

On the differentiability of the minimal and maximal solution maps of elliptic quasi-variational inequalities

Amal Alphonse*

Michael Hintermüller†

Carlos N. Rautenberg‡

October 11, 2021

Abstract

In this note, we prove that the minimal and maximal solution maps associated to elliptic quasi-variational inequalities of obstacle type are directionally differentiable with respect to the forcing term and for directions that are signed. Along the way, we show that the minimal and maximal solutions can be seen as monotone limits of solutions of certain variational inequalities and that the aforementioned directional derivatives can also be characterised as the monotone limits of sequences of directional derivatives associated to variational inequalities. We conclude the paper with some examples and an application to thermoforming.

Keywords: Quasi-variational inequality, obstacle problem, directional differentiability, minimal and maximal solutions, ordered solutions.

1 Introduction

Quasi-variational inequalities (QVIs) are variational inequalities (VIs) where the constraint set over which the solution is sought also depends on the solution itself. As such, QVI problems are highly nonlinear and nonconvex and in sharp contrast to the usual setting for VIs, QVIs usually possess multiple solutions. In certain situations, the set of solutions can be ordered in the sense that there exist minimal and maximal solutions. In this paper, we address the directional differentiability of the maps taking the source term of a QVI into the minimal and maximal solutions. The above-mentioned quirks of QVIs endow their study with substantial technical issues to overcome when examining questions of stability analysis and differential sensitivity.

QVIs were first formulated by Bensoussan and Lions [12, 25] in the modelling of stochastic impulse controls. Applications of QVIs are ubiquitous. Among some, we mention thermoforming processes [1, 5], the formation and growth of lakes, rivers and sandpiles [33, 10, 31, 11], generalised Nash equilibrium games [19, 16, 30], and magnetisation of superconductors [23, 9, 32, 35]. Additional details and references can be found in our survey paper [2] and the book [8].

We focus on elliptic QVIs of obstacle type (these are also known as implicit obstacle problems). The precise formulation is as follows. Let $V \subset H$ be a continuous and dense embedding of separable Hilbert spaces and denote the inner product on H by $(\cdot, \cdot) = (\cdot, \cdot)_H$. Suppose that there exists an ordering to elements of H via a closed convex cone H_+ that satisfies

$$H_+ = \{h \in H : (h, g) \geq 0 \quad \forall g \in H_+\}.$$

The ordering is defined by: $h_1 \leq h_2$ if and only if $h_2 - h_1 \in H_+$. This endows an ordering for V in the obvious way and we write $V_+ := \{v \in V : v \geq 0\}$. It also induces one for the dual space V^* via

$$V_+^* := \{f \in V^* : \langle f, v \rangle \geq 0 \quad \forall v \in V_+\},$$

where $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{V^*, V}$ is the standard duality product. We write h^+ for the orthogonal projection of $h \in H$ onto the space H_+ and we use the decomposition $h = h^+ - h^-$. Suppose that $v \in V$ implies that $v^+ \in V$ and that there exists a constant $C > 0$ such that $\|v^+\|_V \leq C \|v\|_V$ for all $v \in V$.¹

Let $A: V \rightarrow V^*$ a linear operator that satisfies the following properties for all $u, v \in V$:

$$\langle Au, v \rangle \leq C_b \|u\|_V \|v\|_V, \quad (\text{boundedness})$$

$$\langle Au, u \rangle \geq C_a \|u\|_V^2, \quad (\text{coercivity})$$

$$\langle Au^+, u^- \rangle \leq 0. \quad (\text{T-monotonicity})$$

The authors thank the referees for their excellent comments which helped to improve the paper. AA and MH were partially supported by the DFG through the DFG SPP 1962 Priority Programme *Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization* within project 10. MH and CNR acknowledge the support of Germany's Excellence Strategy - The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689) within project AA4-3. In addition, MH acknowledges the support of SFB-TRR154 within subproject B02, and CNR was supported by NSF grant DMS-2012391.

*Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany (alphonse@wias-berlin.de)

†Weierstrass Institute, Mohrenstrasse 39, 10117 Berlin, Germany (hintermueller@wias-berlin.de)

‡Department of Mathematical Sciences and the Center for Mathematics and Artificial Intelligence (CMAI), George Mason University, Fairfax, VA 22030, USA (crautenb@gmu.edu)

¹Some concrete realisations of this setup will be presented in Examples 6.1.

We consider an obstacle map

$$\Phi: H \rightarrow V$$

and the associated constraint set $\mathbf{K}: H \rightrightarrows V$ defined by

$$\mathbf{K}(y) := \{v \in V : v \leq \Phi(y)\}.$$

Given $f \in V^*$, we study QVIs of the form

$$\text{find } y \in \mathbf{K}(y) : \langle Ay - f, y - v \rangle \leq 0 \quad \forall v \in \mathbf{K}(y). \quad (1)$$

In this work, we always take Φ to be increasing, i.e.,

$$u \leq v \text{ implies } \Phi(u) \leq \Phi(v),$$

and we suppose also that (unless stated otherwise) it satisfies the property

$$\text{if } \{v_n\} \subset V \text{ is decreasing with } v_n \rightarrow v \text{ in } V \text{ and } v \leq \Phi(v_n), \text{ then } v \leq \Phi(v). \quad (2)$$

This holds, for example, if Φ is linear and continuous, or if $\Phi: V \rightarrow V$ is weakly sequentially continuous (that is, $v_n \rightarrow v$ in V implies $\Phi(v_n) \rightarrow \Phi(v)$ in V).

We define $\mathbf{Q}$ to be the source-to-solution map associated to (1), so that the inequality reads $y \in \mathbf{Q}(f)$. To prove that the set $\mathbf{Q}(f)$ is non-empty (and indeed to properly define the problem under study in this article) we need some additional details.

1.1 Existence of (extremal) solutions

Fixing an obstacle $\varphi \in V$, consider the VI

$$u \in \mathbf{K}(\varphi) : \langle Au - f, u - v \rangle \leq 0 \quad \forall v \in \mathbf{K}(\varphi)$$

and denote its solution map $S: V^* \times H \rightarrow V$ so that the above can be stated as $u = S(f, \varphi)$. It follows that $\mathbf{Q}(f)$ is the set of fixed points of $\varphi \rightarrow S(f, \varphi)$. In order to show the presence of fixed points, we are going to assume the existence of a *subsolution* $\underline{u}$ and a *supersolution* $\bar{u}$ for $S(f, \cdot)$, that is,

$$\exists \underline{u}, \bar{u} \in V \quad \text{such that} \quad \underline{u} \leq S(f, \underline{u}) \quad \text{and} \quad \bar{u} \geq S(f, \bar{u}),$$

and such that the subsolution lies below the supersolution:

$$\underline{u} \leq \bar{u}.$$

Remark 1.1. For a supersolution, we can take any $\bar{u} \in V$ satisfying $\bar{u} \geq A^{-1}f$ where the right-hand side is (by definition) the solution of the equation $Az = f$. This is a valid choice since $A^{-1}f$ can be considered as the solution² of the VI with source f and obstacle equal to ∞ , which by the comparison principle [34, Theorem 5.1, §4:5], exceeds $S(f, \bar{u})$. If $f \geq 0$ and $\Phi(0) \geq 0$, then we may take $\underline{u} = 0$ to be a subsolution: $0 = S(0, 0) \leq S(f, 0)$.

Under these circumstances, we can apply the Birkhoff–Tartar theory [36, 13] (see also [7, Chapter 15.2.2] and [27, Chapter 2.5]) of fixed points in vector lattices to obtain not only that

$$\mathbf{Q}(f) \cap [\underline{u}, \bar{u}] \neq \emptyset$$

(i.e., (1) has solutions on $[\underline{u}, \bar{u}]$), but moreover, there exists a minimal solution $\mathbf{m}(f)$ and a maximal solution $\mathbf{M}(f)$ in this interval with respect to the ordering introduced above. These satisfy

$$\mathbf{m}(f) \leq u \leq \mathbf{M}(f) \quad \forall u \in \mathbf{Q}(f) \cap [\underline{u}, \bar{u}].$$

Remark 1.2. The condition (2) is not required for the above results.

One should bear in mind that these maps $\mathbf{m}$ and $\mathbf{M}$ are associated a particular interval $[\underline{u}, \bar{u}]$ and sometimes we may show the dependence on the interval explicitly by writing $\mathbf{m}_{[\underline{u}, \bar{u}]}$ and $\mathbf{M}_{[\underline{u}, \bar{u}]}$, or more generally, provided they exist, we denote the minimal and maximal solutions on some interval A by $\mathbf{m}_A$ and $\mathbf{M}_A$ respectively.

