An inexact bundle method and subgradient computations for optimal control of deterministic and stochastic obstacle problems

Lukas Hertlein, Anne-Therese Rauls, Michael Ulbrich and Stefan Ulbrich

Abstract. The aim of this work is to develop an inexact bundle method for nonsmooth nonconvex minimization in Hilbert spaces and to investigate its application to optimal control problems with deterministic or stochastic obstacle problems as constraints. A central requirement is that (approximate) subgradients can be obtained at given points. The second part of the paper thus studies in detail how subgradients can be obtained for optimal control problems governed by (stochastic) obstacle problems.

Mathematics Subject Classification (2000). Primary 26A24, 49J40, 49J52, 65K05; Secondary 49K20, 90C56.

Keywords. Optimal control, nonsmooth optimization, generalized derivatives, bundle method, obstacle problem, stochastic obstacle problem, variational inequalities.

1. Introduction

We consider optimal control problems governed by obstacle problems of the form

$$\text{Find } y \in K_\psi, \quad \langle Ly - F(u), z - y\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \geq 0 \quad \text{for all } z \in K_\psi \tag{VI}$$

as well as by stochastic versions of (VI), see (**VI**$_\text{s}$) below. Here, $\Omega \subseteq \mathbb{R}^d$ is a bounded open domain and K_ψ denotes the closed convex set $K_\psi := \{z \in H_0^1(\Omega) : z \geq \psi \text{ q.e. on } \Omega\}$, where ψ is a given quasi upper-semicontinuous obstacle such that $K_\psi \neq \emptyset$. Additional regularity assumptions on ψ are stated if necessary. The abbreviation "q.e." stands for "quasi-everywhere" and will be used frequently in this paper. It describes that the respective property holds everywhere except on a subset of Ω which has capacity zero. For the notion of capacity and the corresponding definitions, we refer the reader to, e.g., [1, 2, 8, 21]. Furthermore, $L \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$ is a coercive and T-monotone operator. The operator $F\colon U \to H^{-1}(\Omega)$ is assumed to be Lipschitz continuous on bounded sets, continuously differentiable and monotone, defined on a partially ordered Banach space U. The precise assumptions on the space U are given in Section 5, prototypes include

$U = L^2(\Omega)$, $U = H^{-1}(\Omega)$, or $U = \mathbb{R}^n$. Due to the operator F and the assumptions on U, the variational inequality (VI) represents a general class of obstacle problems. It is well known that for each $u \in U$ the variational inequality (VI) has a unique solution. We denote the solution operator by $S_F\colon U \to H_0^1(\Omega)$. If F is the identity map on $H^{-1}(\Omega)$, we omit the subscript and write $S\colon H^{-1}(\Omega) \to H_0^1(\Omega)$.

We consider the following optimal control problem governed by (VI):

$$\min_{u \in U_{\text{ad}}} J(S_\iota(u)) + \tfrac{\alpha}{2}\|u\|_U^2, \tag{P}$$

where $U_{\text{ad}} \subset U$ is a closed convex subset of the Hilbert space U, $J : H_0^1(\Omega) \to \mathbb{R}$ is the objective function and $\iota \in \mathcal{L}(U, H^{-1}(\Omega))$ is a compact and injective operator. Here, S_ι denotes the solution operator of (VI) when choosing $F = \iota$.

The nonconvexity and nondifferentiability of the solution operator S_ι requires the application of nonsmooth optimization methods. An alternative is to view the VI as a constraint, which results in a mathematical progam with equilibrium constraints (MPEC). Most methods for MPECs use regularization or smoothing; we refer to [39] for a survey of numerical methods for the optimal control of elliptic variational inequalities.

Here, we propose to use a variant of the bundle method developed in [22] which is tailored for this use. This method is posed in an appropriate function space setting and can handle inexact function values, inexact subgradients and inexact solutions of the bundle subproblem. We extend the method of [22], allowing for quite general sets of approximate subgradients. Furthermore, we provide a global convergence result for general locally Lipschitz functions, provided there exists a subsequence of iterations in which the new model is sufficiently much improved over the old model (cf. Theorem 2.6). To ensure this, one usually requires approximate convexity [22, 31] or semismoothness [28]. Our generalization is motivated by the fact that, in general, there can exist points where these properties do not hold. Already in finite dimensions, there is not much literature on bundle methods for nonconvex optimization with inexact function values and subgradients [19, 27, 31]. Our work in this paper and in [22] is inspired by [31] and seems to be the only inexact bundle method for infinite dimensional nonconvex problems.

The bundle method requires an approximate subgradient of the reduced objective function in each iteration. Therefore, we derive a formula for an element of a generalized differential for the solution operator of the obstacle problem in each point in U, from which a Clarke subgradient for the reduced objective function can be extracted. The generalized derivative that we construct for the solution operator S_F of (VI) in an arbitrary $u \in U$ is the operator $\Sigma_F(u;\cdot) \in \mathcal{L}(U, H_0^1(\Omega))$, where $\Sigma_F(u;h) = \eta$ solves

$$\text{Find } \eta \in H_0^1(I(u)),\ \langle L\eta - F'(u;h), z\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} = 0 \ \text{ for all } z \in H_0^1(I(u)). \tag{1.1}$$

Here, $I(u) := \{\omega \in \Omega : S_F(u)(\omega) > \psi(\omega)\}$ is the inactive set, which is a quasi-open subset of Ω and $H_0^1(I(u)) = \{v \in H_0^1(\Omega) : v = 0 \text{ q.e. outside } I(u)\}$ is a closed subspace of $H_0^1(\Omega)$. The variational equation (1.1) is also a characterization of the directional derivative in points where S_F is Gâteaux differentiable. In such points, (1.1) is obtained from the variational inequality for the directional derivative of S_F established by Mignot [29]. Since the generalized differentials we consider for S_F contain limits of Gâteaux derivatives w.r.t. certain topologies in points $(u_n)_{n\in\mathbb{N}}$ converging to u, we derive the generalized derivative by pursuing a convergence analysis for such problems considering appropriate sequences $(u_n)_{n\in\mathbb{N}}$. We are not aware of any work that establishes generalized derivatives for the solution operator of (VI) in infinite dimensions apart from our presentation in this paper and in [35]. For the case that F is the identity mapping on $H^{-1}(\Omega)$, a characterization of the entire generalized differential is possible. We review the results of [36] for this problem.

We also are interested in the optimal control of the stochastic obstacle problem. Let $(\Xi, \mathcal{A}, P)$ be a probability space and denote by $\mathbf{Y} := L^2(\Xi, H_0^1(\Omega))$ the Bochner space of square integrable functions with values in $H_0^1(\Omega)$ (cf. [24, Def. 1.2.15]). The stochastic obstacle problem (**VI**$_\text{s}$) is given by the variational inequality

$$\text{Find } \mathbf{y} \in \mathbf{K}_{\boldsymbol{\psi}}, \quad \langle \mathbf{L}\mathbf{y} - \mathbf{b}, \mathbf{z} - \mathbf{y} \rangle_{\mathbf{Y}^*,\mathbf{Y}} \geq 0 \quad \text{for all } \mathbf{z} \in \mathbf{K}_{\boldsymbol{\psi}} \tag{$\mathbf{VI}_\text{s}$}$$

where $\mathbf{L} \in \mathcal{L}(\mathbf{Y}, \mathbf{Y}^*)$, $\mathbf{b} \in \mathbf{Y}^*$, $\boldsymbol{\psi} \in \bar{\mathbf{Y}} := L^2(\Xi, H^1(\Omega))$ and

$$\mathbf{K}_{\boldsymbol{\psi}} := \{\mathbf{y} \in \mathbf{Y} : \mathbf{y}(\xi) \in K_{\psi_\xi} \text{ for } P\text{-almost all } (P\text{-a.a.})\ \xi \in \Xi\}. \tag{1.2}$$

