationary, then Theorem 6.1.7 and the identity $\operatorname{dom}(Q) = \operatorname{dom}(\partial Q) = U$ yield that (6.7) and (6.8) hold, and we may pass to the limit $t \searrow 0$ in (6.7) to obtain the additional inequality in the last line of (6.10). This proves that (6.3) entails (6.10). The reverse implication follows immediately from part (iii) of Theorem 6.1.7. □

Corollary 6.1.10. *Consider the situation in Assumption 6.1.1 and suppose that an $\bar{f} \in \mathcal{F}_{ad}$ is given such that Assumption 6.1.3 is satisfied and such that Q is the characteristic function of a U-closed, convex, non-empty cone $L \subset U$. Then, $\bar{f}$ is Bouligand stationary in the sense of* (6.3) *if and only if there exist a $\bar{p} \in U$ and an $\bar{\eta} \in U^*$ with*

$$\begin{aligned}
A^* \bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\
\bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\
\bar{p} \in L, \quad \langle \bar{\eta}, u \rangle_U &\geq 0 \quad \forall u \in L.
\end{aligned} \tag{6.11}$$

Proof. From part (iii) of Theorem 6.1.7 and the identity $\mathrm{dom}(\partial Q) = \mathrm{dom}(Q) = L$, we immediately obtain that (6.11) entails (6.3). To obtain the reverse implication, we prove that, for every arbitrary but fixed $\bar{f} \in \mathcal{F}_{ad}$ satisfying the assumptions of the corollary and the Bouligand stationarity condition (6.3), we have $\bar{p} := -\partial_f \mathcal{J}(\bar{w}, \bar{f}) \in L$. Let $u := T(A\bar{p})$ denote the solution to (6.4) with right-hand side $g = A\bar{p}$, i.e., the solution to

$$u \in L, \quad \langle Au, z - u\rangle_U \geq \langle A\bar{p}, z - u\rangle_U \quad \forall z \in L. \tag{6.12}$$

Then, we may choose the test functions $z = 0 \in L$ and $z = 2u \in L$ in (6.12) to deduce that

$$\langle Au - A\bar{p}, u\rangle_U = 0.$$

Using this identity in (6.12) yields

$$\langle 0, z - 0\rangle_U \geq \langle A\bar{p} - Au, z - 0\rangle_U \quad \forall z \in L,$$

i.e., it holds $T(A\bar{p} - Au) = 0$, where T is again the solution operator to (6.4). On the other hand, we know from (6.5) that

$$\big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), T(g)\big\rangle_U - \langle g, \bar{p}\rangle_U \geq 0 \qquad \forall g \in U^*.$$

The above implies

$$0 = \big\langle \partial_w \mathcal{J}(\bar{w}, \bar{f}), T(A\bar{p} - Au)\big\rangle_U \geq \langle A\bar{p} - Au, \bar{p}\rangle_U = \langle A\bar{p} - Au, \bar{p} - u\rangle_U,$$

and, due to the strong monotonicity of A, $u = T(A\bar{p}) = \bar{p}$. This shows that $\bar{p} := -\partial_f \mathcal{J}(\bar{w}, \bar{f})$ is indeed an element of $L = \mathrm{dom}(\partial Q)$. The claim now follows straightforwardly from Theorem 6.1.7(i), (ii). □

6.1.1 Some Tangible Examples

To develop intuition for the optimality conditions (6.7) and (6.8), and to see that it indeed makes sense to talk about strong stationarity in Theorem 6.1.7, in what follows, we apply the results of the last section to some of the variational inequalities that we have considered throughout this work. In view of the historical background, it makes sense to start with:

Corollary 6.1.11 (Strong Stationarity for the Classical Obstacle Problem)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a non-empty, open, bounded set and let $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ be defined as before. Assume that two Borel measurable functions $\psi_1, \psi_2 : \Omega \to [-\infty, \infty]$ are given such that*

$$K := \Big\{ v \in H_0^1(\Omega) \;\Big|\; \psi_1 \leq v \leq \psi_2 \ \mathcal{L}^d\text{-a.e. in } \Omega \Big\} \neq \emptyset,$$

and let $S : L^2(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, denote the restriction of the solution operator of the classical obstacle problem

$$w \in K, \qquad \langle -\Delta w, v - w\rangle \geq \langle f, v - w\rangle \qquad \forall v \in K \tag{6.13}$$

to $L^2(\Omega)$. Suppose that $\mathcal{J} : H_0^1(\Omega) \times L^2(\Omega) \to \mathbb{R}$ is a Fréchet differentiable function and that $\mathcal{F}_{ad}$ is a non-empty, closed and convex subset of $L^2(\Omega)$. Then, an $\bar{f} \in \mathcal{F}_{ad}$ with $\mathrm{cl}_{L^2}(\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})) = L^2(\Omega)$ is Bouligand stationary (in the sense of Proposition 6.1.2) for the optimal control problem

$$\begin{aligned} &\min\ \mathcal{J}(w, f) \\ &\text{s.t. } w = S(f), \quad f \in \mathcal{F}_{ad}, \end{aligned}$$

if and only if there exist an adjoint state $\bar{p} \in H_0^1(\Omega)$ and a multiplier $\bar{\eta} \in H^{-1}(\Omega)$ such that $\bar{f}$ and its state $\bar{w} := S(\bar{f})$ satisfy

$$\begin{aligned} -\Delta\bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\ \bar{p} \in \mathcal{T}_K(\bar{w}) \cap \ker(\bar{f} + \Delta\bar{w}), \qquad \langle \bar{\eta}, u\rangle &\geq 0 \quad \forall u \in \mathcal{T}_K(\bar{w}) \cap \ker(\bar{f} + \Delta\bar{w}). \end{aligned} \tag{6.14}$$

Proof. From Corollary 3.4.3 and the embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$, we obtain that the solution operator $S : L^2(\Omega) \to H^1_0(\Omega)$ to (6.13) is Hadamard directionally differentiable in all $f \in L^2(\Omega)$ in all directions $g \in L^2(\Omega)$, and that the directional derivative $\delta := S'(\bar{f}; g)$ in a point $\bar{f} \in L^2(\Omega)$ with state $\bar{w} := S(\bar{f})$ in a direction $g \in L^2(\Omega)$ is uniquely characterized by the variational inequality

$$\delta \in \mathcal{T}_K(\bar{w}) \cap \ker(\bar{f} + \Delta\bar{w}), \qquad \langle -\Delta\delta, z - \delta\rangle \geq \langle g, z - \delta\rangle \qquad \forall z \in \mathcal{T}_K(\bar{w}) \cap \ker(\bar{f} + \Delta\bar{w}).$$

The claim is now a straightforward consequence of Corollary 6.1.10 with $H = L^2(\Omega)$, $V = U = H^1_0(\Omega)$, $V^* = H^{-1}(\Omega)$, $L = \mathcal{T}_K(\bar{w}) \cap \ker(\bar{f} + \Delta\bar{w})$, $Q = \chi_L$ and $A = -\Delta \in L(H^1_0(\Omega), H^{-1}(\Omega))$. □

Note that (6.14) corresponds precisely to the optimality system obtained in [Mignot and Puel, 1984, Theorem 2.2]. This proves that Theorem 6.1.7 and Corollary 6.1.10, respectively, indeed extend the classical results of Mignot and Puel.

As a second example, we consider the non-smooth partial differential equation (4.2) that we have studied in Section 4.1. For the sake of simplicity, we confine our analysis to the case $\alpha = 0$, $\beta = 1$.