However, when discussing the minimal solution map, only the lower bound of the interval under consideration is pertinent due to the trivial observation that if $w \leq a \leq b$, $\mathbf{m}_{[w, a]}(f) = \mathbf{m}_{[w, b]}(f)$ whenever these exist, and similarly, only the upper bound of the interval for the maximal solution map is important since if $a \leq b \leq z$ then $\mathbf{M}_{[a, z]}(f) = \mathbf{M}_{[b, z]}(f)$.

²Informally, $A^{-1}f = S(f, \Phi^{-1}(\infty))$.

1.2 Aim of the article

In this paper, we are interested in the directional differentiability of $f \mapsto \mathbf{m}(f)$ and $f \mapsto \mathbf{M}(f)$. We will show that, under some assumptions, these maps are indeed directionally differentiable for a subset (that we will specify below) of directions belonging to V^* . That is, we prove the existence of the following limits:

$$\lim_{s \rightarrow 0^+} \frac{\mathbf{m}(f + sd) - \mathbf{m}(f)}{s} \quad \text{and} \quad \lim_{s \rightarrow 0^+} \frac{\mathbf{M}(f + sd) - \mathbf{M}(f)}{s}. \quad (3)$$

We also give a characterisation of the derivative in terms of a related QVI. This builds upon our previous works [1, 4, 3] in two ways. In [1, 4] we showed that $\mathbf{Q}$ has a directional derivative in the sense that we proved that for every $u \in \mathbf{Q}(f)$, given a direction $d \in V^*$, there exists $u^s \in \mathbf{Q}(f + sd)$ and (a directional derivative) $\alpha \in V$ such that

$$\lim_{s \rightarrow 0^+} \frac{u^s - u}{s} = \alpha$$

(with the limit taken in V). In [4], we derived further existence results for (1) and procedures to iteratively approach solutions of the QVI. Furthermore, we also obtained stationarity systems for optimal control problems with QVI constraints. On the other hand, in [3], we studied continuity properties of the extremal operators $\mathbf{m}$ and $\mathbf{M}$. This work then can be considered as a bridge between these two sets of papers.

The motivations for this study are manifold:

- (i) the mathematical problem itself is challenging and interesting,
- (ii) suitable characterisations of the directional derivative could be used to develop efficient bundle-free numerical solvers and solution algorithms along the lines of [21],
- (iii) in applications, it can be important to know the influence of the source term on the solution, for example in thermoforming (which can be modelled through QVIs, see [1, §6] and also §6.3 later), manufacturers may wish to understand the sensitivity of the desired shapes or products to changes in heating processes under their control,
- (iv) for optimal control problems with control-to-state maps involving $\mathbf{m}$ or $\mathbf{M}$, differentiability of the maps is necessary in order to state certain first-order stationarity conditions satisfied by the optimal control,
- (v) in applications involving optimal control problems with QVI constraints, as typically there are many states associated to a single (optimal) control (due to non-uniqueness of solutions), it can be important to minimise, for instance, the difference $\mathbf{M}(f) - \mathbf{m}(f)$. In the case of thermoforming, manufacturers may wish to reproduce shapes or products that are within a certain acceptable tolerance value.

Note that continuity properties of these maps (studied under a similar functional setting in [3]) are vital for the existence results for optimal control problems.

The idea is to base our developments on the differentiability results obtained in [4]; let us recall this and set the scene in the next section. First, some words on our notation and conventions.

Notation. Throughout the rest of the paper, we shall use the notation $o(\cdot)$ to denote a remainder term, i.e., $s^{-1}o(s) \rightarrow 0$ in V as $s \rightarrow 0^+$. We will write $v_n \nearrow v$ to mean $v_n \rightarrow v$ and $v_{n+1} \geq v_n$ for all n (i.e., $\{v_n\}$ is an increasing sequence that converges to v). Similarly, $\searrow$ will refer to decreasing convergent sequences. The notation $B_R(z)$ will be used to mean the closed ball in V of radius R centred at z .

2 Background on directional differentiability for QVIs

Recall from §1 the constraint set mapping $\mathbf{K}: H \rightrightarrows V$ which is closed (since V_+ is closed) and convex valued, and associated to this, we define the *tangent cone* of $\mathbf{K}(w)$ at a point $u \in \mathbf{K}(w)$ by

$$\mathcal{T}_{\mathbf{K}(w)}(u) := \overline{\mathcal{R}_{\mathbf{K}(w)}(u)} \quad \text{where} \quad \mathcal{R}_{\mathbf{K}(w)}(u) := \{h \in V : \exists s^* > 0 \text{ such that } u + sh \in \mathbf{K}(w) \forall s \in [0, s^*]\}.$$

The latter is known as the *radial cone* and can be regarded as the set of all feasible directions. See for example [14, §2.2.4] for more details on these concepts.

A fundamental result that we shall need is the following, which, under certain circumstances, tells us that $\mathbf{Q}$ has a directional derivative in a certain sense and provides a characterisation.

Theorem 2.1 ([4, Theorem 3.2]). *Given $f \in V^*$ and $d \in V^*$, for every $u \in \mathbf{Q}(f)$, under the local assumptions*

$$\text{there exists } \epsilon > 0 \text{ such that } \Phi: B_\epsilon(u) \rightarrow V \text{ is Lipschitz with Lipschitz constant } C_\Phi < C_a/(C_a + C_b), \quad (4)$$

$$\Phi: V \rightarrow V \text{ is Hadamard directionally differentiable at } u, \quad (5)$$

there exists $u^s \in \mathbf{Q}(f + sd) \cap B_R(u)$ (where $0 < R \leq \epsilon$) and $\alpha = \alpha(d) \in V$ such that

$$u^s = u + s\alpha + o(s),$$

where α satisfies the QVI

$$\begin{aligned} \alpha \in \mathcal{K}^u(\alpha) : \langle A\alpha - d, \alpha - v \rangle &\leq 0 \quad \forall v \in \mathcal{K}^u(\alpha), \\ \mathcal{K}^u(\alpha) &:= \Phi'(u)(\alpha) + \mathcal{T}_{\mathbf{K}(u)}(u) \cap [f - Au]^\perp. \end{aligned} \quad (6)$$

The directional derivative $\alpha = \alpha(d)$ is positively homogeneous in d .

It is worth emphasising that in fact, [4, Theorem 3.2] is posed in a much more general functional setting than under consideration in this paper (described in the introduction).

Remark 2.2. *It is not too difficult to see that [4, Lemma 3.12] if Φ is Hadamard differentiable on $B_\epsilon(u)$, then the condition*

$$\exists \epsilon > 0 : \|\Phi'(z)(v)\|_V \leq C_\Phi \|v\|_V \quad \forall z \in B_\epsilon(u), \forall v \in V, \text{ where } C_\Phi < C_a/(C_a + C_b)$$

implies (4); this can in certain situations be easier to verify.

The set $\mathcal{K}^u(\alpha)$ appearing in (6) is called a *critical cone* and we can be a bit more precise about its structure if additional assumptions on the functional setting are available as the next remark shows.

Remark 2.3 (Dirichlet space case). *Suppose that $H := L^2(X; \mu)$ where X is a locally compact, separable metric space and μ is a positive Radon measure on X with full support³, and let $V \subset H$ be a dense subspace (the ordering is given by the usual a.e. ordering of functions).*

Assume that

- (i) *there exists a symmetric, positive semidefinite bilinear form $\xi : V \times V \rightarrow \mathbb{R}$ such that V is a Hilbert space when endowed with*

$$(\cdot, \cdot)_V := (\cdot, \cdot)_H + \xi(\cdot, \cdot),$$

- (ii) *if $u \in V$ then $\hat{u} := \min(u^+, 1) \in V$ and $\xi(\hat{u}, \hat{u}) \leq \xi(u, u)$ (this is the Markov property),*

- (iii) *the following density conditions hold:*

$$V \cap C_c(X) \xrightarrow{d} C_c(X) \quad \text{and} \quad V \cap C_c(X) \xrightarrow{d} V.$$

Under these circumstances, V is known as a regular Dirichlet space and the notions of capacity, quasi-continuity and related concepts are well defined, see [17, §2.1] and [18, §3] for more details. Furthermore, from the work⁴ of Mignot [26, Theorem 3.2, Lemma 3.2], the critical cone appearing in (6) can be expressed explicitly as

$$\mathcal{K}^u(w) = \{\varphi \in V : \varphi \leq \Phi'(u)(w) \text{ q.e. on } \mathcal{A}(u) \text{ and } \langle Au - f, \varphi - \Phi'(u)(w) \rangle = 0\}.$$

Here, ‘q.e.’ means quasi-everywhere, i.e., everywhere except on a set of capacity zero, and $\mathcal{A}(u)$ refers to the active or coincidence set⁵ of the solution y to the QVI, i.e.,

$$\mathcal{A}(u) := \{x \in X : u(x) = \Phi(u)(x)\} \quad \text{for } u \in V.$$

Note that in the context of V and H being Sobolev or Lebesgue spaces over a domain Ω , in certain cases the space $X = \overline{\Omega}$ and not merely Ω ! See Examples 6.1 for more details.