For the rest of this paper, bold notation refers to the variables in the stochastic setting, whereas non-bold variables refer to the deterministic setting. Under suitable assumptions on the data, cf. Section 7, the Lions-Stampacchia theorem [26, Thm. 2.1] implies that the stochastic obstacle problem admits a unique solution and the solution operator $\mathbf{S} : \mathbf{Y}^* \to \mathbf{Y}$ is Lipschitz continuous (cf. Theorem 7.3). For P-a.e. $\xi \in \Xi$, this defines the operators $S_\xi : Z = H^{-1}(\Omega) \to H_0^1(\Omega)$ via $S_\xi(z) := \mathbf{S}(\boldsymbol{\iota} z)(\xi)$, where $(\boldsymbol{\iota} z)(\xi) := z$, $z \in H^{-1}(\Omega)$. We study the following class of optimal control problems for the stochastic obstacle problem:

$$\min_{u \in U_{\text{ad}}} \mathbb{E}\left[J_\xi(S_\xi(\iota u))\right] + \tfrac{\alpha}{2}\|u\|_U^2, \tag{$\mathbf{P}_\text{s}$}$$

where $J_\xi : H_0^1(\Omega) \to \mathbb{R}$ is the parametric objective function such that $\xi \mapsto J_\xi(S_\xi(z))$ is integrable for all $z \in H^{-1}(\Omega)$, $\mathbb{E}\left[J_\xi(S_\xi(\cdot))\right]$ is locally Lipschitz and $\mathbb{E}$ denotes the expectation with respect to ξ. The goal is to find a Clarke-stationary point, i.e. a point $\bar{u} \in U$ which satisfies

$$0 \in \partial_C(\mathbb{E}\left[J_\xi(S_\xi(\iota(\cdot)))\right])(\bar{u}) + \alpha\bar{u} + N_{U_{\text{ad}}}(\bar{u}), \tag{1.3}$$

where ∂_C denotes Clarke's subdifferential. The chain rule [11, Thm. 2.3.10] implies that

$$\partial_C(\mathbb{E}\left[J_\xi(S_\xi(\iota(\cdot)))\right])(u) \subset \iota^*\partial_C(\mathbb{E}\left[J_\xi(S_\xi(\cdot))\right])(\iota u) \quad \text{for } P\text{-a.e. } \xi \in \Xi \text{ and all } u \in U.$$

If $J_\xi \circ S_\xi$ or $-J_\xi \circ S_\xi$ is regular at ιu in the sense of Clarke (cf. [11, Def. 2.3.4]), then equality holds at this point. Under suitable assumptions, [11, Thm. 2.7.2] implies

$$\partial_C(\mathbb{E}\left[J_\xi(S_\xi(\cdot))\right])(z) \subset \mathbb{E}\left[\partial_C(J_\xi(S_\xi(\cdot)))(z)\right] \qquad \text{for all } z \in Z$$

with equality if $J_\xi \circ S_\xi$ or $-J_\xi \circ S_\xi$ is regular at z for each $\xi \in \Xi$. Here, the set $\mathbb{E}\left[\partial_C(J_\xi(S_\xi(\cdot)))(z)\right] \subset Z^*$ is defined as

$$\{\mathbb{E}\left[g(\xi)\right] : g \in L^1(\Xi, Z^*) \text{ is a measurable selection of } \partial_C(J_\xi(S_\xi(\cdot)))(z)\}. \tag{1.4}$$

This formula allows to reuse the subgradients (1.1) of the deterministic problem. However, the reduced objective function might not be regular at all admissible points. In this case, the available calculus rules for the Clarke subdifferential, which often take the form of inclusions, make it difficult to calculate the subdifferential $\partial_C(\mathbb{E}\,[J_\xi(S_\xi(\iota(\cdot)))])(\bar{u})$. Thus, we search for weak stationary points (cf. [43]), i.e. points $\bar{u} \in U$ which fulfill

$$0 \in \iota^*\mathbb{E}\,[\partial_C(J_\xi(S_\xi(\cdot)))(\iota\bar{u})] + \alpha\bar{u} + N_{U_{\text{ad}}}(\bar{u}). \tag{1.5}$$

However, under additional assumptions on the regularity of the data, in Section 7.4, we give a formula for exact subgradients $g \in \partial_C(\mathbb{E}\,[J_\xi(S_\xi(\iota(\cdot)))])(\bar{u})$.

The rest of the paper is organized as follows: In Section 2 we present a variant of the bundle method of [22] to solve both problems (P) and (**P**$_\text{s}$). In Section 3, we introduce sets of generalized derivatives that will be used in this article for operators between infinite dimensional spaces. Section 4 deals with the obstacle problem (VI) and its properties, in particular, properties concerning monotonicity and differentiability. We derive a formula for a generalized derivative for the solution operator of the obstacle problem in Section 5. In Section 6, characterizations of the entire generalized differentials are established for an easier instance of the obstacle problem. In Section 7 we discuss the stochastic obstacle problem and derive both formulas for exact subgradients $g \in \partial_C(\mathbb{E}\,[J_\xi(S_\xi(\iota(\cdot)))])(\bar{u})$ as well as conditions under which the weak subgradients $g \in \iota^*\mathbb{E}\,[\partial_C(J_\xi(S_\xi(\cdot)))(\iota\bar{u})]$ can be used in the bundle method.

2. Inexact bundle method

Since the optimal control problems (P) and (**P**$_\text{s}$) for the deterministic and stochastic obstacle problem are nonsmooth, nonconvex optimization problems in Hilbert spaces, we employ a tailored bundle method to solve them. We adopt the approach of [22], which itself draws from ideas in [31], to the given setting. In particular, we allow for more general choices of approximate subgradients and we outline a convergence theory for functions which are not approximately convex. Our problem setting is as follows:

$$\min_{u\in U}\; f(u) + w(u) \quad \text{s.t.} \quad u \in \mathcal{F},$$

where $f(u)$ corresponds to the cost term involving the state and $w(u)$ to the regularization term. The feasible set $\mathcal{F} \subset U$ is nonempty, closed, convex, and U is a Hilbert space. The function $f : \mathcal{F}_U \to \mathbb{R}$, $\mathcal{F}_U \supset \mathcal{F}$ convex and open in U, has the form $f = p \circ \iota$. Here $\iota \in \mathcal{L}(U, Z)$ is a compact and injective operator into the Hilbert space Z and $p : \mathcal{F}_Z \to \mathbb{R}$ is Lipschitz on bounded sets with $\mathcal{F}_Z \subset Z$ convex and open, $\iota(\mathcal{F}_U) \subset \mathcal{F}_Z$. Further, let $w : \mathcal{F}_U \to \mathbb{R}$ be continuously differentiable, Lipschitz on bounded sets, and μ-strongly convex, $\mu > 0$, i.e., for all $u \in U$ with $w(u) < \infty$ there holds

$$w(u + s) - w(u) \geq \langle w'(u), s\rangle_{U^*,U} + \tfrac{\mu}{2}\|s\|_U^2 \quad \text{for all } s \in U.$$

Note that this implies that w is also weakly sequentially lower semicontinuous.

This setting is applicable to a quite comprehensive class of optimal control problems, in particular, it includes both optimal control problems (P) and (**P**$_\text{s}$) by setting $w := \frac{\alpha}{2}\|\cdot\|_U$, $\mathcal{F} := U_{\text{ad}}$, $Z := H^{-1}(\Omega)$, $p := J(S(\cdot))$, $f := J(S_\iota(\cdot))$, or for the stochastic problem $p := \mathbb{E}\,[J_\xi(S_\xi(\cdot))]$, $f := \mathbb{E}\,[J_\xi(S_\xi(\iota(\cdot)))]$.

To find stationary points, bundle methods use subgradient information to build a local model of the nonsmooth part p around the current iterate ιu. Usually, a subgradient

g at a point $\iota u \in Z$ is an element of a subdifferential $G(\iota u) \subset Z^*$ such as Clarke's subdifferential $G(\iota u) = \partial_C p(\iota u)$ or the convex subdifferential $G(\iota u) = \partial p(\iota u)$ if p is convex. However, in certain situations it might not be possible to calculate such an element. Therefore, we pose minimal requirements that a multifunction $G : Z \rightrightarrows Z^*$ has to fulfill for being a suitable approximate subdifferential (cf. Assumption 2.1). The multifunction G then also appears in the stationarity condition of our convergence result.

Bundle methods often require a certain regularity of the objective function f beyond Lipschitz continuity. For example, in [28] the objective function is assumed to be semismooth while in [22, 31] approximate convexity [13] is required (being related to the lower C^1 property). These assumptions are needed to ensure that *all possible subgradients* in a neighborhood of the serious iterate improve the quality of the local model. However, for convergence of the algorithm, it is sufficient that the *computed subgradients* improve the local model. We thus introduce a measure for the quality of the local model, cf. (2.13). Depending on this measure we prove convergence to approximate stationary points in Theorem 2.6. This concept also justifies why the bundle method often returns good results, even when applied to problems which do not satisfy regularity beyond Lipschitz continuity. To increase the flexibility of the algorithm, we also allow for subgradients to be drawn at arbitrary points in a neighborhood of the trial iterate. This implies that there always exists a model with sufficiently high quality to guarantee convergence to stationary points (cf. Remark 2.7).