Corollary 6.1.12 (Strong Stationarity for a Non-Smooth Semilinear PDE)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be a bounded domain and let $H^1_0(\Omega)$ and $H^{-1}(\Omega)$ be defined as before. Let $S : L^2(\Omega) \to H^1_0(\Omega)$, $f \mapsto w$, denote the restriction of the solution operator of the PDE*

$$w \in H^1_0(\Omega), \qquad -\Delta w + \max(0, w) = f \in H^{-1}(\Omega) \tag{6.15}$$

to $L^2(\Omega)$. Suppose that $\mathcal{J} : H^1_0(\Omega) \times L^2(\Omega) \to \mathbb{R}$ is a Fréchet differentiable function and that $\mathcal{F}_{ad}$ is a non-empty, closed and convex subset of $L^2(\Omega)$. Then, an $\bar{f} \in \mathcal{F}_{ad}$ with $\mathrm{cl}_{L^2}(\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})) = L^2(\Omega)$ is Bouligand stationary (in the sense of Proposition 6.1.2) for the optimal control problem

$$\begin{aligned} &\min\ \mathcal{J}(w, f) \\ &\text{s.t. } w = S(f), \quad f \in \mathcal{F}_{ad}, \end{aligned}$$

if and only if there exist an adjoint state $\bar{p} \in H^1_0(\Omega)$ and a function $\bar{\sigma} \in L^\infty(\Omega)$ such that $\bar{f}$ and its state $\bar{w} := S(\bar{f})$ satisfy

$$\begin{aligned} -\Delta\bar{p} + \bar{\sigma}\bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}), \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\ \bar{\sigma} \in \partial \max(0, \cdot)(\bar{w})\ \ &\mathcal{L}^d\text{-a.e. in } \Omega, \\ \bar{p} \leq 0\ \ &\mathcal{L}^d\text{-a.e. in } \{\bar{w} = 0\}. \end{aligned} \tag{6.16}$$

Proof. From the analysis in Section 4.1, it follows straightforwardly that the map $S : L^2(\Omega) \to H^1_0(\Omega)$ is directionally differentiable, and that the directional derivatives $\delta = S'(f; g) \in H^1_0(\Omega)$, $f, g \in L^2(\Omega)$, are uniquely characterized by the variational inequalities

$$\begin{aligned} &\langle -\Delta\delta, z - \delta\rangle \\ &\quad + \frac{1}{2}\int_\Omega \left(\mathbb{1}_{\{w=0\}} \max(0, z)^2 + \mathbb{1}_{\{w>0\}} z^2\right) \mathrm{d}\mathcal{L}^d - \frac{1}{2}\int_\Omega \left(\mathbb{1}_{\{w=0\}} \max(0, \delta)^2 + \mathbb{1}_{\{w>0\}} \delta^2\right) \mathrm{d}\mathcal{L}^d \\ &\qquad\qquad \geq \langle g, z - \delta\rangle \qquad \forall z \in H^1_0(\Omega), \end{aligned}$$

where $w := S(f)$ again denotes the state associated with f. If we use the above in Corollary 6.1.9 with $H = L^2(\Omega)$, $V = U = H^1_0(\Omega)$, $V^* = H^{-1}(\Omega)$, $A = -\Delta \in L(H^1_0(\Omega), H^{-1}(\Omega))$ and

$$Q(z) = \int_\Omega \left(\mathbb{1}_{\{\bar{w}=0\}} \max(0, z)^2 + \mathbb{1}_{\{\bar{w}>0\}} z^2\right) \mathrm{d}\mathcal{L}^d,$$

then we obtain that, for every $\bar{f} \in \mathcal{F}_{ad}$ with $\mathrm{cl}_{L^2}(\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})) = L^2(\Omega)$, Bouligand stationarity is equivalent to the existence of a $\bar{p} \in H_0^1(\Omega)$ and an $\bar{\eta} \in H^{-1}(\Omega)$ with

$$\begin{aligned} -\Delta\bar{p} &= \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) &= 0, \\ \langle \bar{\eta}, u \rangle \geq \int_\Omega \mathbb{1}_{\{\bar{w}=0\}} \max(0, u)\bar{p} + \mathbb{1}_{\{\bar{w}>0\}} u\bar{p} \,\mathrm{d}\mathcal{L}^d \qquad &\forall u \in H_0^1(\Omega). \end{aligned} \tag{6.17}$$

Note that the theorem of Hahn-Banach and the density of $H_0^1(\Omega)$ in $L^2(\Omega)$ imply that every tuple $(\bar{p}, \bar{\eta})$ with (6.17) satisfies $\bar{\eta} \in L^2(\Omega)$,

$$0 \geq \int_\Omega \mathbb{1}_{\{\bar{w}=0\}} \max(0, u)\bar{p} + \mathbb{1}_{\{\bar{w}>0\}} u\bar{p} - \bar{\eta}u \,\mathrm{d}\mathcal{L}^d \qquad \forall u \in L^2(\Omega)$$

and, as a consequence,

$$\bar{\eta} = 0 \ \ \mathcal{L}^d\text{-a.e. in } \{\bar{w} < 0\}, \qquad \bar{\eta} = \bar{p} \ \ \mathcal{L}^d\text{-a.e. in } \{\bar{w} > 0\}, \qquad \bar{p} \leq \bar{\eta} \leq 0 \ \ \mathcal{L}^d\text{-a.e. in } \{\bar{w} = 0\}.$$

By defining

$$\bar{\sigma} := \frac{\bar{\eta}}{\bar{p}} \mathbb{1}_{\{\bar{p} \neq 0\}},$$

we now obtain an L^∞-function with

$$\bar{\eta} = \bar{\sigma}\bar{p} \ \ \mathcal{L}^d\text{-a.e. in } \Omega, \qquad \bar{\sigma} \in \partial \max(0, \cdot)(\bar{w}) \ \ \mathcal{L}^d\text{-a.e. in } \Omega,$$

and

$$-\Delta\bar{p} + \bar{\sigma}\bar{p} = \partial_w \mathcal{J}(\bar{w}, \bar{f}), \qquad \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) = 0.$$

This proves that (6.17) implies (6.16). If, conversely, we begin with the system (6.16), then we may use exactly the same steps as above backwards to obtain that $\bar{p}$ and $\bar{\eta} := \bar{\sigma}\bar{p}$ satisfy (6.17). The claim now follows immediately. □

We point out that (6.16) is precisely the strong stationarity system studied in [Christof et al., 2017, Theorem 4.12, Proposition 4.13]. This demonstrates that Theorem 6.1.7 can also reproduce those results that have been obtained for non-smooth elliptic partial differential equations in the literature. Next, let us consider the EVI of static elastoplasticity (cf. Corollary 4.3.5):

Corollary 6.1.13 (Strong Stationarity for Static Elastoplasticity in Primal Formulation)**.** *Assume that a bounded Lipschitz domain $\Omega \subset \mathbb{R}^3$ and a relatively open and non-empty set $\Gamma_D \subset \partial\Omega$ are given. Define*

$$\begin{aligned} H_D^1(\Omega, \mathbb{R}^3) &:= \left\{ z \in H^1(\Omega)^3 \;\middle|\; \mathrm{tr}(z) = 0 \ \mathcal{H}^2\text{-a.e. on } \Gamma_D \right\}, \\ \mathbb{R}^{3\times3}_{dev} &:= \left\{ q \in \mathbb{R}^{3\times3}_{sym} \;\middle|\; q_{11} + q_{22} + q_{33} = 0 \right\}, \\ V &:= H_D^1(\Omega, \mathbb{R}^3) \times L^2\left(\Omega, \mathbb{R}^{3\times3}_{dev}\right), \end{aligned}$$

and suppose that

$$\begin{gathered} \|\cdot\|_F : \mathbb{R}^{3\times3} \to \mathbb{R}, \qquad a : V \times V \to \mathbb{R}, \qquad P : V \to L^2(\Omega, \mathbb{R}^{3\times3}_{dev}), \qquad j : V \to \mathbb{R}, \\ \mathbb{C} \in L^\infty(\Omega, L(\mathbb{R}^{3\times3}_{sym}, \mathbb{R}^{3\times3}_{sym})), \qquad \mathbb{H} \in L^\infty(\Omega, L(\mathbb{R}^{3\times3}_{dev}, \mathbb{R}^{3\times3}_{dev})) \end{gathered}$$

and σ_0 satisfy the conditions in Assumption 4.3.4. Identify $L^2(\Omega, (\mathbb{R}^{3\times3}_{dev})^*)$ with $L^2(\Omega, \mathbb{R}^{3\times3}_{dev})$, denote with $S : V^* \to V$, $f \mapsto w$, the solution operator of the EVI

$$w \in V, \qquad a(w, v - w) + j(v) - j(w) \geq \langle f, v - w \rangle_V \qquad \forall v \in V,$$

and consider an optimal control problem of the type

$$\begin{aligned} &\min \mathcal{J}(w,f) \\ &\text{s.t. } w = S(f), \quad f \in \mathcal{F}_{ad}, \end{aligned} \tag{6.18}$$