3 Strategy of proof and some preliminary results

Let us sketch the procedure we will undertake for showing directional differentiability for the minimal solution map m . Our aim is to show that, given a source term f and a direction d , there exists an element $m'(f)(d)$ such that

$$m(f + sd) = m(f) + sm'(f)(d) + o(s).$$

As we have seen, Theorem 2.1 states that under certain assumptions, given $u \in \mathbf{Q}(f)$, there exists $u^s \in \mathbf{Q}(f + sd)$ and $\alpha \in V$ such that

$$u^s = u + s\alpha + o(s).$$

We may select u to be the minimal solution $m(f)$ and it remains to prove that the selection mechanism of the theorem that furnishes the u^s is indeed such that $u^s \equiv m(f + sd)$. To do this, we need to take a closer look at the method of proof of the cited theorem. The proof relies on

- (i) creating an iterative sequence of solutions of VIs:

$$\begin{aligned} u_n^s &= S(f + sd, u_{n-1}^s), \\ u_0^s &= u, \end{aligned}$$

(recall the definition of S from §1.1)

- (ii) obtaining, by applying the sensitivity results for VIs [26, 37], expansion formulas of the type

$$u_n^s = u + s\alpha_n + o_n(s) \tag{7}$$

for these elements, and then

- (iii) passing to the limit $n \rightarrow \infty$ and identifying the limits of $\{u_n^s\}$, $\{\alpha_n\}$ and $\{o_n\}$.

Thus, it is clear that we need to show that the limit of $\{u_n^s\}$ (which exists) is indeed $m(f + sd)$. For this purpose, we need to prove some properties of m and M which we shall do so in a series of lemmas. Before we begin, it is important to record that in [4, Proposition 3.7], under the local Lipschitz assumption (4), we showed (via a Banach fixed point theorem argument) that

$$u_n^s \rightarrow u^s \text{ in } V \text{ where } u^s \in \mathbf{Q}(f + sd).$$

³This means that μ is a non-negative Borel measure which is finite on compact sets and strictly positive on non-empty open sets.

⁴More specifically, we have the polyhedricity of sets of obstacle type and the differentiability of VI solution maps associated to these types of constraint sets in [26, Theorem 3.3].

⁵Strictly speaking, we take the quasi-continuous representatives of the functions appearing in the definition of the active set so that $\mathcal{A}(y)$ is quasi-closed and defined up to sets of capacity zero.

3.1 Monotone convergence to extremal solutions

Let $f \in V^*$ be a source term. In this section, we assume that we are given a subsolution $\underline{u}$ and a supersolution $\bar{u}$ for $S(f, \cdot)$ with $\underline{u} \leq \bar{u}$ (this is the same setup as in §1.1). We introduce the assumption

$$\exists v_0 \in V : v_0 \leq \Phi(\underline{u}). \quad (8)$$

Remark 3.1. The condition (8) essentially asks for $\mathbf{K}(\underline{u})$ to be non-empty. A typical situation is where $f \in V_+^*$ is non-negative and Φ is taken such that $\Phi(0) \geq 0$, in which case $\underline{u} := 0$ is a subsolution (as shown in Remark 1.1) so that $v_0 \equiv 0$ is a possibility.

The following lemma demonstrates that, under some compactness on Φ , starting at a subsolution we can converge to the minimal solution. The result is not strictly necessary for our main theorem yet we provide it for interest and completeness (note that the required condition (9) is much stronger than (2)).

Lemma 3.2. Assume (8) and that

$$\Phi : V \rightarrow V \text{ is completely continuous.} \quad (9)$$

Then the sequence

$$\begin{aligned} u_n &:= S(f, u_{n-1}), \\ u_0 &:= \underline{u}. \end{aligned} \quad (10)$$

satisfies $u_n \nearrow \mathbf{m}(f)$ in V .

Proof. By definition, $u_0 \leq \mathbf{m}(f)$. By definition of subsolution and by using the comparison principle [34, Theorem 5.1, §4:5], $u_0 \leq S(f, u_0) = u_1 \leq S(f, \mathbf{m}(f)) = \mathbf{m}(f)$. Arguing in a similar way, $u_0 \leq u_n \leq u_{n+1} \leq \mathbf{m}(f)$ for all n .

Since Φ is increasing, it follows that $v_0 \leq \Phi(u_n)$ for each n . Hence, we may test the VI (for u_n)

$$\langle Au_n - f, u_n - v \rangle \leq 0 \quad \forall v \in V : v \leq \Phi(u_{n-1})$$

with v_0 and use basic manipulations to obtain a uniform bound on u_n , which in combination with the fact that $\{u_n\}$ is monotonic, leads to the existence of some $u \in V$ such that $u_n \rightarrow u$ in V for the full sequence [4, Lemma 2.3]. Now, in the VI for u_n , test with $v_n := v^* - \Phi(u) + \Phi(u_{n-1})$ where $v^* \in V$ satisfies $v^* \leq \Phi(u)$. This is clearly a feasible test function such that $v_n \rightarrow v^*$ in V (strongly by (9)), and we can pass to the limit in the resulting inequality to find exactly that $u \in \mathbf{Q}(f)$. It follows also that $u \in [\underline{u}, \mathbf{m}(f)]$ and therefore $u = \mathbf{m}(f)$. The strong convergence of $\{u_n\}$ is a result of (9) along with a standard continuous dependence estimate (e.g., see [1, Equation (21)])

$$\|u_n - u\|_V \leq C \|\Phi(u_{n-1}) - \Phi(u)\|_V. \quad \square$$

In (10), if we instead start with the initial iterate at a supersolution of $S(f, \cdot)$, we are able to provide analogous results. Indeed, a similar argument to the proof of Lemma 3.2 proves the next lemma but note that complete continuity of Φ is not needed thanks to the structure of the unilaterally bounded constraint sets under consideration.

Lemma 3.3. Assume (8). Then the sequence

$$\begin{aligned} u_n &:= S(f, u_{n-1}), \\ u_0 &:= \bar{u}, \end{aligned} \quad (11)$$

satisfies $u_n \searrow \mathbf{M}(f)$ weakly in V . The convergence is strong under (9).

Proof. By definition, $u_0 \geq \mathbf{M}(f)$. By definition of supersolution and by using the comparison principle, $u_0 \geq S(f, u_0) = u_1 \geq S(f, \mathbf{M}(f)) = \mathbf{M}(f)$ and hence $u_0 \geq u_n \geq u_{n+1} \geq \mathbf{M}(f)$ for all n .

Clearly, $u_n \geq \underline{u}$, thus it follows from (8) that $v_0 \leq \Phi(u_n)$ for each n . Hence, we may test the VI for u_n with v_0 to obtain a uniform bound on u_n and therefore we have $u_n \rightarrow u$ in V for some u (due to the monotonicity of $\{u_n\}$, as in the previous proof).

Now, take an arbitrary $v^* \in V$ with $v^* \leq \Phi(u)$. Since $\{u_n\}$ is decreasing, we have $u_n \leq u_j$ for all $n \geq j$ with j fixed, i.e., $u_j - u_n \in V_+$. Passing to the limit $n \rightarrow \infty$ (using the fact that V_+ is closed and convex, hence weakly sequentially closed), we find $u_j - u \in V_+$ so that $u \leq u_j$. Using the fact that Φ is increasing, we see that $v^* \leq \Phi(u_j)$ for any $j \in \mathbb{N}$, i.e., v^* is a valid test function for the VI for u_n . Taking the limit in the VI for u_n allows us to deduce that u satisfies

$$\langle Au - f, u - v \rangle \leq 0 \quad \forall v \in V : v \leq \Phi(u).$$

To conclude that $u \in \mathbf{Q}(f)$, we still need to show that $u \leq \Phi(u)$, but this is a consequence of (2).