The general procedure of the bundle method is as follows: In outer iteration j and inner iteration k, a finite set $\mathcal{M}_{j,k}$ of affine linear functions, called cutting planes, is selected. The convex function $\phi_{j,k} := \max\{m(\cdot) : m \in \mathcal{M}_{j,k}\}$ is chosen as the local model of f at the serious iterate u_j. The bundle method subproblem is given by

$$\min_{y \in \mathcal{F}} \phi_{j,k}(y) + w(y) + \tfrac{1}{2}\langle (Q_j + \tau_{j,k} E)\iota(y - u_j), \iota(y - u_j)\rangle_{Z^*, Z}.$$

Here, $\tau_{j,k} > 0$ is the proximity parameter, $Q_j \in \mathcal{L}(Z, Z^*)$ may represent curvature information of p at u_j and $E \in \mathcal{L}(Z, Z^*)$ denotes the Riesz map. $\tau_{j,k}$ and Q_j are chosen such that the third term in the cost function of the bundle subproblem is strictly convex w.r.t. y, cf. Section 2.3. The unique minimizer of the subproblem $y_{j,k}$ is called *inner iterate*. Often it is difficult or impossible to calculate an exact solution of the bundle method subproblem. Therefore, we introduce the *trial iterate* $\tilde{y}_{j,k}$ as an approximation of $y_{j,k}$. If this trial iterate $\tilde{y}_{j,k}$ fulfills a certain decrease condition, it is accepted as the new serious iterate u_{j+1} and the inner loop is terminated. Otherwise, a new cutting plane is selected which enriches the old model. If the new model is not sufficiently improved, the proximity parameter is increased to gather more cutting plane information close to the serious iterate u_j. Then the next inner iteration is started.

When the j index is clear from the context, we often drop this index and refer to the quantities introduced above by u, ϕ_k, $\mathcal{M}_k$, Q, τ_k, y_k and $\tilde{y}_k$, respectively.

2.1. Trial iterates, function values and subgradients

Typically, bundle methods use function values and subgradients (or approximations thereof) to build a model of the objective function. In this paper, we work with a general concept of approximate function values and subgradients. Given a point $y_k \in \mathcal{F}$, we need to find a point $\tilde{y}_k$, called *trial iterate*, in a neighborhood of y_k at which we can compute a function value approximation and an approximate subgradient. The trial iterate $\tilde{y}_k$ has to satisfy

$$\tilde{y}_k \in \bar{B}_U(y_k, R) \cap \mathcal{F} \quad \text{and} \quad \|\iota(\tilde{y}_k - y_k)\|_Z \le \min\{M\|\iota(y_k - u)\|_Z, a_k\}, \tag{2.1}$$

where $R, M \ge 0$ are fixed constants and $(a_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ is a forcing sequence such that $a_k \to 0$ as $k \to \infty$. Furthermore, $\tilde{y}_k$ needs to achieve at least a fraction $0 < \theta < 1$ of the model reduction provided by y_k:

$$\Phi_k(u) - \Phi_k(\tilde{y}_k) \ge \theta\left(\Phi_k(u) - \Phi_k(y_k)\right). \tag{2.2}$$

A *function value approximation* $f_{\tilde{y}_k} \in \mathbb{R}$ of $f(\tilde{y}_k)$ is assumed to fulfill

$$|f_{\tilde{y}_k} - f(\tilde{y}_k)| \le \Delta, \quad \text{where } \Delta > 0 \text{ is a constant.} \tag{2.3}$$

Similar to the trial iterate we introduce the point $v_k \in \bar{B}_U(y_k, \hat{R}) \cap \mathcal{F}$, $\hat{R} \ge 0$, called *subgradient base point*, at which approximate subgradients are drawn. The enlargement of the set of points at which subgradients can be obtained might be helpful to find new subgradients which improve the local model. However, although it is possible to draw subgradients at points v_k, which can be far away from the trial iterate $\tilde{y}_k$, these subgradients might not be useful. See also Remark 2.7 for a discussion on this topic. Denote by $\mathcal{V} \subset U$ the set of all subgradient base points. We define an *approximate subgradient* of the function p at the point $v \in \hat{\mathcal{V}} := \operatorname{cl}\iota(\mathcal{V})$ as an element $\tilde{g} \in G(v)$, where G fulfills:

Assumption 2.1. The multifunction $G : \hat{\mathcal{V}} \rightrightarrows Z^*$ has the following properties:
1. For all $v \in \hat{\mathcal{V}}$, the image $G(v)$ is nonempty and convex.
2. For all bounded sets $B \subset Z$, the set $G(B \cap \hat{\mathcal{V}}) := \cup_{v\in B\cap\hat{\mathcal{V}}} G(v)$ is bounded in Z^*.
3. G has a weakly closed graph, i.e., for all sequences $(v_n)_{n\in\mathbb{N}} \subset \hat{\mathcal{V}}$ and $(g_n)_{n\in\mathbb{N}} \subset Z^*$ such that $v_n \to \bar{v}$ in Z, $g_n \rightharpoonup g$ in Z^* and $g_n \in G(v_n)$ $\forall n \in \mathbb{N}$, it holds $g \in G(\bar{v})$.

These are exactly the requirements on the subgradients needed to prove convergence of the bundle algorithm. In Section 7.2, we show that (1.4) fulfills this assumption.

Remark 2.2. For a function $p : Z \to \mathbb{R}$ that is Lipschitz on bounded sets, Clarke's differential $\partial_C p : Z \rightrightarrows Z^*$ satisfies Assumption 2.1. This follows from [11, Prop. 2.1.2 and Prop. 2.1.5]. In [22], $G = \partial_C p + C$ is used, where $C \subset Z^*$ is a closed convex set with $0 \in C$.

2.2. The cutting plane model

For $u, \tilde{y}, v \in U$ and $\tilde{g} \in Z^*$ define the *downshift* $s_{\tilde{y},v,\tilde{g},u} \in \mathbb{R}$, the *tangent* $t_{\tilde{y},\tilde{g},u} : U \to \mathbb{R}$, and the *downshifted tangent* $m_{\tilde{y},v,\tilde{g}}(\cdot, u) : U \to \mathbb{R}$ by

$$\begin{aligned} s_{\tilde{y},v,\tilde{g},u} &:= [f_{\tilde{y}} + \langle \tilde{g}, \iota(u - \tilde{y})\rangle_{Z^*,Z} - f_u]_+ + c\|\iota(v - u)\|_Z^2, \\ t_{\tilde{y},\tilde{g},u}(\cdot) &:= f_{\tilde{y}} + \langle \tilde{g}, \iota(\cdot - \tilde{y})\rangle_{Z^*,Z}, \quad m_{\tilde{y},v,\tilde{g}}(\cdot, u) := t_{\tilde{y},\tilde{g},u}(\cdot) - s_{\tilde{y},v,\tilde{g},u}. \end{aligned} \tag{2.4}$$

Here, the downshift parameter $c > 0$ is fixed. At the serious iterate u we compute the *exactness subgradient* $\tilde{g}_0 \in G(\iota(u))$ and define the *exactness plane* $m_0(\cdot, u) : U \to \mathbb{R}$ by

$$m_0(\cdot, u) := m_{u,u,\tilde{g}_0}(\cdot, u) = f_u + \langle \tilde{g}_0, \iota(\cdot - u)\rangle_{Z^*,Z}.$$

Let $\mathcal{B}_k$ denote the set of all *bundle information of previous iterations* (including all information in previous outer iterations), i.e. all triples of the form $(\tilde{y}_k, v_k, \tilde{g}_k)$ where $\tilde{y}_k$, v_k and $\tilde{g}_k$ are the trial iterate, the base point and the subgradient of iteration k. Let $\mathcal{D}_k$ denote the set of *previous downshifted tangents*:

$$\mathcal{D}_k := \{m_{\tilde{y},v,\tilde{g}}(\cdot, u) : (\tilde{y}, v, \tilde{g}) \in \mathcal{B}_k\}. \tag{2.5}$$

We choose a finite subset $\mathcal{M}_k$ of co $\mathcal{D}_k$ (co $=$ convex hull) to build the *cutting plane model* $\phi_k : U \to \mathbb{R}$ by

$$\phi_k(y) := \max\{m(y) : m \in \mathcal{M}_k\}.$$

Choosing $\mathcal{M}_{k+1} = \{m_\nu(\cdot, u), \nu = 0, \dots, k\}$ yields the full model

$$\phi_{k+1}^{\text{full}} := \max\{m_\nu(\cdot, u),\ \nu = 0, \dots, k\}.$$

However, large k might lead to an expensive cutting plane model. Therefore, we allow $\mathcal{M}_k \subset$ co $\mathcal{D}_k$ to be chosen according to Assumption 2.3 below.