with a Fréchet differentiable objective function $\mathcal{J} : V \times \big(L^2(\Omega,\mathbb{R}^3) \times L^2(\Omega,\mathbb{R}^{3\times 3}_{dev})\big) \to \mathbb{R}$ *and a non-empty, closed and convex admissible set* $\mathcal{F}_{ad} \subset L^2(\Omega,\mathbb{R}^3) \times L^2(\Omega,\mathbb{R}^{3\times 3}_{dev})$*. Then, for every* $f \in V^*$ *with associated solution* $w = S(f) \in V$ *there exists a unique multiplier* $\lambda \in L^2(\Omega,\mathbb{R}^{3\times 3}_{dev})$ *with*

$$\lambda \in \partial\|\cdot\|_F(Pw) \;\; \mathcal{L}^3\text{-a.e. in } \Omega \quad \text{and} \quad \sigma_0 P^*\lambda = \varphi \;\text{ for }\; \varphi(\cdot) := f(\cdot) - a(w,\cdot) \in V^*,$$

and, for every $\bar f \in \mathcal{F}_{ad}$ *with* $\mathrm{cl}_{L^2}(\mathbb{R}^+(\mathcal{F}_{ad} - \bar f)) = L^2(\Omega,\mathbb{R}^3) \times L^2(\Omega,\mathbb{R}^{3\times 3}_{dev})$*, Bouligand stationarity for* (6.18) *is equivalent to the existence of a* $\bar p$ *and an* $\bar\eta$ *with*

$$\begin{aligned} \bar p \in U, \quad \bar\eta \in U^*, \\ A^*\bar p = \partial_w \mathcal{J}(\bar w, \bar f) - \bar\eta, \\ \bar p + \partial_f \mathcal{J}(\bar w, \bar f) = 0, \\ \bar p \in L, \quad \langle \bar\eta, u\rangle_U \geq 0 \quad \forall u \in L. \end{aligned} \tag{6.19}$$

Here, $\bar w := S(\bar f)$ *is the state associated with* $\bar f$*,* U *is the Hilbert space*

$$U := H^1_D(\Omega,\mathbb{R}^3) \times \left\{ q \in L^2(\Omega,\mathbb{R}^{3\times 3}_{dev}) \;\middle|\; \int_{\{P\bar w \neq 0\}} \frac{\|P\bar w\|_F^2 \|q\|_F^2 - (P\bar w : q)^2}{\|P\bar w\|_F^3} \,\mathrm{d}\mathcal{L}^3 < \infty \right\}$$

endowed with the norm

$$\|u\|_U := \left(\|u\|_V^2 + \int_{\{P\bar w \neq 0\}} \frac{\|P\bar w\|_F^2 \|Pu\|_F^2 - (P\bar w : Pu)^2}{\|P\bar w\|_F^3} \,\mathrm{d}\mathcal{L}^3 \right)^{1/2},$$

$A \in L(U,U^*)$ *is the operator defined by*

$$\langle Au_1, u_2\rangle_U = a(u_1,u_2) + \sigma_0 \int_{\{P\bar w \neq 0\}} \frac{\|P\bar w\|_F^2 \big(Pu_1 : Pu_2\big) - (P\bar w : Pu_1)\big(P\bar w : Pu_2\big)}{\|P\bar w\|_F^3} \,\mathrm{d}\mathcal{L}^3$$

for all $u_1, u_2 \in U$*, and* L *is the set*

$$L := \big\{ u \in U \;\big|\; \bar\lambda : Pu = \|Pu\|_F \text{ a.e. in } \{P\bar w = 0\} \big\},$$

where $\bar\lambda$ *is the multiplier associated with the subgradient* $\bar\varphi(\cdot) := \bar f(\cdot) - a(\bar w,\cdot) \in V^*$*.*

Proof. The existence of a unique multiplier λ for all f follows straightforwardly from Corollary 4.3.5. It remains to prove the assertion concerning the Bouligand stationarity. To this end, we note that the map S is directionally differentiable in every $\bar f \in \mathcal{F}_{ad}$ by Corollary 4.3.5, and that, according to (4.30), the directional derivatives $\delta := S'(\bar f; g)$, $g \in V^*$, are uniquely characterized by the EVI

$$\delta \in L, \quad \langle A\delta, z - \delta\rangle_U \geq \langle g, z - \delta\rangle_V \quad \forall z \in L.$$

Using Corollary 6.1.10 with $H := L^2(\Omega,\mathbb{R}^3) \times L^2(\Omega,\mathbb{R}^{3\times 3}_{dev})$ and U, V, L as defined above, the claim now follows immediately. □

Observe that (6.19) does not involve the operator $V \ni v \mapsto a(v,\cdot) \in V^*$ but the modified map A, and that the identities and variational inequalities in (6.19) are identities and variational inequalities in the space U and not in the original space V (in contrast to (6.14) and (6.16)). Similar effects also occur when Mosolov's problem or the PDE in Section 4.3.5 are considered (see also Corollary 6.1.14 below). We remark that, in the dual setting, strong stationarity results for the problem of static elastoplasticity have already been obtained in [Herzog et al., 2013].

As a final example, we study the H^1_0-elliptic variational inequality from Section 5.2:

Corollary 6.1.14 (Strong Stationarity for an H_0^1-elliptic Variational Inequality of the Second Kind)**.** *Let $\Omega \subset \mathbb{R}^d$, $d \geq 2$, be a bounded Lipschitz domain and let $S : L^2(\Omega) \to H_0^1(\Omega)$, $f \mapsto w$, denote the restriction of the solution operator of the variational inequality*

$$w \in H_0^1(\Omega), \quad \int_\Omega \nabla w \cdot \nabla(v-w)\mathrm{d}\mathcal{L}^d + \int_\Omega |v|\mathrm{d}\mathcal{L}^d - \int_\Omega |w|\mathrm{d}\mathcal{L}^d \geq \langle f, v-w\rangle \quad \forall v \in H_0^1(\Omega) \tag{6.20}$$

to $L^2(\Omega)$. Suppose that $\mathcal{J} : H_0^1(\Omega) \times L^2(\Omega) \to \mathbb{R}$ is a Fréchet differentiable function, that $\mathcal{F}_{ad}$ is a non-empty, closed and convex subset of $L^2(\Omega)$, and that a right-hand side $\bar{f} \in \mathcal{F}_{ad}$ is given such that $\mathrm{cl}_{L^2}(\mathbb{R}^+(\mathcal{F}_{ad} - \bar{f})) = L^2(\Omega)$ holds and such that the state $\bar{w} := S(\bar{f})$ satisfies the conditions in Assumption 5.2.4. Then, $\bar{f}$ is Bouligand stationary for the optimal control problem

$$\begin{aligned} &\min\ \mathcal{J}(w,f) \\ &\text{s.t. } w = S(f), \quad f \in \mathcal{F}_{ad}, \end{aligned}$$

if and only if there exist a $\bar{p}$ and an $\bar{\eta}$ with

$$\begin{aligned} \bar{p} \in U, \quad \bar{\eta} \in U^*, \\ A^*\bar{p} = \partial_w \mathcal{J}(\bar{w}, \bar{f}) - \bar{\eta}, \\ \bar{p} + \partial_f \mathcal{J}(\bar{w}, \bar{f}) = 0, \\ \bar{p} \in L, \quad \langle \bar{\eta}, u\rangle_U \geq 0 \quad \forall u \in L. \end{aligned} \tag{6.21}$$

Here, U is the Hilbert space

$$U := \left\{ u \in H_0^1(\Omega) \;\middle|\; \int_{\{\bar{w}=0\}\cap\{\nabla\bar{w}\neq 0\}} \frac{\mathrm{tr}(u)^2}{\|\nabla\bar{w}\|_2}\mathrm{d}\mathcal{H}^{d-1} < \infty \right\}$$

endowed with the norm

$$\|u\|_U := \left(\|u\|_{H^1}^2 + \int_{\{\bar{w}=0\}\cap\{\nabla\bar{w}\neq 0\}} \frac{\mathrm{tr}(u)^2}{\|\nabla\bar{w}\|_2}\mathrm{d}\mathcal{H}^{d-1} \right)^{1/2},$$