It follows also that $u \in [\mathbf{M}(f), \bar{u}]$ and therefore $u = \mathbf{M}(f)$. The statement regarding strong convergence follows again by the continuous dependence estimate as above. $\square$

4 The minimal solution map

We will show in this section that the minimal solution map is directionally differentiable under some assumptions by utilising the strategy explained in the previous section.

As before, we take $\underline{u}$ to be a subsolution for $S(f, \cdot)$. Let $s \geq 0$ be a small parameter and take a direction $d \in V_+^*$. Since $\underline{u} \leq S(f, \underline{u}) \leq S(f + sd, \underline{u})$, $\underline{u}$ is also a subsolution for $S(f + sd, \cdot)$. In the other direction, we suppose, as a standing assumption in this section, that

$$\text{there exists } s_0 > 0 \text{ such that for all } s \leq s_0, \bar{u} \text{ is a supersolution for } S(f + sd, \cdot) \quad (12)$$

($\bar{u}$ should remain independent of s). Then, by the argument in §1.1, we have the non-emptiness of the set $\mathbf{Q}(f + sd) \cap [\underline{u}, \bar{u}]$. Furthermore, $\mathbf{m}(f + sd)$ exists on $[\underline{u}, \bar{u}]$ (for small s).

Remark 4.1. Asking for $\bar{u}$ to be a supersolution for the perturbed problem (parametrised by s) is not a stringent requirement since any supersolution for $S(f + sd, \cdot)$ is also a supersolution for $S(f + rd, \cdot)$ for $s \geq r \geq 0$, and because we will be sending $s \rightarrow 0$, we may always start by taking, for example, $\bar{u}$ to be a supersolution of $S(f + d, \cdot)$.

Example 4.2. As indicated in Remark 1.1, a concrete choice of a supersolution satisfying (12) is

$$\bar{u} = A^{-1}(f + d).$$

Lemma 4.3. Let $d \in V_+^*$ and (8) hold. Then $\mathbf{m}(f + sd) \geq \mathbf{m}(f)$.

Proof. If Φ is, in addition, completely continuous, this is easy to show. Indeed, defining

$$\begin{aligned} y_n^s &:= S(f + sd, y_{n-1}^s), \\ y_0^s &:= \underline{u}, \end{aligned}$$

and with u_n defined as in (10), we see that $y_1^s \geq u_1$ since $d \geq 0$. This implies that $y_n^s \geq u_n$ and hence, passing to the limit with the help of Lemma 3.2, we have $\mathbf{m}(f + sd) \geq \mathbf{m}(f)$.

Let us prove the lemma (without the additional assumption). Define the sequence

$$\begin{aligned} y_n &:= S(f, y_{n-1}), \\ y_0 &:= \mathbf{m}(f + sd), \end{aligned}$$

and note that $y_1 = S(f, y_0) \leq S(f + sd, y_0) = y_0$. This implies that $\{y_n\}$ is a decreasing sequence.

Since y_0 is the minimal solution on $[\underline{u}, \bar{u}]$, $y_0 \geq \underline{u}$ and from $y_1 = S(f, y_0) \geq S(f, \underline{u}) \geq \underline{u}$ (by definition of subsolution), each $y_n \geq \underline{u}$. This implies that v_0 from assumption (8) satisfies $v_0 \leq \Phi(y_n)$ for each n , and hence is a feasible test function for the VI for y_n . Just like in the proof of Lemma 3.2, testing with v_0 , we easily obtain a uniform bound on $\{y_n\}$ in V , and therefore we have $y_n \rightarrow y$ in V for some y , which in addition to the fact that $\{y_n\}$ is decreasing, leads, via (2), to

$$y_n \rightarrow y \in \mathbf{Q}(f)$$

in exactly the same way as in the proof of Lemma 3.3.

Furthermore, we have seen that $y \in [\underline{u}, y_0] = [\underline{u}, \mathbf{m}(f + sd)] \subset [\underline{u}, \bar{u}]$. This, along with $y \in \mathbf{Q}(f)$ implies that $\mathbf{m}(f) \leq y \leq \mathbf{m}(f + sd)$. $\square$

Now, let us define (as sketched at the start of §3) a sequence starting at $\mathbf{m}(f)$ with perturbed source term as follows:

$$\begin{aligned} u_n^s &:= S(f + sd, u_{n-1}^s), \\ u_0^s &:= \mathbf{m}(f). \end{aligned} \quad (13)$$

It is worth remarking that since $\mathbf{m}(f)$ acts as a subsolution for $S(f + sd, \cdot)$, if Φ is completely continuous, $u_n^s \nearrow \mathbf{m}|_{[\mathbf{m}(f), \bar{u}]}(f + sd)$ by Lemma 3.2 (the notation $\mathbf{m}|_A(f + sd)$ refers to the minimal solution on A). But in fact, the limit is the minimal solution on the full interval $[\underline{u}, \bar{u}]$ as the next result shows (which, we note, dispenses with the need for complete continuity).

Proposition 4.4. Let $d \in V_+^*$, (8) and

$$\text{there exists } \epsilon > 0 \text{ such that } \Phi: B_\epsilon(\mathbf{m}(f)) \rightarrow V \text{ is Lipschitz with Lipschitz constant } C_\Phi < C_a/(C_a + C_b) \quad (14)$$

hold. Then $u_n^s \nearrow \mathbf{m}(f + sd)$ in V .

Proof. We have $u_n^s \rightarrow u^s$ in V for some $u^s \in \mathbf{Q}(f + sd)$ by [4, Proposition 3.7], thanks to (14) (as explained at the start of §3). Since d is non-negative, the sequence $\{u_n^s\}$ is increasing and hence $u^s \in [\mathbf{m}(f), \infty)$, but we also see that $u_1^s = S(f + sd, \mathbf{m}(f)) \leq S(f + sd, \mathbf{m}(f + sd)) = \mathbf{m}(f + sd)$ (with the inequality by Lemma 4.3) and thus $u^s \in [\mathbf{m}(f), \mathbf{m}(f + sd)] \subset [\underline{u}, \mathbf{m}(f + sd)]$. The claim follows. $\square$

In the above proposition, if $\Phi: V \rightarrow V$ is completely continuous, then (14) can be omitted. With all the preparations complete, we are ready to state the differentiability result.

Theorem 4.5 (Directional differentiability of $\mathbf{m}$). *For $f \in V^*$, let $\underline{u}$ be a subsolution for $S(f, \cdot)$ and given $d \in V_+^*$, let $\bar{u}$ satisfy (12) with $\underline{u} \leq \bar{u}$. In addition to (8) and (14), assume that*

$$\Phi: V \rightarrow V \text{ is Hadamard directionally differentiable at } \mathbf{m}(f). \quad (15)$$

Then the map $\mathbf{m}: V^ \rightarrow V$ is directionally differentiable at f in the direction d :*

$$\lim_{s \rightarrow 0^+} \frac{\mathbf{m}(f + sd) - \mathbf{m}(f)}{s} = \mathbf{m}'(f)(d).$$

Furthermore, $\mathbf{m}'(f)(d)$ satisfies the QVI

$$\begin{aligned} \alpha \in \mathcal{K}_{\mathbf{m}}(\alpha) : \langle A\alpha - d, \alpha - v \rangle &\leq 0 \quad \forall v \in \mathcal{K}_{\mathbf{m}}(\alpha), \\ \mathcal{K}_{\mathbf{m}}(\alpha) &:= \Phi'(\mathbf{m}(f))(\alpha) + \mathcal{T}_{\mathbf{K}(\mathbf{m}(f))}(\mathbf{m}(f)) \cap [f - A\mathbf{m}(f)]^\perp. \end{aligned} \quad (16)$$

Proof. The proof is as described at the start of this section. Indeed, a straightforward application of Theorem 2.1 gives the existence of $\alpha \in V$ satisfying (6) such that the limit u^s of $\{u_n^s\}$ (defined in (13)) satisfies

$$u^s = \mathbf{m}(f) + s\alpha + o(s).$$

Proposition 4.4 can be applied and it tells us that $u^s = \mathbf{m}(f + sd)$. $\square$

We reiterate that assumption (12) in the above theorem is satisfied by simply taking $\bar{u} = A^{-1}(f + d)$ as explained in Example 4.2 and hence this is, in some sense, not an assumption at all; we include it for full generality and in order to have $\mathbf{m}$ clearly defined (recall that $\mathbf{m}$ and $\mathbf{M}$ require the identification of an ordered interval to be sensible).