2.3. Proximity control

If there is curvature information of $p : Z \to \mathbb{R}$ around $\iota(u)$ available we want to incorporate this into the model. Fix the constants $0 < \underline{q} < \bar{q}$ and denote by $E \in \mathcal{L}(Z, Z^*)$, $Ev = (v, \cdot)_Z$, the Riesz map. We assume that $Q \in \mathcal{L}(Z, Z^*)$ and $q \in (\underline{q}, \bar{q})$ are chosen such that

$$\langle (Q + qE)v, v\rangle_{Z^*,Z} \geq \underline{q}\|v\|_Z^2 \quad \text{for all } v \in Z \qquad \text{and} \qquad \|Q\|_{\mathcal{L}(Z,Z^*)} \leq \bar{q}, \tag{2.6}$$

and that Q is symmetric, i.e., $\langle Qx, y\rangle_{Z^*,Z} = \langle Qy, x\rangle_{Z^*,Z}$ for all $x, y \in Z$. For any *proximity parameter* $\tau \geq q$, the positive definite symmetric bilinear form $\langle (Q + \tau E)\cdot, \cdot\rangle_{Z^*,Z}$ defines a norm on Z via $\|\cdot\|_{Q+\tau E}^2 := \langle (Q + \tau E)\cdot, \cdot\rangle_{Z^*,Z}$.

2.4. The subproblem of the bundle method

The *subproblem of the bundle method* is given by

$$\min_{y \in \mathcal{F}} \Psi_k(y) := \phi_k(y) + w(y) + \tfrac{1}{2}\|\iota(y - u)\|_{Q+\tau_k E}^2. \tag{2.7}$$

Since Ψ_k is strongly convex on $\mathcal{F}$, this problem has a unique minimum $y_k \in \mathcal{F}$ which is called *inner iterate*. Denote the indicator function of $\mathcal{F}$ by $\delta_\mathcal{F} : U \to \mathbb{R} \cup \{\infty\}$. We define the *local model* $\Phi_k : U \to \mathbb{R} \cup \{\infty\}$ via $\Phi_k := \phi_k + w + \delta_\mathcal{F}$. The sum rule of convex analysis [5, Cor. 16.50] can be applied and yields $\partial\Phi_k = \partial\phi_k + w' + N_\mathcal{F}$, where $N_\mathcal{F} = \partial\delta_\mathcal{F}$ is the normal cone of $\mathcal{F}$ and ∂ denotes the convex subdifferential. The fact that y_k minimizes the subproblem of the bundle method can equivalently be expressed as

$$\begin{aligned} 0 &\in \partial(\Phi_k + \tfrac{1}{2}\|\iota(\cdot - u)\|_{Q+\tau_k E}^2)(y_k) \\ &= \partial\phi_k(y_k) + w'(y_k) + N_\mathcal{F}(y_k) + \iota^*(Q + \tau_k E)\iota(y_k - u). \end{aligned}$$

Therefore there exist elements $g_k^* \in \partial\phi_k(y_k)$ and $n_k \in N_\mathcal{F}(y_k)$ such that

$$e_k := \iota^*(Q + \tau_k E)\iota(u - y_k) = g_k^* + w'(y_k) + n_k \in \partial\Phi_k(y_k). \tag{2.8}$$

For $m \in \mathcal{M}_k$ denote by $g_m := m'(0) \in U^*$ the derivative of the affine linear function $m : U \to \mathbb{R}$. As the set $\mathcal{M}_k$ is finite, by [11, Prop. 2.3.12] it holds for all $y \in U$ that

$$\partial\phi_k(y) = \text{co}\,\{g_m : m \in \mathcal{M}_k,\ m(y) = \phi_k(y)\}. \tag{2.9}$$

Since $g_k^* \in \partial\phi_k(y_k)$, there exist numbers $\lambda_m \geq 0$ with $\sum_{m\in\mathcal{M}_k} \lambda_m = 1$ and $g_k^* = \sum_{m\in\mathcal{M}_k} \lambda_m g_m$. We define the *aggregate cutting plane* $m_k^* \in \text{co}\,\mathcal{D}_k$ by

$$m_k^*(\cdot, u) := \sum_{m\in\mathcal{M}_k} \lambda_m m(\cdot). \tag{2.10}$$

For the convergence analysis we only require the following properties of the models ϕ_k:

Assumption 2.3. For each $k \geq 0$ the set $\mathcal{M}_{k+1} \subset \text{co}\,\mathcal{D}_{k+1}$ is chosen such that

$$\text{a) } m_0(\cdot, u) \leq \phi_{k+1}(\cdot), \quad \text{b) } m_k^*(\cdot, u) \leq \phi_{k+1}(\cdot), \tag{2.11}$$

where we set $m_0^*(\cdot, u) := m_0(\cdot, u)$.

In Algorithm 1, the inexact bundle method is presented.

2.5. Global convergence result

Definition 2.4. A point $\bar{u} \in U$ is called ϵ-stationary, $\epsilon \geq 0$, if $0 \in w'(\bar{u}) + N_{\mathcal{F}}(\bar{u}) + \iota^*(G(\iota(\bar{u})) + \bar{B}_{Z^*}(0, \epsilon))$. A point which is 0-stationary is called stationary.

Remark 2.5. If $G = \partial_C p$ is the Clarke subdifferential and p or $-p$ is regular at $\iota\bar{u}$, the chain rule [11, Thm. 2.3.10] implies that $\partial_C f(\bar{u}) = \iota^* \partial_C p(\iota\bar{u})$ and stationarity is equivalent to $0 \in \partial_C f(\bar{u}) + w'(\bar{u}) + N_{\mathcal{F}}(\bar{u})$.

For each "bad" iteration (j, k), i.e. (we return to double indexing)

$$\rho_{j,k} < \gamma \text{ and } \tilde{\rho}_{j,k} \geq \tilde{\gamma}, \tag{2.12}$$

define the accuracy measure

$$\epsilon_{j,k} := \frac{J_{\tilde{y}_{j,k}} - \Phi_{j,k+1}(\tilde{y}_{j,k})}{\|\iota(\tilde{y}_{j,k} - u_j)\|_Z}. \tag{2.13}$$

Theorem 2.6 (Convergence of the bundle method). *Let the initial point $u_1 \in \mathcal{F}$ be such that $\mathcal{F}_1 := \{x \in \mathcal{F} : J(x) \leq J(u_1) + 2\Delta\}$ is bounded in U and define $\epsilon_{j,k}$ as in* (2.13).

1. *If Algorithm 1 produces only finitely many serious iterates and the sequence of proximity parameters $(\tau_k)_k$ is bounded, then the last serious iterate u is stationary.*
2. *If Algorithm 1 produces only finitely many serious iterates and $(\tau_k)_k$ is unbounded, then there exists a subsequence of iterations $((j, k_i))_{i\in\mathbb{N}}$ of the type* (2.12) *such that $\tau_{j,k_i} \to \infty$, $\tilde{y}_{j,k_i} \rightharpoonup u$ and u is ϵ-stationary with $\epsilon = (M+1)/(\theta(\tilde{\gamma}-\gamma)) \liminf_i \epsilon_{j,k_i}$.*
3. *If Algorithm 1 generates infinitely many serious iterates, $(u_{j_i})_{i\in\mathbb{N}}$ is a subsequence converging weakly to $\bar{u}$, and $\liminf_i \|e_{j_i,k(i)}\|_{U^*} = 0$, cf.* (2.8)*, where $k(i)$ is the last inner iteration in outer iteration j_i (i.e., $u_{j_i+1} = \tilde{y}_{j_i,k(i)}$), then $\bar{u}$ is stationary.*
4. *If Algorithm 1 generates infinitely many serious iterates, $(u_{j_i})_{i\in\mathbb{N}}$ is a subsequence converging weakly to $\bar{u}$, and $\liminf_i \|e_{j_i,k(i)}\|_{U^*} > 0$ with $k(i)$ as in part 3, then for all i sufficiently large there exists a largest k_i such that (j_i, k_i) is of type* (2.12) *and $\bar{u}$ is ϵ-stationary, where $\epsilon = (M + 1)/(\theta(\tilde{\gamma} - \gamma)) \liminf_i \epsilon_{j_i,k_i}$.*

Algorithm 1: Inexact bundle method

Parameters : $0 < \gamma < \tilde{\gamma} < 1$, $0 < \theta < 1$, $\Delta > 0$, $R > 0$, $M > 0$, $0 < \underline{q} < \bar{q} \leq T$. Forcing sequence $(a_k)_{k \in \mathbb{N}}$, gradient approximation multifunction $G : \mathcal{F} \rightrightarrows Z^*$ fulfilling Assumption 2.1.