$A \in L(U, U^*)$ is the operator defined by

$$\langle Au_1, u_2\rangle_U = \langle -\Delta u_1, u_2\rangle_{H_0^1} + 2\int_{\{\bar{w}=0\}\cap\{\nabla\bar{w}\neq 0\}} \frac{\mathrm{tr}(u_1)\,\mathrm{tr}(u_2)}{\|\nabla\bar{w}\|_2}\mathrm{d}\mathcal{H}^{d-1}$$

for all $u_1, u_2 \in U$, L is the set

$$\begin{aligned} L := \Big\{ u \in U \;\Big|\; \mathrm{tr}(u)^- = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_+,\ \mathrm{tr}(u)^+ = 0\ \mathcal{H}^{d-1}\text{-a.e. on } \mathcal{N}_-, \\ |u| = \bar{\lambda} u\ \mathcal{L}^d\text{-a.e. in } \{\bar{w}=0\} \Big\}, \end{aligned}$$

$\bar{\lambda} \in L^2(\Omega)$ is the multiplier associated with the subgradient $\bar{\varphi} := \Delta\bar{w} + \bar{f}$ as in (5.39)*, and $\mathcal{N}_+$, $\mathcal{N}_-$ are defined as in Assumption 5.2.4.*

Proof. The proof is completely along the same lines as that of Corollary 6.1.13. □

We point out that Corollary 6.1.14 generalizes [De los Reyes and Meyer, 2016, Theorem 5.3]. Note that, in this paper, the authors to do not observe the emergence of the surface integral and the change in function space that appears in (6.21) due to their more restrictive assumptions on the state $\bar{w}$, cf. point (ii) in Remark 5.2.16.

Let us conclude this section with some general remarks:

Remark 6.1.15.

(i) The optimality systems in Corollaries 6.1.11 to 6.1.14 are typically overdetermined and can, in general, not be solved directly by numerical algorithms. However, for particular problems, strong stationarity conditions may still serve as a point of departure for the development of practical solution techniques. In the case of the PDE in Corollary 6.1.12, for example, one can simply drop the inequality $\bar{p} \leq 0$ $\mathcal{L}^d$-a.e. in $\{\bar{w} = 0\}$ in (6.16) *and solve the resulting well-posed system with a semismooth Newton method, see [Christof et al., 2017, Section 5]. (Note that the system that is obtained in this way is still necessary for Bouligand stationarity.)*

(ii) Strong stationarity systems are often well-suited to assess the strength of optimality conditions that are obtained by regularization. If we consider, e.g., again the PDE (6.15)*, then an approximation approach yields precisely the system* (6.16) *without the condition $\bar{p} \leq 0$ $\mathcal{L}^d$-a.e. in $\{\bar{w} = 0\}$, see [Christof et al., 2017, Section 4.1]. In this case, the mollification of the non-smooth terms thus causes a loss of information about the sign of the adjoint state $\bar{p}$. Note that, e.g., for the EVI* (6.20) *a regularization procedure is necessarily more "destructive" because it can never reproduce the change in the bilinear form that is observed in* (6.21) *and that is directly linked to the non-smoothness of the involved L^1-norm.*

6.2 Remarks on Second-Order Conditions and Other Fields

Before we close this chapter, we would like to give some additional comments on alternative applications of our differentiability results and the relationship between the analysis of Chapter 1 and other branches of optimization and optimal control.

Let us first point out that, in the situation of Section 6.1, the theorems of Chapters 1 to 5 can also be used to construct a first-order model of the reduced objective $f \mapsto \mathcal{J}(S(f), f)$. Such models are, e.g., essential for the design of trust-region-type algorithms for non-smooth optimal control problems of the form (O). For details on this topic, we refer to [Christof et al., 2018; Qi and Sun, 1994] and the references therein.

Further, it should be noted that the sensitivity analysis of elliptic variational inequalities is not the only field where the functional $Q_j^{v,\varphi} : V \to [0, \infty]$ and the notion of second-order epi-differentiability are relevant. If we consider, e.g., constrained optimization problems, then we may prove:

Theorem 6.2.1. *Suppose that a Hilbert space V, a twice Fréchet differentiable function $\mathcal{J} : V \to \mathbb{R}$ and a closed, convex non-empty set $K \subset V$ are given, and consider the optimization problem*

$$\min \ \mathcal{J}(v), \qquad \text{s.t. } v \in K.$$

Assume that the map $V \ni z \mapsto \mathcal{J}''(v)z^2 \in \mathbb{R}$ is a Legendre form for all $v \in K$ (in the sense of [Bonnans and Shapiro, 2000, Definition 3.73]) and that the characteristic function $\chi_K : V \to [0, \infty]$ is twice epi-differentiable in all $v \in K$ for all $\varphi \in \mathcal{N}_K(v)$. Then, for every $\bar{v} \in K$, it holds

$$-\mathcal{J}'(\bar{v}) \in \mathcal{N}_K(\bar{v}), \qquad Q_K^{\bar{v},-\mathcal{J}'(\bar{v})}(z) + \mathcal{J}''(\bar{v})z^2 > 0 \qquad \forall z \in V \setminus \{0\}$$

if and only if there exist constants $c > 0$ and $r > 0$ with

$$\mathcal{J}(v) \geq \mathcal{J}(\bar{v}) + \frac{c}{2}\|v - \bar{v}\|_V^2 \qquad \forall v \in K \cap B_r(\bar{v}).$$

Here, $B_r(\bar{v})$ is the closed ball in V of radius r around $\bar{v}$ and $Q_K^{v,\varphi}$ is again short for $Q_{\chi_K}^{v,\varphi}$.

Proof. The claim follows directly from [Christof and Wachsmuth, 2017b, Theorem 4.5, Lemma 5.1]. □

As the above result shows, the second subderivative is also the "right" substitute for the second Fréchet derivative when necessary and sufficient second-order optimality conditions for functionals of the form $\mathcal{J} + \chi_K$ are studied. We remark that similar effects can also be observed in other applications. See, e.g., [Christof and Wachsmuth, 2017b; Rockafellar and Wets, 1998] for details.

Concluding Remarks

In view of the results of Chapters 1 to 5, we can conclude that the notion of second-order epi-differentiability is the fundamental concept in the sensitivity analysis of elliptic variational inequalities of the first and the second kind. It allows to derive a criterion for the Hadamard directional differentiability of the solution map - namely Theorem 1.4.1 - that is sharp, that only requires minimal assumptions on the appearing operators and functionals, and that covers the classical results of Mignot, Haraux, Bonnans and Shapiro in a natural way. As we have seen in Lemma 1.3.13, the property of polyhedricity introduced in [Haraux, 1977; Mignot, 1976] is moreover just a special instance of a more general sufficient condition for second-order epi-differentiability that may also be used when functions with non-zero curvature are considered, cf. Section 4.3.

From the results of Chapter 5, we may deduce further that, while relatively easy for non-smooth functionals on Bochner-Lebesgue spaces, checking the condition of second-order epi-differentiability is typically rather hard as soon as weak derivatives and Sobolev spaces are involved. Compare, e.g., with the results for the EVI of static elastoplasticity in Corollaries 4.3.5 and 4.3.6 in this context, where the directional differentiability of the solution map could be obtained without any additional assumptions, and with the rather cumbersome analysis in Sections 5.1 and 5.2. As already mentioned, the basic problem that arises when non-differentiable terms on, e.g., the spaces $H^m(\Omega)$, $m \geq 1$, are considered is that distributional curvature effects come into play that are very difficult to analyze properly. This becomes apparent, e.g., in Theorems 3.4.7, 5.2.14 and 5.2.15 where the chain rule in Theorem 2.4.8 and the concept of polyhedricity fail for fairly non-trivial reasons.