The QVI (16) satisfied by the derivative α is determined as the monotone limit of the sequence $\{\alpha_n\}$ (see (7)) of solutions of VIs where each α_n satisfies

$$\alpha_n \in \mathcal{K}_{\mathbf{m}}(\alpha_{n-1}) : \langle A\alpha_n - d, \alpha_n - \varphi \rangle \leq 0 \quad \forall \varphi \in \mathcal{K}_{\mathbf{m}}(\alpha_{n-1}).$$

A direct consequence of the monotonicity of $\{u_n^s\}$ allows us to conclude that $\alpha^n \nearrow \alpha$ in V .

5 The maximal solution map

This section is almost a mirror image of §4 but not entirely because our constraint set is unilateral, that is, it involves (only) the $\leq$ operator. Here, we reverse the sign of the direction term in order to enforce monotonicity of a certain sequence.

We start with $\bar{u}$ a given supersolution of $S(f, \cdot)$. Again, let us take $s \geq 0$. Observe that for any $d \in V^*$ with $d \leq 0$, $\bar{u}$ is a supersolution for $S(f + sd, \cdot)$ too: $\bar{u} \geq S(f, \bar{u}) \geq S(f + sd, \bar{u})$ by the sign on d . Akin to the previous section, we are going to make the standing assumption

$$\text{there exists } s_0 > 0 \text{ such that for all } s \leq s_0, \underline{u} \text{ is a subsolution for } S(f + sd, \cdot) \quad (17)$$

so that $\mathbf{Q}(f + sd) \cap [\underline{u}, \bar{u}]$ is non-empty and $\mathbf{M}(f + sd)$ is well defined on $[\underline{u}, \bar{u}]$.

Remark 5.1. *This assumption is more inconvenient to fulfill than the corresponding condition (12) for the minimal solution map. Nevertheless, if $\Phi(0) \geq 0$ and there exists an $r_0 \geq 0$ such that $f + r_0 d \geq 0$, then 0 is a subsolution satisfying (17) for sufficiently small s . Indeed, for $s \leq r_0$, we have $S(f + sd, 0) \geq S(f + r_0 d, 0) \geq 0$.*

Lemma 5.2. *Let $d \in -V_+^*$, and (8) hold. Then $\mathbf{M}(f) \geq \mathbf{M}(f + sd)$.*

Proof. Define

$$\begin{aligned} y_n^s &:= S(f + sd, y_{n-1}^s), \\ y_0^s &:= \bar{u}. \end{aligned}$$

Taking the sequence $\{u_n\}$ from (11), it follows that $u_1 \geq y_1^s$ and therefore $u_n \geq y_n^s$. Passing to the weak limit in this inequality by using Lemma 3.3, we see that $\mathbf{M}(f) \geq \mathbf{M}(f + sd)$ (note that (17) is needed to obtain $y_n^s \geq \underline{u}$ and hence also $\lim_n y_n^s \geq \underline{u}$). $\square$

It is not difficult to see that $\mathbf{M}(f)$ is a supersolution for $S(f + sd, \cdot)$ for non-positive d . This allows us to construct a perturbed sequence starting at $\mathbf{M}(f)$ and we obtain the next result.

Proposition 5.3. *Let $d \in -V_+^*$, (8) and*

$$\text{there exists } \epsilon > 0 \text{ such that } \Phi: B_\epsilon(\mathbf{M}(f)) \rightarrow V \text{ is Lipschitz with Lipschitz constant } C_\Phi < C_a/(C_a + C_b) \quad (18)$$

hold. With

$$\begin{aligned} u_n^s &:= S(f + sd, u_{n-1}^s), \\ u_0^s &:= \mathbf{M}(f), \end{aligned}$$

we have $u_n^s \searrow \mathbf{M}(f + sd)$ in V .

Proof. Firstly, as indicated in the proof of Proposition 4.4, (18) allows us to obtain that $u_n^s \rightarrow u^s$ strongly in V to some $u^s \in \mathbf{Q}(f + sd)$. We see that, using Lemma 5.2, $u_1^s \geq S(f + sd, M(f + sd)) = M(f + sd)$, implying $u_n^s \geq M(f + sd)$. Since $M(f)$ is a supersolution for $S(f + sd, \cdot)$, we obtain, by Lemma 3.3, (the following convergence being weak) $u_n^s \searrow u^s = M_{[\underline{u}, M(f)]}(f + sd)$. Lemma 5.2 tells us that in fact $M(f + sd)$ belongs to $[\underline{u}, M(f)]$ so we must have $u^s = M(f + sd)$ because $M(f + sd)$ is also the largest element on $[\underline{u}, M(f)]$. $\square$

Finally, with the same reasoning as in the proof of Theorem 4.5, due to Theorem 2.1 and Proposition 5.3 we can obtain the next result.

Theorem 5.4 (Directional differentiability of M). *For $f \in V_+^*$, let $\bar{u}$ be a supersolution for $S(f, \cdot)$ and given $d \in V_+^*$, let $\underline{u}$ satisfy (17) and $\underline{u} \leq \bar{u}$. In addition to (8) and (18), assume that*

$$\Phi: V \rightarrow V \text{ is Hadamard directionally differentiable at } M(f). \quad (19)$$

Then the map $M: V^ \rightarrow V$ is directionally differentiable at f in the direction d :*

$$\lim_{s \rightarrow 0^+} \frac{M(f + sd) - M(f)}{s} = M'(f)(d). \quad (20)$$

Furthermore, $M'(f)(d)$ satisfies the QVI

$$\begin{aligned} \alpha \in \mathcal{K}_M(\alpha) : \langle A\alpha - d, \alpha - v \rangle &\leq 0 \quad \forall v \in \mathcal{K}_M(\alpha), \\ \mathcal{K}_M(\alpha) &:= \Phi'(M(f))(\alpha) + \mathcal{T}_{\mathbf{K}(M(f))}(M(f)) \cap [f - AM(f)]^\perp. \end{aligned}$$

In a similar way to §4, we find that $\alpha_n \searrow \alpha$ in V where

$$\alpha_n \in \mathcal{K}_M(\alpha_{n-1}) : \langle A\alpha_n - d, \alpha_n - \varphi \rangle \leq 0 \quad \forall \varphi \in \mathcal{K}_M(\alpha_{n-1}).$$

6 Examples

In this section, we illustrate the applicability of the theory developed in this paper with some examples and applications of QVI problems to which we can apply the above results. We start in §6.1 with a simple and somewhat academic 1D example, the purpose of which is to simply demonstrate that there are indeed non-trivial QVI problems satisfying the requirements of the above theorems. In §6.2, we study an example where Φ is related to a solution mapping of a PDE. We conclude in §6.3 with a real world application involving a PDE coupled with a VI; this also serves to demonstrate that maps Φ like the simpler one in §6.2 have uses in physical models and applications.

Before we proceed, let us give some concrete examples of function spaces V and H and operators A such that all of the assumptions required (stated in the introduction) are satisfied. We will also point out the examples that lead to regular Dirichlet spaces so that the improved characterisation of the critical cone (see Remark 2.3) is available.

Examples 6.1 (Examples of the functional setting). *Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and $n \geq 1$. In the following examples, we will consider H to be an L^2 -type space and the ordering $\leq$ will be taken to be the usual one: $u \leq v$ if and only if $u \leq v$ a.e. on the set under consideration. The cone H_+ will be the set of a.e. non-negative functions in H .*

- (i) *Set $V = H_0^1(\Omega)$ and $H = L^2(\Omega)$ with the standard norms. Take A to be a second-order linear elliptic operator of the form*

$$\langle Au, v \rangle = \sum_{i,j=1}^n \int_{\Omega} a_{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} + \sum_{i=1}^n \int_{\Omega} b_i \frac{\partial u}{\partial x_i} v + \int_{\Omega} c_0 uv \quad \forall u, v \in V, \quad (21)$$

with $a_{ij} = a_{ji} \in C^{0,1}(\bar{\Omega})$, $b_i \in W^{1,\infty}(\Omega)$, $c_0 \in L_+^\infty(\Omega)$ and

$$\sum_{i,j=1}^n a_{ij} \xi_i \xi_j \geq C |\xi|^2 \quad \text{a.e.}$$

for some $C > 0$ and such that A is coercive. The prototypical case is the selection $a_{ij} \equiv \delta_{ij}$ and $b_i \equiv c_0 \equiv 0$ which leads to $A = -\Delta$ being the Dirichlet Laplacian.