Initialization : Choose an initial iterate $u_1 \in \mathcal{F}$ and function value approximation $f_{u_j} \in \bar{B}(f(u_j), \Delta)$.

for $j = 1, \ldots$ **do**

  Compute $\tilde{g}_0 \in G(\iota(u_j))$ and set $J_{u_j} = f_{u_j} + w(u_j)$. Choose a symmetric operator $Q_j \in \mathcal{L}(Z, Z^*)$ and $q_j \in (\underline{q}, \bar{q})$ satisfying (2.6). Choose $\tau_1 \in [q_j, T]$. Set $m_0(\cdot, u_j) = f_{u_j} + \langle \tilde{g}_{j,0}, \iota(\cdot - u_j) \rangle_{Z^*, Z}$ and $\Phi_1 = m_0(\cdot, u_j) + w$.

  for $k = 1, \ldots$ **do**

    Trial iterate generation. Define the inner iterate y_k by

$$y_k := \arg\min_{y \in \mathcal{F}} \phi_k(y) + w(y) + \tfrac{1}{2} \|\iota(y - u_j)\|^2_{Q_j + \tau_k E}.$$

    Find a trial iterate $\tilde{y}_k \in \bar{B}_U(y_k, R) \cap \mathcal{F}$ which fulfills (2.1) and (2.2). Compute $f_{\tilde{y}_k} \in \bar{B}(f(\tilde{y}_k), \Delta)$ and set $J_{\tilde{y}_k} = f_{\tilde{y}_k} + w(\tilde{y}_k)$.

    Stop if $\tilde{y}_k = u_j$.

    Acceptance test. Set

$$\rho_k = \frac{J_{u_j} - J_{\tilde{y}_k}}{J_{u_j} - \Phi_k(\tilde{y}_k)}.$$

    if $\rho_k \geq \gamma$ **then**

      Set $u_{j+1} = \tilde{y}_k$, $f_{u_{j+1}} = f_{\tilde{y}_k}$ and quit the inner loop.

    end

    Update local model. Enrich the set of bundle information $\mathcal{B}_{k+1}$ by computing a function value approximation $f_{\tilde{y}_k}$ at the trial iterate and a subgradient $\tilde{g}_k \in G(v_k)$ at the base point v_k. Possibly add more bundle information to $\mathcal{B}_{k+1}$. Possibly delete or aggregate old cutting planes such that the new cutting planes $\mathcal{M}_{k+1}$ fulfill (2.11). Set $\Phi_{k+1} = \max\{m : m \in \mathcal{M}_{k+1}\} + w$.

    Update proximity parameter. Set $\tilde{\rho}_k = \frac{J_{u_j} - \Phi_{k+1}(\tilde{y}_k)}{J_{u_j} - \Phi_k(\tilde{y}_k)}$ and update

$$\tau_{k+1} = \begin{cases} 2\tau_k & \text{if } \tilde{\rho}_k \geq \tilde{\gamma} \\ \tau_k & \text{if } \tilde{\rho}_k < \tilde{\gamma} \end{cases}.$$

  end

end

Proof. Due to space limitations, it is not possible to give a proof here. The result can be shown by adapting and extending our convergence theory in [22]. □

Remark 2.7. 1. As in [22, Rem. 5.7], Theorem 2.6 still holds true if the function w is set to zero ($w \equiv 0$) and the feasible set $\mathcal{F}$ is bounded in U.

2. In the setting of [22], which is a special case of the situation considered here, we recover the statement of [22, Thm. 5.6]. There, for parts 2 and 4, ϵ-stationarity with $\epsilon \leq (M+1)/(\theta(\tilde{\gamma} - \gamma))(\epsilon_1 + \epsilon_2)$ is obtained under the following assumptions: The function value approximation condition

$$f_{\tilde{y}_k} - f(\tilde{y}_k) \leq f_u - f(u) + \epsilon_1 \|\iota(\tilde{y}_k - u)\|_Z + \Xi \|\iota(\tilde{y}_k - u)\|^2_Z$$

holds for $\epsilon_1 \geq 0$, $\Xi \geq 0$ with $\epsilon_1 + \Xi > 0$; $p : Z \to \mathbb{R}$ is approximately convex [13, 22] at $\iota(u)$ (part 2) or $\iota(\bar{u})$ (part 4); $G = \partial_C p + \bar{B}_{Z^*}(0, \epsilon_2)$, where $\epsilon_2 \geq 0$; $\|\iota(v_k - u)\|_Z \leq \hat{M}\|\iota(y_k - u)\|_Z$ with $\hat{M} \geq 0$; and $\Phi_k(\cdot) \geq m_k(\cdot, u)$.

3. Theorem 2.6 gives an indicator on how to refine the model. If in line 11 of Algorithm 1 the term $\epsilon_{j,k}$ is not sufficiently small, refine the function value approximation and the approximate subgradients or calculate more cutting planes to enrich the new model.
4. If exact function values (i.e. $f_u = f(u)\ \forall u \in \mathcal{F}$) and exact subgradients (i.e. $G(u) = \partial_C p(u)\ \forall u \in \iota(\mathcal{F})$) are used, there always exists a new model Φ_{k_i+1} such that the limit point $\hat{u}$ in Theorem 2.6 (i.e. $\hat{u} = u$ or $\hat{u} = \bar{u}$) is stationary. In fact, at iteration (j_i, k_i) there then exists $v_{j_i,k_i} \in [u_{j_i}, \tilde{y}_{j_i,k_i}]$ and $\tilde{g}_{j_i,k_i} \in G(\iota(v_{j_i,k_i}))$ with $p(\iota(\tilde{y}_{j_i,k_i})) - p(\iota(u_j)) \leq \langle \tilde{g}_{j_i,k_i}, \iota(\tilde{y}_{j_i,k_i} - u_{j_i})\rangle_{Z^*,Z}$ (Lebourg's mean value theorem [11, Thm. 2.3.7] shows that even "=" can be achieved). Assume we can find v_{j_i,k_i} and $\tilde{g}_{j_i,k_i}$ with this property and that the cutting plane $m^i := m_{\tilde{y}_{j_i,k_i}, v_{j_i,k_i}, \tilde{g}_{j_i,k_i}}(\cdot, u_{j_i})$ is included in the new model, i.e., $\Phi_{j_i,k_i+1} \geq m^i$, then we find

$$
\begin{aligned}
&J_{\tilde{y}_{j_i,k_i}} - \Phi_{j_i,k_i+1}(\tilde{y}_{j_i,k_i}) \leq f_{\tilde{y}_{j_i,k_i}} - m^i(\tilde{y}_{j_i,k_i}) \\
&= [f(\tilde{y}_{j_i,k_i}) + \langle \tilde{g}_{j_i,k_i}, \iota(u_{j_i} - \tilde{y}_{j_i,k_i})\rangle_{Z^*,Z} - f(u_{j_i})]_+ + c\|\iota(v_{j_i,k_i} - u_{j_i})\|_Z^2 \\
&\leq c\|\iota(\tilde{y}_{j_i,k_i} - u_{j_i})\|_Z^2.
\end{aligned}
$$

In the case that $\tilde{y}_{j_i,k_i} \rightharpoonup \hat{u}$, this shows that $\liminf \epsilon_{j_i,k_i} \leq \liminf c\|\iota(\tilde{y}_{j_i,k_i} - u_{j_i})\|_Z = 0$. According to Theorem 2.6, this means that $0 \in \iota^*\partial_C p(\iota\hat{u}) + w'(\hat{u}) + N_{\mathcal{F}}(\hat{u})$.

3. Generalized derivatives

In this section, we will define the sets of generalized derivatives that we will consider for the solution operator of the obstacle problem which is defined between infinite dimensional spaces. A generalization of so called subdifferentials for functions mapping to $\mathbb{R}$ is necessary. Due to the choice of weak and strong topologies in infinite dimensional spaces, we obtain four different generalized differentials as generalizations of the Bouligand subdifferential consisting of different combinations of topologies on the involved spaces and there is no unique generalization of the concepts in finite dimension, see also [10], [36]. For the finite dimensional case, see, e.g., [32, Def. 2.12], [14, Def. 4.6.2].