Given the results of Section 6.1 and Chapter 1, we may come to the conclusion that the concept of strong stationarity is very well suited for the analysis of optimal control problems that are governed by elliptic variational inequalities of the first or the second kind. As Proposition 1.3.5 shows, the directional derivatives $\delta := S'(f;g)$ of the solution operator S to an EVI of the type (P) are, if existent, always characterized by an auxiliary problem of the form

$$\delta \in V, \quad \langle A\delta, z-\delta\rangle + \frac{1}{2}Q(z) - \frac{1}{2}Q(\delta) \geq \langle g, z-\delta\rangle \quad \forall z \in V$$

and this is precisely what is needed in Theorem 6.1.7. Note that, by combining Theorems 1.4.1 and 6.1.7, we obtain in particular that, if we consider an optimal control problem of the type (O) which satisfies $\mathcal{F}_{ad} = H$ and which is governed by an EVI with a directionally differentiable solution operator S, then every Bouligand stationary point $\bar{f}$ has to satisfy a strong stationarity condition of the form (6.7) (since Assumption 6.1.3 is automatically satisfied). In this case, we thus obtain at a multiplier system without any additional assumptions.

Lastly, we would like to point out that differentiability results similar to that in Proposition 1.3.10 can also be proved when perturbations of the operator A and the functional j in (P) are considered. See, e.g., [Christof and Wachsmuth, 2017a, Theorem 4.1], [Christof and Meyer, 2016, Theorem 4.14] and [Adly and Bourdin, 2017, Theorem 41] for some examples. However, the sensitivity analysis for such perturbations is far from straightforward and, at least to the author's best knowledge, it is currently completely unclear if, e.g., an equivalence as in Theorem 1.4.1 can be obtained for problems of this type. The same holds true for the extension of our differentiability results to evolution variational inequalities as considered, e.g., in [Jarušek et al., 2003]. Further research is necessary here.

Bibliography

Adams, R. A. (1975). *Sobolev Spaces*. Academic Press, New York.

Adly, S. and Bourdin, L. (2017). Sensitivity analysis of variational inequalities via twice epi-differentiability and proto-differentiability of the proximity operator. Preprint, arXiv:1707.08512.

Alexandrov, A. D. (1939). Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected to it. *Leningrad State Univ. Ann. (Uchenye Zapiski) Math. Ser.*, (6):3–35.

Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford University Press, Oxford/New York.

Ambrosio, L., Pinamonti, A., and Speight, G. (2014). Weighted Sobolev spaces on metric measure spaces. Preprint, arXiv:1406.3000.

Aronszajn, N. (1976). Differentiability of Lipschitzian mappings between Banach spaces. *Studia Math.*, 57(2):147–190.

Attouch, H. (1984). *Variational Convergence for Functions and Operators*. Pitman, Boston, MA.

Attouch, H., Buttazzo, G., and Michaille, G. (2006). *Variational Analysis in Sobolev and BV Spaces*. SIAM, Philadelphia.

Beardon, A. F. (1991). *Iteration of Rational Functions: Complex Analytic Dynamical Systems*, volume 132 of *Graduate Texts in Mathematics*. Springer Verlag.

Betz, T. (2015). *Optimal control of two variational inequalities arising in solid mechanics*. PhD thesis, Technische Universität Dortmund.

Beurling, A. and Deny, J. (1959). Dirichlet spaces. *Proc. Natl. Acad. Sci. USA*, 45(2):208–215.

Bonnans, J. F., Cominetti, R., and Shapiro, A. (1998). Sensitivity analysis of optimization problems under second order regular constraints. *Math. Oper. Res.*, 23(4):806–831.

Bonnans, J. F. and Shapiro, A. (2000). *Perturbation Analysis of Optimization Problems*. Springer Verlag, New York.

Borwein, J. M. and Vanderwerff, J. D. (2010). *Convex Functions: Constructions, Characterizations and Counterexamples*. Cambridge University Press, Cambridge.

Borwein, J. M. and Zhu, Q. J. (2005). *Techniques of Variational Analysis*. Springer Verlag, Berlin/Heidelberg/New York.

Borwein, J. W. and Noll, D. (1994). Second order differentiability of convex functions in Banach spaces. *Trans. Amer. Math. Soc.*, 342(1):43–81.

Brezis, H. (1971). Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations. *Contrib. Nonlinear Funct. Anal.*, pages 101–156.

Calatroni, L., Chung, C., De Los Reyes, J. C., Schönlieb, C. B., and Valkonen, T. (2015). Bilevel approaches for learning of variational imaging models. Preprint, arXiv:1505.02120.

Carstensen, C., Reddy, B. D., and Schedensack, M. (2016). A natural nonconforming FEM for the Bingham flow problem is quasi-optimal. *Numer. Math.*, 133(1):37–66.

Casas, E., Herzog, R., and Wachsmuth, G. (2012). Optimality conditions and error analysis of semilinear elliptic control problems with L^1-cost functional. *SIAM J. Optim.*, 22(3):795–820.

Chambolle, A. and Darbon, J. (2009). On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vis.*, 84:288–307.

Chan, T. F. and Shen, J. (2005). *Image Processing and Analysis*. SIAM, Philadelphia, PA.

Christof, C. (2017). L^∞-error estimates for the obstacle problem revisited. *Calcolo*, 54(4):1243–1264.

Christof, C., Clason, C., Meyer, C., and Walther, S. (2017). Optimal control of a non-smooth semilinear elliptic equation. Preprint, arXiv:1705.00939.

Christof, C., De Los Reyes, J. C., and Meyer, C. (2018). A non-smooth trust-region method for locally Lipschitz functions with application to optimization problems constrained by variational inequalities. Preprint, arXiv:1711.03208v2.

Christof, C. and Meyer, C. (2016). Sensitivity analysis for a class of H_0^1-elliptic variational inequalities of the second kind (aka. Differentiability properties of the solution operator to an elliptic variational inequality of the second kind). Preprint, SPP1962-012.

Christof, C. and Wachsmuth, G. (2017a). Differential sensitivity analysis of variational inequalities with locally Lipschitz continuous solution operators. Preprint, arXiv:1711.02720.

Christof, C. and Wachsmuth, G. (2017b). No-gap second-order conditions via a directional curvature functional. Preprint, arXiv:1707.07579.

Christof, C. and Wachsmuth, G. (2017c). On the non-polyhedricity of sets with upper and lower bounds in dual spaces. Preprint, arXiv:1711.02588.

Cioranescu, D., Girault, V., and Rajagopal, K. R. (2016). *Mechanics and Mathematics of Fluids of the Differential Type*, volume 35 of *Advances in Mechanics and Mathematics*. Springer Verlag.

Cottle, R. W., Pang, J. S., and Stone, R. E. (2009). *The Linear Complementarity Problem*, volume 60 of *Classics in Applied Mathematics*. SIAM.

De los Reyes, J. C., Herzog, R., and Meyer, C. (2016). Optimal control of static elastoplasticity in primal formulation. *SIAM J. Control Optim.*, 54(6):3016–3039.

De los Reyes, J. C. and Meyer, C. (2016). Strong stationarity conditions for a class of optimization problems governed by variational inequalities of the second kind. *J. Optim. Theory Appl.*, 168(2):375–409.

Dean, E. J., Glowinski, R., and Guidoboni, G. (2007). On the numerical simulation of Bingham viscoplastic flow: Old and new results. *J. Non-Newton. Fluid Mech.*, 142:36–62.

Ding, Z. (1996). A proof of the trace theorem of Sobolev spaces on Lipschitz domains. *Proc. Amer. Math. Soc.*, 124(2):591–600.

Do, C. N. (1992). Generalized second-order derivatives of convex functions in reflexive Banach spaces. *Trans. Amer. Math. Soc.*, 334(1):281–301.

Drábek, P. and Milota, J. (2007). *Methods of Nonlinear Analysis: Applications to Differential Equations*. Birkhäuser Verlag, Basel/Boston/Berlin.

Dudley, R. M. and Norvaiša, R. (2010). *Concrete Functional Calculus*. Springer Monographs in Mathematics. Springer Verlag.

Edmunds, D. E. and Evans, W. D. (1987). *Spectral Theory and Differential Operators*. Oxford University Press, Oxford/New York.

Egert, M., Haller-Dintelmann, R., and Rehberg, J. (2015). Hardy's inequality for functions vanishing on a part of the boundary. *Potential Anal.*, 43(1):49–78.

Ekeland, I. and Temam, R. (1976). *Convex Analysis and Variational Problems*. North-Holland Publishing Company.

Evans, L. C. (2010). *Partial Differential Equations*. AMS, Providence, RI, second edition.

Evans, L. C. and Gariepy, R. F. (2015). *Measure Theory and Fine Properties of Functions*. CRC Press, Boca Raton, FL, revised edition.