- *V is a regular Dirichlet space [17, §1.2, Examples 1.2.3] with $\xi(u, v) = \int_{\Omega} \nabla u \cdot \nabla v$.*
- (ii) *Take H as above but now $V = H^1(\Omega)$. Let A be as in (21) with additionally $c_0 \geq \lambda > 0$ a.e. where λ is a constant such that A is coercive. If $a_{ij} \equiv \delta_{ij}$, $b_i \equiv 0$ and $c_0 > 0$ is a constant, we have $A = -\Delta + c_0 I$ with $-\Delta$ the Neumann Laplacian.*
- *Setting instead $H = L^2(\bar{\Omega})$ implies that the density condition in Remark 2.3 holds and gives rise to a regular Dirichlet form, see [17, §1.6, Example 1.6.1]. In the context of Remark 2.3, we made the choice $X = \bar{\Omega}$.*
- (iii) *Let $V = H^s(\Omega)$ for $s \in (0, 1)$ where $H^s(\Omega)$ is the fractional Sobolev space consisting of functions in $L^2(\Omega)$ such that the quantity*

$$\|u\|_{H^s(\Omega)}^2 := \int_{\Omega} u^2(x) \, dx + \int_{\Omega} \int_{\Omega} \frac{|u(x) - u(y)|^2}{|x - y|^{n+2s}} \, dx \, dy$$

is finite. We may take $H = L^2(\bar{\Omega})$ so that $X = \bar{\Omega}$ again. This is a regular Dirichlet space (see e.g. [15, §6.5, Example 6]) with the form

$$\xi(u, v) = \int_{\Omega} \int_{\Omega} \frac{(u(x) - u(y))(v(x) - v(y))}{|x - y|^{n+2s}} dx dy.$$

An example for A is $\langle Au, v \rangle = (u, v)_{H^s(\Omega)}$.

These are examples of possible choices of function spaces and operators A . We now come to examples of obstacle mappings Φ and the related QVI problems.

6.1 One-dimensional toy problem

Construct an increasing, smooth function $\Phi: \mathbb{R} \rightarrow \mathbb{R}$ with the following properties: (a) $\Phi(0) \geq 0$, (b) for a given sequence $\{y_j\}_{j=1}^N \subset \mathbb{R}_+$ (for some $N \in \mathbb{N}$, $N > 1$) such that $y_1 \leq y_2 \leq \dots \leq y_N$, we have the existence of some $\epsilon > 0$ such that for all $j \in \{1, \dots, N\}$, we have $\Phi|_{B_\epsilon(y_j)} \equiv y_j$.

Set $f := \max(0, Ay_1, \dots, Ay_N)$. Then $Ay_j - f \leq 0$ and if v is such that $v \leq \Phi(y_j)$ then $y_j - v = \Phi(y_j) - v \geq 0$, hence each $y_j \in \mathbf{Q}(f)$ is a solution.

Taking $d \geq 0$, it is evident that y_1 is a subsolution for $S(f, \cdot)$ and $S(f + sd, \cdot)$, and $A^{-1}(f + d)$ is a supersolution (see Example 4.2) with $m(f) = y_1$ the minimal on $[y_1, A^{-1}(f + d)]$. Clearly, all assumptions of Theorem 4.5 are satisfied and we get the directional differentiability of m in all non-negative directions.

6.2 Obstacle maps related to inverse of elliptic operators

Let $L: V \rightarrow V^*$ be a bounded, linear⁶, coercive (with coercivity constant K_a) and T-monotone operator and take a Hadamard differentiable map $g: H \rightarrow V^*$ such that

- (i) $g(0) = 0$,
- (ii) $g: B_\epsilon(0) \rightarrow V^*$ is Lipschitz for some $\epsilon > 0$ with Lipschitz constant C_g satisfying $C_g \leq K_a C_\Phi$ where C_Φ is as in (14),
- (iii) $u \leq v$ in H implies that $g(u) \leq g(v)$ (i.e., g is increasing).

Define $\Phi: H \rightarrow V$ by $\phi = \Phi(u)$ if and only if

$$L\phi = g(u).$$

That is, $\Phi(u)$ is the solution of a PDE. It is standard that this is well defined: for every $u \in H$, there exists a unique $\phi \in V$ satisfying this equation by the Lax–Milgram lemma.

If $u \leq v$, denoting $\phi_u = \Phi(u)$ and $\phi_v = \Phi(v)$, we have

$$\langle L(\phi_u) - L(\phi_v), (\phi_u - \phi_v)^+ \rangle = \langle g(u) - g(v), (\phi_u - \phi_v)^+ \rangle \leq 0$$

since g is increasing. By T-monotonicity of L , $\phi_v \leq \phi_u$, showing that Φ is an increasing map. It is also not difficult to see that $\Phi(0) = 0$.

Concerning differentiability, take $u, h \in V$. Denoting $\varphi := \Phi(u + sh) = L^{-1}g(u + sh)$ and $\phi := \Phi(u) = L^{-1}g(u)$, we see that, using the Hadamard differentiability of g ,

$$\varphi = L^{-1}(g(u) + sg'(u)(h) + o(s)) = \phi + sL^{-1}g'(u)(h) + L^{-1}o(s),$$

hence Φ is Hadamard differentiable with $\Phi'(u)(h) = L^{-1}g'(u)(h)$ given as the unique solution of the equation

$$L\delta = g'(u)(h).$$

Now, take $f \in V^*$. If $v \in V$ is such that $v \leq \Phi(0) = 0$ then $\langle A0 - f, 0 - v \rangle = \langle f, v \rangle \leq 0$, i.e., $0 \in \mathbf{Q}(f)$.

Let $d \in V^*$. From Example 4.2, $A^{-1}(f + d)$ is a supersolution with $A^{-1}(f + d) \geq 0$. It follows that $\mathbf{Q}(f) \cap [0, A^{-1}(f + d)]$ is non-empty and there exist minimal and maximal solutions on this interval. Let us point out that if in addition

$$A^{-1}f \leq L^{-1}g(A^{-1}f),$$

it is easy to see that $A^{-1}f \in \mathbf{Q}(f)$ is a solution (and not just a supersolution of $S(f, \cdot)$). This condition is satisfied if for example, $L \equiv A$ and $g(A^{-1}f) \geq f$, as can be seen via the weak comparison principle.

Now, take $u, v \in B_\epsilon(0)$ for some ϵ . Then $\Phi(u) - \Phi(v) = L^{-1}g(u) - L^{-1}g(v)$ and hence

$$\begin{aligned} \|\Phi(u) - \Phi(v)\|_V &\leq K_a^{-1} \|g(u) - g(v)\|_{V^*} \\ &\leq K_a^{-1} C_g \|u - v\|_V \end{aligned}$$

which, by assumption (ii) on g above, shows that (14) is satisfied. Therefore, we again have the directional differentiability of m for all $d \in V^*$.

⁶This can be readily generalised to nonlinear, Fréchet differentiable operators L with appropriate changes to the calculations that follow.

6.3 Application to thermoforming

We study now an application in thermoforming that we set out in [1]; the background and introductory details below are slightly revised from there but afterwards we will also improve and expand previous results and show that the minimal solution map associated to this example is differentiable.

Thermoforming is a process used to manufacture products by heating a membrane or plastic sheet to its pliable temperature and then forcing the membrane (by means of vacuum or gas pressure) onto a mould which is typically made of aluminium or an alloy; this makes the membrane deform and take on the shape of the mould. The process is applied to form structures on a variety of physical scales from car panels to microfluidic structures (e.g. channels on the range of micrometers). The importance of the applications and the necessity of precision of the details in this process has sparked research into its modelling and accurate numerical simulation, e.g., [22, 38].