Definition 3.1. Let the operator $T\colon X \to Y$ be a locally Lipschitz continuous operator between a separable Banach space X and a Hilbert space Y. Denote the subset of X on which T is Gâteaux differentiable by D_T and let T' be the respective Gâteaux derivative. We define the following sets of (Bouligand) generalized derivatives of T in u

$$
\begin{aligned}
\partial_B^{ij} T(u) := \{&\Sigma \in \mathcal{L}(X, Y) : T'(u_n) \to \Sigma \text{ in the sense of } \mathrm{OT}(i) \\
&\text{for some } (u_n)_{n\in\mathbb{N}} \subset D_T \text{ with } u_n \to u \text{ in the sense of } \mathrm{T}(j)\}.
\end{aligned}
$$

Here, $i, j \in \{s, w\}$ and $T(s), T(w)$ means convergence in the strong, respective weak, sense in X, while $OT(s)$ means convergence in the strong operator topology and $OT(w)$ convergence in the weak operator topology and, in addition, $T(u_n) \rightharpoonup T(u)$ in Y.

Recall that convergence of operators $(T_n)_{n\in\mathbb{N}} \subseteq \mathcal{L}(X,Y)$ in the strong operator topology means pointwise convergence of $(T_n)_{n\in\mathbb{N}}$ to T in Y, convergence in the weak operator topology means pointwise weak convergence of $(T_n)_{n\in\mathbb{N}}$ to T in Y.

Remark 3.2. 1. Assume that T fulfills the assumptions of Definition 3.1 and let, in addition, Y be separable. Then the following relations between the differentials hold for all $u \in X$, see also [36, Prop. 2.11],

$$\partial_B^{ss} T(u) \subseteq \partial_B^{sw} T(u) \subseteq \partial_B^{ww} T(u) \quad \text{and} \quad \partial_B^{ss} T(u) \subseteq \partial_B^{ws} T(u) \subseteq \partial_B^{ww} T(u).$$

2. The set $\partial_B^{sw} T(u)$, and thus also $\partial_B^{ww} T(u)$, is nonempty. See also [35, Rem. 1.1].
3. Let $S\colon X \to Y$ be a solution operator of a partial differential equation or of a variational inequality, which is Lipschitz continuous on bounded sets. Let $J\colon Y \times X \to \mathbb{R}$ be a continuously differentiable objective function and denote by $\hat{J}(u) = J(S(u), u)$ the corresponding reduced objective function. Then
$$\{\Sigma^* J_y\left(S(u), u\right) + J_u\left(S(u), u\right) : \Sigma \in \partial_B^{sw} S(u)\} \subset \partial_B^{sw} \hat{J}(u) \subset \partial_C \hat{J}(u).$$
4. It directly follows that if T is Gâteaux differentiable in u with Gâteaux derivative $T'(u)$, then $T'(u)$ belongs to all generalized differentials defined in Definition 3.1.

4. Properties of the obstacle problem

In this section, we deal with the variational inequality

$$\text{Find } y \in K_\psi, \quad \langle Ly - F(u), z - y\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \geq 0 \quad \text{for all } z \in K_\psi. \tag{VI}$$

The variational inequality (VI) is a basic but, due to the operator F, quite general form of an obstacle problem. We assume that $\Omega \subseteq \mathbb{R}^d$ is an open and bounded domain and that $L \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$ is a coercive and T-monotone operator, i.e., the inequality $\langle Ly, y\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \geq C_L \|y\|^2_{H_0^1(\Omega)}$ holds for some positive constant $C_L > 0$, as well as $\langle L(y-z), (y-z)_+\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} > 0$ for all $y, z \in H_0^1(\Omega)$ with $(y-z)_+ \neq 0$, see [37].

Furthermore, $F\colon U \to H^{-1}(\Omega)$ is a continuously differentiable and monotone operator on a separable partially ordered Banach space U which is Lipschitz continuous on bounded subsets of U. We will specify the precise assumptions on U in Assumption 5.5, but let us note that the class of Banach spaces U we consider includes the important examples $H^{-1}(\Omega)$, $L^2(\Omega)$, or $\mathbb{R}^n$. The closed convex set K_ψ is of the form $K_\psi := \left\{z \in H_0^1(\Omega) : z \geq \psi\right\}$ and the quasi upper-semicontinuous obstacle ψ is chosen such that K_ψ is nonempty. The inequality "$z \geq \psi$" is to be understood pointwise quasi-everywhere (q.e.) in Ω (see e.g. [1, 2, 8, 21]).

It is well known that under these assumptions the obstacle problem (VI) has a unique solution and that the solution operator $S_F\colon U \to H_0^1(\Omega)$ that assigns the solution of the variational inequality to a given $u \in U$ is Lipschitz continuous on bounded sets, see, e.g., [4], [26], [15]. If F is the identity mapping on $H^{-1}(\Omega)$, we just write $S\colon H^{-1}(\Omega) \to H_0^1(\Omega)$ for the corresponding solution operator. Note that $S_F = S \circ F$.

The following lemma establishes monotoncity properties of the solution operator S_F. This result can be found in [15, Prob. 3, p. 30] and [37, Thm. 5.1].

Lemma 4.1. *The solution operator $S_F\colon U \to H_0^1(\Omega)$ of the obstacle problem* (VI) *is increasing: If u_1, u_2 are elements of U such that $u_1 \geq u_2$, then the inequality $S_F(u_1) \geq S_F(u_2)$ holds a.e. and q.e. in Ω.*

In the following sections, we often write $\langle \cdot, \cdot \rangle$ for the dual pairing between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$, omitting the subscript specifying the spaces.

4.1. Differentiability of the solution operator

A classical result by Mignot [29] states that the solution operator $S\colon H^{-1}(\Omega) \to H_0^1(\Omega)$ is directionally differentiable and the directional derivative is given by a variational inequality. Based on the directional differentiability of S in the sense of Hadamard, see, e.g., [38] for this notion of directional differentiability and its relation to other notions, we can apply a chain rule and obtain the directional derivative $S_F'(u;h)$ in $u \in U$ and in direction $h \in U$ for the composite mapping $S_F = S \circ F$

$$\text{Find } \eta \in \mathcal{K}_{K_\psi}(F(u)), \quad \langle L\eta - F'(u;h), z - \eta \rangle \geq 0 \quad \text{for all } z \in \mathcal{K}_{K_\psi}(F(u)). \tag{4.1}$$

Here, $\mathcal{K}_{K_\psi}(F(u)) := \mathcal{T}_{K_\psi}(S_F(u)) \cap \mu^\perp$ is called the critical cone and $\mathcal{T}_{K_\psi}(S_F(u))$ denotes the tangent cone of K_ψ at $S_F(u) \in K_\psi$, the set $\mu^\perp = \{z \in H_0^1(\Omega) : \langle \mu, z \rangle = 0\}$ is the annihilator with respect to the functional $\mu = LS_F(u) - F(u) \in H^{-1}(\Omega)$. With the help of capacity theory, one can find the following characterization of the critical cone

$$\begin{aligned} \mathcal{K}_{K_\psi}(F(u)) &= \left\{ z \in H_0^1(\Omega) : z \geq 0 \text{ q.e. on } A(u), \langle \mu, z \rangle = 0 \right\} \\ &= \left\{ z \in H_0^1(\Omega) : z \geq 0 \text{ q.e. on } A(u), z = 0 \text{ q.e. on } A_s(u) \right\}. \end{aligned} \tag{4.2}$$

Here, $A(u) := \{\omega \in \Omega : S_F(u)(\omega) = \psi(\omega)\}$ denotes the active set, the set where $S_F(u)$ touches the obstacle ψ, and $A_s(u)$ denotes the strictly active set. The second characterization in (4.2) gives an implicit representation of the strictly active set, while it can also be defined explicitly as the fine support of the regular Borel measure associated with $\mu = LS_F(u) - F(u) \in H^{-1}(\Omega)^+$. For details we refer to [29], [8, Sect. 6.4], [42, App. A]. Both sets, the active set as well as the strictly active set, are quasi-closed subsets of Ω that are defined up to a set of zero capacity.

We now specify the behavior of S_F' in points where S_F is Gâteaux differentiable. Therefore, we cite the following lemma from [35, Lem. 3.3].

Lemma 4.2. *Suppose that S_F is Gâteaux differentiable in $u \in U$ and let $h \in U$ be arbitrary. Then the directional derivative $S_F'(u;h)$ is determined by the solution to the problem*

$$\textit{Find } \eta \in H_0^1(D(u)), \quad \langle L\eta - F'(u;h), z \rangle = 0 \quad \textit{for all } z \in H_0^1(D(u)). \tag{4.3}$$

Here, any quasi-open set $D(u)$ satisfying $\Omega \setminus A(u) \subseteq D(u) \subseteq \Omega \setminus A_s(u)$ up to a set of zero capacity is admissible in (4.3) *and provides the same solution η.*

Remark 4.3.