Fitzpatrick, S. and Phelps, R. R. (1982). Differentiability of the metric projection in Hilbert space. *Trans. Amer. Math. Soc.*, 270(2):483–501.

Foote, R. L. (1984). Regularity of the distance function. *Proc. Amer. Math. Soc.*, 92(1):153–155.

Fuchs, M. and Seregin, G. (1998). Regularity results for the quasi–static Bingham variational inequality in dimensions two and three. *Math. Z.*, 227(3):525–541.

Fuchs, M. and Seregin, G. (2000). *Variational Methods for Problems from Plasticity Theory and for Generalized Newtonian Fluids*. Springer Verlag, Berlin/Heidelberg/New York.

Fukushima, M., Oshima, Y., and Takeda, M. (2011). *Dirichlet Forms and Symmetric Markov Processes*, volume 19 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, second edition.

Gilbarg, D. and Trudinger, N. S. (2001). *Elliptic Partial Differential Equations of Second Order*. Springer Verlag, Berlin/Heidelberg/New York, reprint of the 1998 edition.

Girault, V. and Raviart, P. A. (1986). *Finite Element Methods for Navier-Stokes Equations: Theory and Algorithms*, volume 5 of *Springer Series in Computational Mathematics*. Springer Verlag.

Glowinski, R. (1980). *Lectures on Numerical Methods for Non-Linear Variational Problems*. Tata Institute of Fundamental Research, Bombay.

Glowinski, R. (2015). *Variational Methods for the Numerical Solution of Nonlinear Elliptic Problems*. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM.

Grisvard, P. (1985). *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston.

Hackbusch, W. (2017). *Elliptic Differential Equations: Theory and Numerical Treatment*, volume 18 of *Springer Series in Computational Mathematics*. Springer Verlag, Berlin/Heidelberg, second edition.

Han, W. and Reddy, B. D. (1999). *Plasticity: Mathematical Theory and Numerical Analysis*. Springer Verlag, New York.

Haraux, A. (1977). How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities. *J. Math. Soc. Japan*, 29(4):615–631.

Heinonen, J., Koskela, P., Shanmugalingam, N., and Tyson, J. T. (2015). *Sobolev Spaces on Metric Measure Spaces: An Approach Based on Upper Gradients*, volume 27 of *New Mathematical Monographs*. Cambridge University Press.

Herzog, R., Meyer, C., and Wachsmuth, G. (2013). B- and strong stationarity for optimal control of static plasticity with hardening. *SIAM J. Optim.*, 23(1):321–352.

Hintermüller, M. and Surowiec, T. M. (2011). First-order optimality conditions for elliptic mathematical programs with equilibrium constraints via variational analysis. *SIAM J. Optim.*, 21(4):1561–1593.

Hintermüller, M. and Surowiec, T. M. (2017). On the directional differentiability of the solution mapping for a class of variational inequalities of the second kind. *Set-Valued Var. Anal.* to appear.

Holm, D. D., Schmah, T., and Stoica, C. (2009). *Geometric Mechanics and Symmetry*. Oxford University Press, Oxford.

Holmes, R. B. (1973). Smoothness of certain metric projections on Hilbert space. *Trans. Amer. Math. Soc.*, 183:87–100.

Hörmander, L. (1990). *The Analysis of Linear Partial Differential Operators I*, volume 256 of *Grundlehren der Mathematischen Wissenschaften*. Springer Verlag, Berlin/Heidelberg/New York, second edition.

Huilgol, R. R. and You, Z. (2005). Application of the augmented Lagrangian method to steady pipe flows of Bingham, Casson and Herschel-Bulkley fluids. *J. Non-Newton. Fluid Mech.*, 128(2):126–143.

Huybrechs, D. and Olver, S. (2009). Highly oscillatory quadrature. In *Highly oscillatory problems*, volume 366 of *London Math. Soc. Lecture Note Ser.*, pages 25–50, Cambridge. Cambridge University Press.

Ioffe, A. (1991). Variational analysis of a composite function: A formula for the lower second order epi-derivative. *J. Math. Anal. Appl.*, 160:379–405.

Iserles, A. and Nørsett, S. P. (2004). On quadrature methods for highly oscillatory integrals and their implementation. *BIT*, 44(4):755–772.

Iserles, A., Nørsett, S. P., and Olver, S. (2006). Highly oscillatory quadrature: The story so far. In *Numerical Mathematics and Advanced Applications*, pages 97–118, Berlin, Heidelberg. Springer Verlag.

Jarušek, J., Krbec, M., Rao, M., and Sokołowski, J. (2003). Conical differentiability for evolution variational inequalities. *J. Differential Equations*, 193(1):131–146.

Joly, J.-L. and Thelin, F. (1976). Convergence of convex integrals in $\mathcal{L}^p$-spaces. *J. Math. Anal. Appl.*, (54):230–244.

Kato, N. (1989). On the second derivatives of convex functions on Hilbert spaces. *Proc. Amer. Math. Soc.*, 106(3):697–705.

Khludnev, A. M. and Sokołowski, J. (1991). *Modelling and Control in Solid Mechanics*. Birkhäuser Verlag, Berlin/Boston/Berlin.

Kikuchi, F., Nakazato, K., and Ushijima, T. (1984). Finite element approximation of a nonlinear eigenvalue problem related to MHD equilibria. *Japan J. Appl. Math.*, 1(2):369–403.

Kikuchi, N. and Oden, J. T. (1988). *Contact Problems in Elasticity*, volume 8 of *SIAM, Studies in Applied Mathematics*. SIAM, Philadelphia.

Kinderlehrer, D. and Stampacchia, G. (2000). *An Introduction to Variational Inequalities and Their Applications*, volume 31 of *Classics in Applied Mathematics*. SIAM.

Kinnunen, J. and Martio, O. (1997). Hardy's inequality for Sobolev functions. *Math. Res. Lett.*, (4):489 –500.

Krantz, S. G. and Parks, H. R. (2012). *A Primer of Real Analytic Functions*. Birkhäuser Advanced Texts Basler Lehrbücher. Birkhäuser Verlag.

Kufner, A. (1980). *Weighted Sobolev Spaces*. Teubner-Texte zur Mathematik.

Levy, A. B. (1999). Sensitivity of solutions to variational inequalities on Banach spaces. *SIAM J. Control Optim.*, 38(1):50–60.

Lindqvist, P. (1987). Stability of solutions of $\operatorname{div}(|\nabla u|^{p-2}\nabla u) = f$ with varying p. *J. Math. Anal. Appl.*, 127:93–102.

Lindqvist, P. (2017). *Notes on the p-Laplace Equation*. Report 161. University of Jyväskylä, Department of Mathematics and Statistics, second edition.

Maz'ya, V. (2011). *Sobolev Spaces*, volume 342 of *Grundlehren der Mathematischen Wissenschaften*. Springer Verlag, second revised edition.

Meyer, C. and Susu, L. M. (2017). Optimal control of nonsmooth, semilinear parabolic equations. *SIAM J. Control Optim.*, 55(4):2206–2234.

Mignot, F. (1976). Contrôle dans les inéquations variationelles elliptiques. *J. Funct. Anal.*, 22(2):130–185.

Mignot, F. and Puel, J. P. (1984). Optimal control in some variational inequalities. *SIAM J. Control Optim.*, 22(3):466–476.

Mosolov, P. P. and Miasnikov, V. P. (1965). Variational methods in the theory of the fluidity of a viscous-plastic medium. *J. Appl. Math. Mech.*, 29(3):545–577.

Mosolov, P. P. and Miasnikov, V. P. (1966). On stagnant flow regions of a viscous-plastic medium in pipes. *PMM*, 30(4):705–717.

Mosolov, P. P. and Miasnikov, V. P. (1967). On qualitative singularities of the flow of a viscoplastic medium in pipes. *J. Appl. Math. Mech.*, 31(3):609–613.

Müller, G. and Schiela, A. (2017). On the control of time discretized dynamic contact problems. *Comput. Optim. Appl.*, 68(2):243–287.

Musina, R. and Nazarov, A. I. (2017). A note on truncations in fractional Sobolev spaces. *Bull. Math. Sci.*, pages 1–8.