Mathematically, the contact problem associated with the membrane and the mould can be described as a variational inequality problem (assuming perfect sliding of the membrane with the mould as described in [6]). However, when the heated sheet is forced into contact with the mould, a complex phenomenon takes place: in principle, the mould and the plastic sheet are at different temperatures (the mould may be cold relative to the membrane), which triggers a heat transfer process with difficult-to-predict consequences (see for example [24], e.g., it changes the polymer viscosity). In practice, the thickness of the thermoformed piece can be controlled locally by the mould structure and its initial temperature distribution [24] and the non-uniform temperature distribution of the polymer sheet has substantial influence on the results [28]. Indeed, a commonly-used material for the mould is aluminum and large heat fluctuations create a substantial difference in the size; aluminum has a relatively high thermal expansion volumetric coefficient and this implies that there is a dynamic change in the obstacle (the mould) as the polymer sheet is forced in contact with it. This suggests that the model can be formulated as a compliant obstacle-type problem (such as the one described in [29]) and hence the overall process is a QVI with underlying nonlinear PDEs determining the heat transfer and the volume change in the obstacle.

We consider this compliant obstacle behavior whilst making simplifying assumptions in order to study a tractable yet meaningful model.

6.3.1 The mathematical model

We restrict the analysis to the 1D case for the sake of simplicity. Let $\Phi_0 : [0, 1] \rightarrow \mathbb{R}^+$ be the (parametrised) mould shape that we wish to reproduce through a sheet or membrane. The membrane lies below the mould and is pushed upwards through some mechanism (as mentioned, usually through vacuum and/or air pressure) denoted as f . We make the following fundamental physical assumptions to simplify the oncoming study:

- (i) the temperature for the membrane u is constant and prescribed,
- (ii) the growth of the mould $\Phi(u)$ is affine with respect to its temperature T ,
- (iii) the temperature (denoted by T , as above) of the mould (a) is subject to diffusion, convection and boundary conditions arising from the insulated boundary and, (b) depends on the vertical distance between the mould $\Phi(u)$ and the membrane u .

Despite the fact that the thermoforming process is an evolutionary process, the setting described by the above assumptions is appropriate for one time step in the time discretisation and fits the mathematical framework of the present paper.

Define $V = H^1(0, 1)$, $H = L^2(0, 1)$ and let $A = -\Delta + I$ where $-\Delta$ is the Neumann Laplacian. The system we consider is the following:

$$u \in V : u \leq \Phi(u), \quad \langle Au - f, u - v \rangle \leq 0 \quad \forall v \in V : v \leq \Phi(u), \quad (22a)$$

$$kT - \Delta T = g(\Phi(u) - u) \quad \text{on } [0, 1], \quad (22b)$$

$$\partial_\nu T = 0 \quad \text{on } \{0, 1\}, \quad (22c)$$

$$\Phi(u) = \Phi_0 + LT \quad \text{on } [0, 1], \quad (22d)$$

where $f \in H_+$, $k > 0$ is a constant, $\Phi_0 \in V$, $L : V \rightarrow V$ is a bounded linear operator such that, with $\Omega := [0, 1]$,

for every $\Omega_0 \subset \Omega$, if $u \leq v$ a.e. on Ω_0 then $Lu \leq Lv$ a.e. on Ω_0 ,

and $g : \mathbb{R} \rightarrow \mathbb{R}$ is decreasing and C^2 with $g(0) = M > 0$ a constant, $0 \leq g \leq M$ with g' bounded⁷ with

$$g' \equiv 0 \text{ on } (1, \infty), \quad (23)$$

and g'' bounded from above. Thus when the membrane and mould are in contact or are close to each other, there is a maximum level of heat transfer onto the mould, whilst when they are sufficiently separated, there is very little or no heat exchange. An example of g to have in mind is a smoothing of the function

$$\tilde{g}(r) = \begin{cases} 1 & : \text{if } r \leq 0 \\ 1 - r & : \text{if } 0 < r < 1 \\ 0 & : \text{if } r \geq 1. \end{cases} \quad (24)$$

⁷Under these circumstances, g maps V into V [20, Theorem 1.18]

Remark 6.2. The system (22) is derived as follows. Consideration of the potential energy of the membrane will show that u solves the inequality (22a) [6] with the novel QVI nature resulting from assuming that heat transfer occurs between the membrane and mould (the membrane and the mould modify each other). If we let $\hat{T}: \Gamma \rightarrow \mathbb{R}$ be the temperature of the mould defined on the curve

$$\Gamma := \{(r, \Phi(u)(r)) : r \in [0, 1]\} \subset \mathbb{R}^2,$$

our modelling assumptions directly imply that $\hat{T}$ solves the PDE

$$\begin{aligned} k\hat{T}(x) - \Delta_\Gamma \hat{T}(x) &= g(x_2 - u(x_1)) \quad \text{for } x = (r, \Phi(u)(r)) \in \Gamma, \\ \frac{\partial \hat{T}}{\partial \nu} &= 0 \quad \text{on } \partial\Gamma, \end{aligned}$$

where $x = (x_1, x_2)$ and $-\Delta_\Gamma$ is the Laplace–Beltrami operator. Reparametrising by $T(r) = \hat{T}(r, \Phi(r))$ and simplifying the Laplace–Beltrami operator leads to (22b), see the appendix of [1] for the details.

6.3.2 Properties for the system

It can be useful to write the equation for T explicitly in terms of u by plugging (22d) into (22b):

$$\begin{aligned} kT - \Delta T &= g(LT + \Phi_0 - u) \quad \text{on } [0, 1], \\ \partial_\nu T &= 0 \quad \text{on } \{0, 1\}. \end{aligned} \tag{25}$$

According to [1, Lemma 16], for every $u \in H$, there exists a unique solution $T \in V$ to this equation. This formulation allows us to deduce the following properties for Φ .

Lemma 6.3 ([1, Lemmas 17 and 18]). *The map $\Phi: H \rightarrow V \subset H$ is increasing and $\Phi(0) \geq 0$ a.e.*

As before, we choose $\underline{u} = 0$ as subsolution and $\bar{u} = A^{-1}(f + d)$ as supersolution. Using these results along with the Birkhoff–Tartar theorem, we have the following existence result.

Theorem 6.4 ([1, Theorem 7]). *There exists a solution $(u, T, \Phi(u))$ to the system (22).*

Testing (25) with the solution T and using the fact that $g \in L^\infty(\mathbb{R})$, we have the *a priori* bound (independent of Φ_0 , L and u)

$$\|T\|_V \leq \|g\|_\infty \min(1, k)^{-1} =: C^*.$$

This next result improves [1, Lemma 19], which needed a smallness condition related to various constants appearing in the model for the conclusion to hold.

Lemma 6.5. *The map $\Phi: V \rightarrow V$ is completely continuous.*

Proof. Let $u_n \rightarrow u$ in V and denote by T_n the associated temperature, i.e.,

$$\begin{aligned} kT_n - \Delta T_n &= g(LT_n + \Phi_0 - u_n), \\ \partial_\nu T_n &= 0. \end{aligned}$$

Using the estimate derived above, we get the uniform bound

$$\|T_n\|_V \leq C^*.$$

Thus, for a subsequence that we shall relabel, $T_n \rightharpoonup T$ in V to some T . The compact embedding $V \hookrightarrow H$ implies that $T_n \rightarrow T$ in H , and by the continuity of g and L and the Dominated Convergence Theorem, $g(LT_n + \Phi_0 - u_n) \rightarrow g(LT + \Phi_0 - u)$ in H , thus we find that T satisfies the expected limiting equation (25). Since solutions of the latter equation are unique, the entire sequence $\{T_n\}$ converges, not just a subsequence. The claim follows (by definition of Φ , see (22d)). $\square$

Complete continuity of Φ is much stronger than what is necessary to enter into the framework of the theory presented in the previous sections but it is a useful property that typically holds in these types of models involving PDEs.

Regarding the differentiability of Φ and a characterisation of the derivative, we have the following, which can be proved via the implicit function theorem.

Theorem 6.6 ([1, Theorem 8]). *The map $\Phi: V \rightarrow V$ is Fréchet differentiable at any solution u . Furthermore, given $d \in V$, denoting by $\delta \in V$ the solution of*

$$\begin{aligned} (k - \Delta)\delta - g'(\Phi(u) - u)L\delta &= g'(\Phi(u) - u)d, \\ \partial_\nu \delta &= 0, \end{aligned}$$

we have $\Phi'(u)(d) = -L\delta$.

It remains for us to verify the Lipschitz condition (14).