1. The sets $H_0^1(D(u))$ are Sobolev spaces on quasi-open domains. For a thorough introduction to such spaces, we refer to [25]. The space $H_0^1(O)$ for a quasi-open set $O \subset \Omega$ can be defined as $H_0^1(O) := \{z \in H_0^1(\Omega) : z = 0 \text{ q.e. outside } O\}$.

2. $H_0^1(\Omega \setminus A(u))$ is the largest linear subset and $H_0^1(\Omega \setminus A_s(u))$ is the linear hull of the critical cone $\mathcal{K}_{K_\psi}(F(u))$. This describes the relation between (4.3) and (4.1).
3. Observe that whenever $A(u) = A_s(u)$ holds up to a set of zero capacity, i.e., when the strict complementarity condition is fulfilled in u, then up to disagreement on a set of capacity zero, there is only one set $D(u) = \Omega \setminus A(u) = \Omega \setminus A_s(u)$ admissible in Lemma 4.2. Nevertheless, due to the generality of the operator F in the variational inequality (VI), there might be points where S_F is Gâteaux differentiable and where the strict complementarity condition does not hold. This cannot happen for S.

The analysis we will carry out relies on the characterization of the Gâteaux derivative of S_F given as the solution of (4.3) with the choice $D(u) = \Omega \setminus A(u)$. In the following, we will write $I(u) := \Omega \setminus A(u)$ for the inactive set. We will find a similar description as in (4.3) also for a generalized derivative of S_F in points where S_F is not Gâteaux differentiable. In such points, different choices of $D(u)$ in (4.3) generally yield different solutions and solution operators. Therefore, we have to distinguish sequences $(u_n)_{n\in\mathbb{N}} \subseteq U$ with different properties and carefully analyze the resulting behavior and the stability of the sets $H_0^1(D(u))$ and the corresponding solution operators of (4.3). The dependency of the solutions of variational inequalities, such as (4.3), on the set of test functions, such as $H_0^1(D(u))$, will be clarified in the first part of the next section.

5. An element of the Bouligand generalized differential

In this section, we will construct an element of $\partial_B^{ss} S_F(u)$. In Lemma 4.2, we have seen that the Gâteaux derivatives of the solution operator S_F in differentiability points evaluated in a direction h solve a variational equation. Since the generalized differentials from Definition 3.1 contain limits of Gâteaux derivatives, we need to study the convergence of solutions (4.3). The tool will be the following definition, see [30] and [37].

Definition 5.1 (Mosco convergence). Let X be a Banach space and denote by $(C_n)_{n\in\mathbb{N}}$ a sequence of nonempty, closed, convex subsets of X. We say that $(C_n)_{n\in\mathbb{N}}$ converges to a closed convex set C in the sense of Mosco if and only if the following conditions hold:
1. For all $c \in C$ there is $(c_n)_{n\in\mathbb{N}}$ with $c_n \in C_n$ for all $n \in \mathbb{N}$ as well as $c_n \to c$ in X.
2. For any sequence $(c_{n_k})_{k\in\mathbb{N}}$ satisfying $c_{n_k} \in C_{n_k}$ for a subsequence $(n_k)_{k\in\mathbb{N}}$ of $(n)_{n\in\mathbb{N}}$ as well as $c_{n_k} \rightharpoonup c$ in X, it follows $c \in C$.

Based on Definition 5.1, the following result can be obtained, see [37, Thm. 4.1].

Proposition 5.2. *Let $L \in \mathcal{L}\left(H_0^1(\Omega), H^{-1}(\Omega)\right)$ be coercive and let $(C_n)_{n\in\mathbb{N}}$, C be closed convex subsets of $H_0^1(\Omega)$. Assume that $C_n \to C$ in the sense of Mosco and $h_n \to h$ in $H^{-1}(\Omega)$, then the unique solutions of*

$$\textit{Find } \eta_n \in C_n, \quad \langle L\eta_n - h_n, z - \eta_n\rangle \geq 0 \quad \textit{for all } z \in C_n$$

converge to the solution η of the limit problem with C_n, h_n replaced by C, h, respectively.

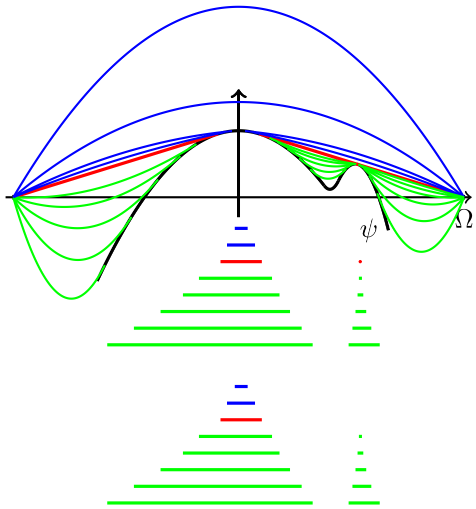

FIGURE 1. Top: An instance of the obstacle problem for a piecewise quadratic obstacle ψ. The solution $S(0)$ is plotted in red, while solutions for $S(u)$ with different parameters for $u \leq 0$ are plotted in green and for $u \geq 0$ in blue. Middle: The corresponding active sets $A(u)$ for the different values of u. Bottom: The corresponding strictly active sets $A_s(u)$ for the different values of u.

5.1. The set-valued map $u \mapsto H_0^1(I(u))$

In this subsection, we analyze the set-valued map $u \mapsto H_0^1(I(u))$ and establish a Mosco convergence result for the spaces $H_0^1(I(u_n))$. We show the convergence of $(S_F'(u_n; h))_{n\in\mathbb{N}} \subset H_0^1(\Omega)$ for Gâteaux differentiability points u_n of S_F converging from below towards u and identify the limit. This will give us an element of the Bouligand generalized differential $\partial_B^{ss} S_F(u)$. At this point, let us recall the variational equation

$$\text{Find } \eta \in H_0^1(I(u)), \quad \langle L\eta - F'(u;h), z\rangle = 0 \quad \text{for all } z \in H_0^1(I(u)). \tag{5.1}$$

for the Gâteaux derivatives that we have developed in Lemma 4.2.

The crucial point to be considered when examining the convergence of solution operators of (5.1), which is by Proposition 5.2 linked to the Mosco convergence of the sets $H_0^1(I(u_n))$, is to avoid sudden jumps in the inactive sets. As Figure 1 shows, these jumps can occur suddenly in the limit active set $I(u)$ and Mosco convergence of $H_0^1(I(u_n))$ to $H_0^1(I(u))$ cannot be expected. Figure 1 also shows the influence of monotonicity of the sequence $(u_n)_{n\in\mathbb{N}}$ on the active and strictly active sets. More precisely, different solutions of the obstacle problem are depicted in Figure 1. The associated values of u_i are chosen constant and equal to zero, respectively > 0 and < 0. We can also see the corresponding active sets $A(u_i)$ and the strictly active sets $A_s(u_i)$ underneath. In $u = 0$ with the respective solution in red, since the isolated point in $A(0)$ belongs to the set $A(0)$, but is not contained in $A_s(0)$, the strict complementarity condition does not hold, i.e., $A(0) \neq A_s(0)$. Note that a single point has capacity strictly positive in the one-dimensional case. Therefore, $u = 0$ is a point where the respective solution operator is potentially non-Gâteaux differentiable.

Example. Let us consider the sets $(H_0^1(I(u_n)))_{n\in\mathbb{N}}$. We argue that the Mosco limit will, in general, not be $H_0^1(I(u))$ for a decreasing sequence $(u_n)_{n\in\mathbb{N}}$ with $u_n \to u$. In the situation of Figure 1, choose an element $v \in H^1(\mathbb{R}^d)$ with $\{v > 0\} = \Omega \setminus A_s(0)$ up to a set of zero capacity, see [41, Prop. 2.3.14] or [18, Lem. 3.6], and define $v_n := v$ for all $n \in \mathbb{N}$. Then, it holds $v_n \in H_0^1(I(u_n))$ for all $n \in \mathbb{N}$ as well as $v_n \to v$. Nevertheless, v is not an element of $H_0^1(I(0))$. Therefore, the Mosco limit of the sequence $(H_0^1(I(u_n)))_{n\in\mathbb{N}}$ is not $H_0^1(I(0))$ (but rather $H_0^1(\Omega \setminus A_s(0))$).

This idea to consider increasing sequences $(u_n)_{n\in\mathbb{N}}$ converging to u in order to obtain Mosco convergence of the sets $H_0^1(I(u_n))$ to $H_0^1(I(u))$ is formalized in the following theorem, which is taken from [35, Thm. 5.2].