Nečas, J. (2012). *Direct Methods in the Theory of Elliptic Equations*. Springer Verlag, Berlin/Heidelberg.

Niculescu, C. P. and Persson, L. E. (2006). *Convex Functions and Their Applications - A Contemporary Approach*. Springer Verlag, New York.

Noll, D. (1995). Directional differentiability of the metric projection in Hilbert space. *Pacific J. Math.*, 170(2):567–592.

Oden, J. T. and Kikuchi, N. (1980). Theory of variational inequalities with applications to problems of flow through porous media. *Internat. J. Engrg. Sci.*, 18(527):1173–1284.

Outrata, J., Jarušek, J., and Stará, J. (2011). On optimality conditions in control of elliptic variational inequalities. *Set-Valued Var. Anal.*, 19(1):23–42.

Peypouquet, J. (2015). *Convex Optimization in Normed Spaces: Theory, Methods and Examples*. Springer Briefs in Optimization. Springer Verlag.

Piat, V. C. and Cassano, F. S. (1994). Some remarks about the density of smooth functions in weighted Sobolev spaces. *J. Convex Anal.*, (2):135–142.

Poliquin, R. A. and Rockafellar, R. T. (1993). A calculus of epi-derivatives applicable to optimization. *Canad. J. Math.*, 45(1):879–896.

Qi, L. and Sun, J. (1994). A trust region algorithm for minimization of locally Lipschitzian functions. *Math. Program.*, 66(1):25–43.

Rappaz, J. (1984). Approximation of a nondifferentiable nonlinear problem related to MHD equilibria. *Numer. Math.*, 45(1):117–133.

Rockafellar, R. T. (1985). Maximal monotone relations and the second derivatives of nonsmooth functions. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 2(3):167–184.

Rockafellar, R. T. (1990). Generalized second derivatives of convex functions and saddle functions. *Trans. Amer. Math. Soc.*, 322(1):51–77.

Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer Verlag, Berlin.

Rodrigues, J. F. (1987). *Obstacle Problems in Mathematical Physics*. Elsevier, Amsterdam.

Ruzicka, M. (2004). *Nichtlineare Funktionalanalysis*. Springer Verlag, Berlin/Heidelberg.

Scarpa, L. (2017). Well-posedness for a class of doubly nonlinear stochastic PDEs of divergence type. *J. Differential Equations*, 263(4):2113–2156.

Schiela, A. and Wollner, W. (2011). Barrier methods for optimal control problems with convex nonlinear gradient state constraints. *SIAM J. Optim.*, 21(1):269–286.

Schirotzek, W. (2007). *Nonsmooth Analysis*. Springer Verlag, Berlin/Heidelberg.

Schweizer, B. (2013). *Partielle Differentialgleichungen*. Springer Verlag, Berlin/Heidelberg.

Shapiro, A. (1990). On concepts of directional differentiability. *J. Optim. Theory Appl.*, 66(3):477–487.

Shapiro, A. (1992). Perturbation analysis of optimization problems in Banach spaces. *Numer. Funct. Anal. Optim.*, 13(1-2):97–116.

Shapiro, A. (1994a). Directionally nondifferentiable metric projection. *J. Optim. Theory Appl.*, 81(1):203–204.

Shapiro, A. (1994b). Existence and differentiability of metric projections in Hilbert spaces. *SIAM J. Optim.*, 4(1):130–141.

Shapiro, A. (1997). On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM J. Optim.*, 7(2):508–518.

Shapiro, A. (2016). Differentiability properties of metric projections onto convex sets. *J. Optim. Theory Appl.*, 169(3):953–964.

Sofonea, M. and Matei, A. (2012). *Mathematical Models in Contact Mechanics*. Cambridge University Press, New York.

Sokołowski, J. (1988). Sensitivity analysis of contact problems with prescribed friction. *Appl. Math. Optim.*, 18(1):99–117.

Sokołowski, J. and Zolésio, J.-P. (1988). Shape sensitivity analysis of contact problems with prescribed friction. *Nonlinear Anal.*, 12(12):1399–1411.

Sokołowski, J. and Zolésio, J.-P. (1992). *Introduction to Shape Optimization*. Springer Verlag, New York.

Surnachev, M. D. (2014). Density of smooth functions in weighted Sobolev spaces with variable exponent. *Dokl. Math.*, 89(2):146–150.

Temam, R. (1975). A non-linear eigenvalue problem: the shape at equilibrium of a confined plasma. *Arch. Rational Mech. Anal.*, 60(1):51–73.

Tröltzsch, F. (2010). *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*. AMS, Providence.

Vinberg, E. B., Minachin, V., Alekseevskij, D. V., Shvartsman, O. V., and Solodovnikov, A. S. (2013). *Geometry II: Spaces of Constant Curvature*, volume 29 of *Encyclopaedia of Mathematical Sciences*. Springer Verlag.

Wachsmuth, G. (2014). Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM J. Optim.*, 24(4):1914–1932.

Wachsmuth, G. (2016). A guided tour of polyhedric sets. Preprint, TU Chemnitz.

Wachsmuth, G. (2017). Strong stationarity for optimization problems with complementarity constraints in absence of polyhedricity. *Set-Valued Var. Anal.*, 25(1):133–175.

Warner, F. W. (1983). *Foundations of Differentiable Manifolds and Lie Groups*, volume 94 of *Graduate Texts in Mathematics*. Springer Verlag.

Werner, J. (1984). *Optimization Theory and Applications*. Vieweg, Braunschweig.

Wong, R. (2001). *Asymptotic Approximations of Integrals*. SIAM, Philadelphia.

Yousept, I. (2017). Hyperbolic Maxwell variational inequalities for Bean's critical-state model in type-II superconductivity. *SIAM J. Numer. Anal.*, 55(5):2444–2464.

Zarantonello, E. H. (1971). Projections on convex sets in Hilbert space and spectral theory I and II. *Contrib. Nonlinear Funct. Anal.*, pages 237–424.

Zhikov, V. V. (1998). Weighted Sobolev spaces. *Sbor. Math.*, 189(8):1139–1170.

Zhikov, V. V. (2013). Density of smooth functions in weighted Sobolev spaces. *Dokl. Math.*, 88(3):669–673.

Ziemer, W. P. (1989). *Weakly Differentiable Functions*. Springer Verlag, New York.

Zowe, J. and Kurcyusz, S. (1979). Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.*, 5(1):49–62.

List of Symbols

Spaces and Sets

$\mathbb{N}$	Natural numbers
$\mathbb{Q}$	Rational numbers
$\mathbb{R}$	Real numbers
$\mathbb{R}^+$	Positive real numbers
$\mathbb{R}^{3\times3}_{sym}$	Symmetric 3×3-matrices
$\mathbb{R}^{3\times3}_{dev}$	See Assumption 4.3.4
$\mathbb{Z}$	Integers
$B_r(x), B_r^X(x)$	Closed ball of radius $r > 0$ around x in a space X
H	Hilbert space
I, J	Intervals
K, L	Convex, non-empty sets (K is typically the domain of j in (P))
$\mathrm{O}(d)$	Orthogonal group in dimension d
U, V	Hilbert spaces
W, W_p	Rectification neighborhood
X, Y	Banach spaces
X^*	Topological dual space of a Banach space X
$C_c(\Omega)$	Space of continuous functions with compact support, see Definition 3.4.1
$C_c^\infty(\Omega)$	Space of smooth functions with compact support
$C^\infty(\Omega)$	Space of smooth functions
$C^{m,q}(\Omega)$	Hölder space, see [Adams, 1975]
$\mathcal{D}'(\mathbb{R})$	Space of distributions
$H^m(\Omega)$	Classical H^m-space, see [Attouch et al., 2006, Section 5.2]
$H_0^1(\Omega)$	Closure of $C_c^\infty(\Omega)$ w.r.t. the H^1-norm, see [Attouch et al., 2006, Section 5.2]
$H_D^1(\Omega, \mathbb{R}^3)$	See Assumption 4.3.4
$H^{1/2}(\partial\Omega)$	Fractional Sobolev space on the boundary, cf. Lemma 5.1.21
$H^{-1}(\Omega)$	Dual of $H_0^1(\Omega)$, see [Attouch et al., 2006, Section 5.2]
$H^{-1/2}(\partial\Omega)$	Dual of $H^{1/2}(\partial\Omega)$
$H(\mathrm{div}; D)$	See (5.17) and [Girault and Raviart, 1986, Section 2.2]
$L(X, Y)$	Space of continuous linear functions from X to Y
$L^0(\Omega, [0, \infty])$	Vector space of (equivalence classes of) non-negative measurable functions
$L^q(\Omega, H)$	Bochner space (also $L^q(\Omega, \mu) := L^q(\Omega) := L^q(\Omega, \mathbb{R})$, case $q = 0$ as above)
$W^{m,q}(\Omega)$	Sobolev space, see [Attouch et al., 2006, Section 5.1]
$W_0^{m,q}(\Omega)$	$W^{m,q}$-space with vanishing trace(s), see [Attouch et al., 2006, Section 5.1]
$\mathcal{A}, \mathcal{A}_i, \mathcal{A}^\circ, \mathcal{A}_i^\circ$	See Definition 5.1.23
$\mathcal{B}, \mathcal{B}_i, \mathcal{B}^\circ$	See Definition 5.1.23
$\mathcal{C}$	Exceptional set in Assumption 5.2.4
$\mathcal{F}_{ad}$	Set of admissible controls, cf. Assumption 6.1.1
$\mathcal{I}, \mathcal{I}_i$	See Definition 5.1.23
$\mathcal{K}_j^{red}(v, \varphi)$	Reduced critical cone, see Definition 1.3.3
$\mathcal{M}, \mathcal{N}$	Submanifold of the Euclidean space
$\mathcal{N}_+$	See Assumption 5.2.4