Lemma 6.7. *Let $\Phi_0 > 1 + C_a^{-1}K\|f\|_{V^*}$ where K denotes the injectivity constant in the embedding $V \hookrightarrow L^\infty(\Omega)$. Then the condition (14) is satisfied.*

Proof. Test the inequality satisfied by $m(f)$ with 0, which is feasible since $0 \leq \Phi(0) \leq \Phi(m(f))$ with the latter inequality by the increasing property of Φ . This immediately leads to the bound

$$\|m(f)\|_V \leq C_a^{-1} \|f\|_{V^*}.$$

For some $\rho > 0$, take $v \in B_{\rho/K}(m(f))$, hence $\|v - m(f)\|_{L^\infty(\Omega)} \leq \rho$. We have, using the fact that $T \geq 0$ and the local increasing property of L ,

$$\begin{aligned} \Phi(v) - v &= \Phi_0 + LT(v) - v \\ &\geq \Phi_0 - v \\ &\geq \Phi_0 - m(f) - \rho \\ &\geq \Phi_0 - C_a^{-1} K \|f\|_{V^*} - \rho. \end{aligned}$$

It follows that if $\Phi_0 > 1 + C_a^{-1} K \|f\|_{V^*}$, we can choose ρ so that the above exceeds 1 and by the condition (23) on g' , we have

$$g'(\Phi(v) - v) = 0.$$

Then for such $v \in B_{\rho/K}(m(f))$, we see from Theorem 6.6 that $L^{-1}\Phi'(u)(d) = 0$ for all directions d , hence (14) is satisfied (via Remark 2.2). $\square$

Applying now Theorem 4.5, we obtain the directional differentiability of m at $f \in H_+$ in the direction $d \in V_+^*$.

7 Conclusion

In this paper, we have given the first directional differentiability results for minimal and maximal solution maps associated to QVIs and have demonstrated the theory with some examples. In conclusion, let us note that the formulation and analysis (such as existence and stationarity theory) of optimal control problems with these types of control-to-state maps (as discussed in §1.2) requires additional effort and care as well as further continuity results. We aim to address these issues in future work.

References

- [1] A. Alphonse, M. Hintermüller, and C. N. Rautenberg. Directional differentiability for elliptic quasi-variational inequalities of obstacle type. *Calc. Var. Partial Differential Equations*, 58(1):Art. 39, 47, 2019.
- [2] A. Alphonse, M. Hintermüller, and C. N. Rautenberg. Recent trends and views on elliptic quasi-variational inequalities. In M. Hintermüller and J. F. Rodrigues, editors, *Topics in Applied Analysis and Optimisation*, pages 1–31, Cham, 2019. Springer International Publishing.
- [3] A. Alphonse, M. Hintermüller, and C. N. Rautenberg. Stability of the solution set of quasi-variational inequalities and optimal control. *SIAM J. Control Optim.*, 58(6):3508–3532, 2020.
- [4] A. Alphonse, M. Hintermüller, and C. N. Rautenberg. Optimal control and directional differentiability for elliptic quasi-variational inequalities. *arXiv e-prints*, page arXiv:2008.13586, May 2021. Available at <https://arxiv.org/abs/2008.13586>.
- [5] A. Alphonse, C. N. Rautenberg, and J. F. Rodrigues. Analysis of a quasi-variational contact problem arising in thermoelasticity. *arXiv e-prints*. Available at <https://arxiv.org/abs/2008.00890>.
- [6] H. Andrá, M. K. Warby, and J. R. Whiteman. Contact problems of hyperelastic membranes: Existence theory. *Mathematical Methods in the Applied Sciences*, 23:865–895, 2000.
- [7] J.-P. Aubin. *Mathematical methods of game and economic theory*. Dover Publications, Inc., Mineola, NY, revised edition, 2007. With a new preface by the author.
- [8] C. Baiocchi and A. Capelo. *Variational and Quasivariational Inequalities*. Wiley-Interscience, 1984.
- [9] J. W. Barrett and L. Prigozhin. A quasi-variational inequality problem in superconductivity. *Math. Models Methods Appl. Sci.*, 20(5):679–706, 2010.
- [10] J. W. Barrett and L. Prigozhin. A quasi-variational inequality problem arising in the modeling of growing sandpiles. *ESAIM Math. Model. Numer. Anal.*, 47(4):1133–1165, 2013.
- [11] J. W. Barrett and L. Prigozhin. Lakes and rivers in the landscape: a quasi-variational inequality approach. *Interfaces Free Bound.*, 16(2):269–296, 2014.
- [12] A. Bensoussan, M. Goursat, and J.-L. Lions. Contrôle impulsif et inéquations quasi-variationnelles stationnaires. *C. R. Acad. Sci. Paris Sér. A-B*, 276:A1279–A1284, 1973.
- [13] G. Birkhoff. *Lattice theory*, volume 25 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, R.I., third edition, 1979.
- [14] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

- [15] Z.-Q. Chen and M. Fukushima. *Symmetric Markov processes, time change, and boundary theory*, volume 35 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2012.
- [16] F. Facchinei and C. Kanzow. Generalized nash equilibrium problems. *4OR*, 5(3):173–210, Sep 2007.
- [17] M. Fukushima, Y. Oshima, and M. Takeda. *Dirichlet forms and symmetric Markov processes*, volume 19 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, extended edition, 2011.
- [18] A. Haraux. 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, 1977.
- [19] P. T. Harker. Generalized nash games and quasi-variational inequalities. *European Journal of Operational Research*, 54(1):81 – 94, 1991.
- [20] J. Heinonen, T. Kilpeläinen, and O. Martio. *Nonlinear potential theory of degenerate elliptic equations*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 1993. Oxford Science Publications.
- [21] M. Hintermüller and T. Surowiec. A bundle-free implicit programming approach for a class of elliptic MPECs in function space. *Math. Program.*, 160(1-2, Ser. A):271–305, 2016.
- [22] W.-G. Jiang, M. K. Warby, J. R. Whiteman, S. Abbot, W. Shorter, P. Warwick, T. Wright, A. Munro, and B. Munro. Finite element modelling of high air pressure forming processes for polymer sheets. *Computational Mechanics*, 31:163–172, 2001.
- [23] M. Kunze and J. F. Rodrigues. An elliptic quasi-variational inequality with gradient constraints and some of its applications. *Math. Methods Appl. Sci.*, 23(10):897–908, 2000.
- [24] J. K. Lee, T. L. Virkler, and C. E. Scott. Effects of rheological properties and processing parameters on abs thermoforming. *Polymer Engineering and Science*, 41(2):240–261, 2001.
- [25] J.-L. Lions. Sur le contrôle optimal des systemes distribuées. *Enseigne*, 19:125–166, 1973.
- [26] F. Mignot. Contrôle dans les inéquations variationelles elliptiques. *J. Functional Analysis*, 22(2):130–185, 1976.
- [27] U. Mosco. Implicit variational problems and quasi variational inequalities. In *Nonlinear Operators and the Calculus of Variations*, pages 83–156. 1976.
- [28] G. J. Nam, K. H. Ahn, and J. W. Lee. Three-dimensional simulation of thermoforming process and its comparison with experiments. *Polymer Engineering and Science*, 40(10):2232–2240, 2000.
- [29] J. Outrata, M. Kocvara, and J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, 1998.
- [30] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized nash equilibria, and multi-leader-follower games. *Computational Management Science*, 2(1):21–56, Jan 2005.
- [31] L. Prigozhin. Sandpiles and river networks: extended systems with non-local interactions. *Phys. Rev. E*, 49:1161–1167, 1994.
- [32] L. Prigozhin. On the Bean critical-state model in superconductivity. *European Journal of Applied Mathematics*, 7:237–247, 1996.
- [33] L. Prigozhin. Variational model of sandpile growth. *European J. Appl. Math.*, 7(3):225–235, 1996.
- [34] J. F. Rodrigues. *Obstacle problems in mathematical physics*, volume 134 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1987. Notas de Matemática [Mathematical Notes], 114.
- [35] J. F. Rodrigues and L. Santos. A parabolic quasi-variational inequality arising in a superconductivity model. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 29(1):153–169, 2000.
- [36] L. Tartar. Inéquations quasi variationnelles abstraites. *C. R. Ac. Sci. Paris*, 178:1193–1196, 1974.
- [37] G. Wachsmuth. A guided tour of polyhedral sets: basic properties, new results on intersections, and applications. *J. Convex Anal.*, 26(1):153–188, 2019.
- [38] M. K. Warby, J. R. Whiteman, W.-G. Jiang, P. Warwick, and W. T. Finite element simulation of thermoforming processes for polymer sheets. *Mathematics and Computes in Simulation*, 61:209–218, 2003.