$\mathcal{N}_-$	See Assumption 5.2.4
$\mathcal{N}_0$	See Remark 5.2.5
$\mathcal{N}_L(x)$	Normal cone, see Definition 1.1.1
$\mathcal{T}_L^{rad}(x)$	Radial cone, see Definition 1.1.1
$\mathcal{T}_L(x)$	Tangent cone, see Definition 1.1.1
$\mathcal{T}_L^{crit}(x)$	Critical cone, see (1.28)
$\mathcal{T}_L^2(x, z)$	Second-order tangent set, see Definition 3.3.1
$\mathcal{W}$	Exceptional set in Lemma 5.1.27
$\mathfrak{B}(\Omega)$	Borel σ-algebra of a topological space Ω
Γ	Graph of a (potentially set-valued) function
Σ	σ-algebra
Ω	Measurable space/open subset of $\mathbb{R}^d$
$\partial\Omega$	Boundary of a set Ω

Functions and Operators

$\lvert\cdot\rvert$	Absolute value function
$\lvert\cdot\rvert_m$	Seminorm in (4.4) and (4.20), respectively
$\lVert\cdot\rVert$	Norm
$\lVert\cdot\rVert_1$	1-norm on the Euclidean space
$\lVert\cdot\rVert_2$	2-norm on the Euclidean space
$\lVert\cdot\rVert_\infty$	Maximum norm on the Euclidean space, see Assumption 3.4.6
$\lVert\cdot\rVert_F$	Frobenius norm
$(\cdot)^+$	Shorthand notation for $\max(0, \cdot)$
$(\cdot)^-$	Shorthand notation for $\min(0, \cdot)$
$\langle\cdot, \cdot\rangle$	Dual pairing
$(\cdot, \cdot)$	Scalar product
$\cdot$	(between two vectors) Standard scalar product on the Euclidean space
$:$	(between two matrices) Inner product associated with the Frobenius norm
∇	(Weak) gradient
Δ	(Weak) Laplace operator
∂_m	(Weak) partial derivative on $\mathbb{R}^d$ w.r.t. the m-th coordinate
∂_v	Partial Fréchet derivative w.r.t. the variable v
∂	Convex subdifferential, see Section 1.1
$\mathbb{C}$	Elasticity tensor, see Assumption 4.3.4
$\mathbb{H}$	Hardening modulus, see Assumption 4.3.4
A, B	Functions mapping a space into its dual space (A is typically the map in (P))
F, G	Vector-valued mappings
F^*	Adjoint of a linear map
$F'(x; \cdot)$	(Hadamard) directional derivative, see Definition 1.1.2
$F'(x)$	First Fréchet derivative, see Definition 1.1.2
$F''(x)$	Second Gâteaux/Fréchet derivative, see [Drábek and Milota, 2007]
Id	Identity map
$Q_j^{v,\varphi}$	Second subderivative, see Definition 1.3.1
S, T	Solution operators
b_m	Bilinear forms in Assumption 4.2.1 and Assumption 4.3.1, respectively
j, k, l	Scalar functions (j is typically the function in (P))
k_m	Scalar functions in (4.4) and (4.20), respectively
$q_j^{w,\varphi}$	Bilinear form generating the second subderivative, see Corollary 1.4.4

$\mathcal{E}$	Dirichlet form, see Definition 3.4.1
$\mathcal{H}^d$	(Restriction of the) d-dimensional Hausdorff measure, see [Attouch et al., 2006]
$\mathcal{J}$	Objective functional
$\mathcal{L}$	Lagrange function, see Remark 3.3.7
$\mathcal{L}^d$	(Restriction of the) d-dimensional Lebesgue measure
aff	Affine hull of a set
cl	Topological closure of a set
conv	Convex hull
dist	Distance in a normed space
dom	Domain of a function, see Section 1.1
epi	Epigraph of a function, see Section 1.1
$\operatorname{ess\,sup}$	Essential supremum
graph	Graph of a (potentially set-valued) function, see Section 1.1
int	Interior of a set
$\ker$	Kernel of a linear function
span	Linear span of a set, see Lemma 1.2.3
$\max(\cdot,\cdot)$	Maximum between two real numbers
$\min(\cdot,\cdot)$	Minimum between two real numbers
sgn	Signum function
$\sin$	Sine function
supp	Support of a function
tr	Trace operator, see [Attouch et al., 2006, Section 5.6]
tr_ν	Normal trace, see Corollary 3.4.4
$\mathbb{1}_D$	Indicator function of a set D, i.e., $\mathbb{1}_D(x) = 1$ if $x \in D$, $\mathbb{1}_D(x) = 0$ if $x \notin D$
ℓ	Level set function in Lemma 5.1.28
χ_D	Characteristic function of a set D, i.e., $\chi_D(x) = 0$ if $x \in D$, $\chi_D(x) = \infty$ if $x \notin D$
γ	Parametrizing Lipschitz function, see Definition 5.1.16
μ	Measure
ι	Riesz isomorphism/Canonical embedding

Miscellaneous

$\rightharpoonup$	Weak convergence
$\searrow$	Convergence from above in $\mathbb{R}$
$\nearrow$	Convergence from below in $\mathbb{R}$
$\exists$	Exists
$\forall$	For all
$(\cdot)^\perp$	Rotated vector, see (5.5)
$F\vert_D$	Restriction of a function F to a set D
$F\vert_{(x,y)}$	Evaluation of a function F in (x, y)
$O(\cdot), o(\cdot)$	Landau notation
d	Spatial dimension
$e_1, ..., e_d$	Standard basis of $\mathbb{R}^d$, see Assumption 3.4.9
$\tilde{e}_1, ..., \tilde{e}_d$	Slanted orthogonal basis of $\mathbb{R}^d$, see Assumption 3.4.9
f	Argument of the solution map
$\bar{f}$	Stationary control
g	Direction of the directional derivative
$\bar{p}$	Adjoint state in the strong stationarity system
u, v, w, x, y, z	Elements of Hilbert/Banach spaces (w is typically the solution of (P))

$\mathfrak{f}$	Function generating the rotationally symmetric f in Assumption 5.1.3
$\mathfrak{w}$	Function generating the rotationally symmetric solution in Section 5.1.2
α, β	Real numbers
δ_t	Difference quotients of the solution map, see (1.18)
δ	Limit of the difference quotients of the solution map
δ_0	Dirac distribution at the origin
$\bar{\eta}$	Multiplier in the strong stationarity system
φ, λ	Subgradients (λ is typically a multiplier as in the chain rule)
σ_0	Yield stress, see Assumption 4.3.4
ω	Lipschitz function in Assumption 5.1.25
$\mathbf{1}$	Vector with one in every entry, see Assumption 3.4.9