Theorem 5.3. *Let $(u_n)_{n\in\mathbb{N}} \subset U$ be an increasing sequence such that $u_n \uparrow u$. Then, the sequence $(H_0^1(I(u_n)))_{n\in\mathbb{N}}$ converges to $H_0^1(I(u))$ in the sense of Mosco. If, furthermore, S_F is Gâteaux differentiable in u_n for all $n \in \mathbb{N}$, then $(S_F'(u_n;\cdot))_{n\in\mathbb{N}}$ converges in the strong operator topology to $\Sigma_F(u;\cdot)$, where, for a given $h \in U$, the element $\Sigma_F(u;h)$ is given by the unique solution of* (5.1).

5.2. Existence of points of Gâteaux differentiability in the positive cone

Next, we argue that an increasing sequence $(u_n)_{n\in\mathbb{N}}$ converging to an arbitrary $u \in U$ in which S_F is Gâteaux differentiable always exists. With this result, we can infer that $\Sigma_F(u;\cdot) \in \mathcal{L}\left(U, H_0^1(\Omega)\right)$ is in $\partial_B^{ss} S_F(u)$. The argument is based on the following theorem.

Theorem 5.4. *Every map from a separable Banach space to a Hilbert space which is Lipschitz continuous on bounded sets is Gâteaux differentiable on a dense subset of its domain.*

A proof can be found in, e.g., [6, Thm. 6.42]. We also refer the reader to [29, Thm. 1.2], where the same result is shown for the case that only Hilbert spaces appear.

In order to ensure that Theorem 5.3 yields an element of $\partial_B^{ss} S_F(u)$, we make the following assumptions on the size of the positive cone in U.

Assumption 5.5. We assume that V is a partially ordered space such that the positive cone $\mathcal{P} := \{v \in V : v \geq 0\}$ has nonempty interior. Let V be embedded into the space U. The embedding $\iota\colon V \to U$ is assumed to be continuous, dense and compatible with the order structures of V and U, i.e., if $v_1, v_2 \in V$ with $v_1 \leq v_2$ then $\iota(v_1) \leq \iota(v_2)$ in U.

Note that Assumption 5.5 is satisfied for, e.g., $U = L^2(\Omega)$, $U = H^{-1}(\Omega)$ and $U = \mathbb{R}^n$. Now, we can show the following proposition, which is taken from [35, Prop. 5.5]

Proposition 5.6. *Let u be an arbitrary element of U and assume that Assumption 5.5 is satisfied for U. Then there exists a sequence $(u_n)_{n\in\mathbb{N}}$ such that the solution operator S_F is Gâteaux differentiable in each u_n and $u_n \uparrow u$.*

5.3. Characterization of a generalized derivative

The preceding results imply the following characterization of a generalized derivative in $\partial_B^{ss} S_F(u) \subseteq \partial_B^{sw} S_F(u)$ for arbitrary elements $u \in U$, see [35, Thm. 5.6].

Theorem 5.7. *Let Assumption 5.5 be fulfilled for U and let $u \in U$ be arbitrary. Then the operator $\Sigma_F(u;\cdot) \in \mathcal{L}(U, H_0^1(\Omega))$, where $\Sigma_F(u;h)$ is given by the unique solution to the variational equation* (5.1)*, is in the Bouligand generalized differential $\partial_B^{ss} S_F(u)$ of S_F in u.*

Let us now consider an optimal control problem where the obstacle problem describes the constraint set, such as $\min_u \hat{J}(u) = J(S_F(u), u)$. Here, $J\colon H_0^1(\Omega) \times U \to \mathbb{R}$ is a continuously differentiable objective function. An element of Clarke's generalized gradient $\partial_C \hat{J}(u)$ at the point $u \in U$ can be obtained in the following way, see [35, Thm.5.7].

Theorem 5.8. *Let q be the unique solution of the variational equation*

$$\text{Find } q \in H_0^1(I(u)), \quad \langle L^* q, v\rangle = \langle J_y\left(S_F(u), u\right), v\rangle \quad \text{for all } v \in H_0^1(I(u)). \tag{5.2}$$

Then, $F'(u)^* q + J_u\left(S_F(u), u\right)$ is in $\partial_C \hat{J}(u)$. Here, J_y and J_u denote the contiuous Fréchet derivatives of J with respect to y and u, respectively, $F'(u)^* \in \mathcal{L}\left(H_0^1(\Omega), U^*\right)$ is the (Banachian) adjoint operator of $F'(u) \in \mathcal{L}\left(U, H^{-1}(\Omega)\right)$ and $L^* \in \mathcal{L}\left(H_0^1(\Omega), H^{-1}(\Omega)\right)$ is the (Banachian) adjoint operator of $L \in \mathcal{L}\left(H_0^1(\Omega), H^{-1}(\Omega)\right)$.

6. Characterization of the entire generalized differentials

Now, we reduce the generality of (VI) and consider the obstacle problem

$$\text{Find } y \in K_\psi, \quad \langle Ly - u, z - y\rangle \geq 0 \quad \text{for all } z \in K_\psi \tag{$\mathrm{VI_{id}}$}$$

and the corresponding solution operator $S\colon H^{-1}(\Omega) \to H_0^1(\Omega)$. Here, $L = -\Delta$, i.e., $\langle Ly, v\rangle = \int_\Omega \nabla y \cdot \nabla v \, dx$. The operator F from (VI) is realized by the identity on $H^{-1}(\Omega)$.

The generalized derivative formula obtained in Theorem 5.7 applies in this special case and we have already computed an element of $\partial_B^{ss} S(u)$ for this setting. Nevertheless, we can make use of the simplified structure of ($\mathrm{VI_{id}}$) and obtain, by abstract arguments, more than just one generalized derivative. Indeed, it is possible to characterize $\partial_B^{ss} S(u)$, $\partial_B^{sw} S(u)$ and $\partial_B^{ws} S(u)$. This section is based on [36].

The preimage space of S is the whole space $H^{-1}(\Omega)$, while the operator F entering the more general variational inequality (VI) often realizes only a smaller subset of $H^{-1}(\Omega)$, since the range of F is, in general, smaller than $H^{-1}(\Omega)$. For arbitrary $u \in H^{-1}(\Omega)$, given a quasi-open set $D(u)$ with $I(u) \subseteq D(u) \subseteq \Omega \setminus A_s(u)$, the availability of all elements in $H^{-1}(\Omega)$ allows to construct a sequence $u_n \to u$ in $H^{-1}(\Omega)$ such that $S'(u_n)$ is the solution operator to

$$\text{Find } \eta \in H_0^1(D(u)), \quad \langle L\eta - h, z\rangle = 0 \quad \text{for all } z \in H_0^1(D(u)),$$

i.e., the sequence $(S'(u_n))_{n\in\mathbb{N}}$ is constant and converges. This is a strategy entering the proof of Theorem 6.1 in [36]. It indicates why we are able to characterize generalized differentials for solution operators of ($\mathrm{VI}_{\mathrm{id}}$), while the situation is much more complicated for the general variational inequality (VI). Already in finite dimensions, the authors of [20] impose a local surjectivity assumption on the analog of the operator F in finite dimension, in order to characterize a generalized differential. We obtain the following characterization of $\partial_B^{ss} S(u)$ and $\partial_B^{ws} S(u)$. For the proofs, see [36].

Theorem 6.1. *Let $u \in H^{-1}(\Omega)$ be arbitrary. The Bouligand generalized differentials $\partial_B^{ss} S(u)$ and $\partial_B^{ws} S(u)$ contain all solution operators of* (4.3) *for any quasi-open set $D(u)$ with $I(u) \subseteq D(u) \subseteq \Omega \setminus A_s(u)$ and any element of $\partial_B^{ss} S(u)$ and $\partial_B^{ws} S(u)$ is of this form.*

Remark 6.2.

1. The characterization of Theorem 6.1 applies independent from differentiability. If and only if there is no gap between $A(u)$ and $A_s(u)$ in the sense of capacity, the operator S is Gâteaux differentiable in u and the differentials $\partial_B^{ss} S(u)$ and $\partial_B^{ws} S(u)$ contain only the Gâteaux derivative.
2. The result in Theorem 6.1 supports the conjecture that by focussing on the sets $\Omega \setminus A_s(u_n)$ instead of $I(u_n)$, carrying out the appropriate analysis and using the approach from Section 5, one would indeed obtain a further generalized derivative, also for the variational inequality (VI) invoking the monotone operator F.

6.1. Capacitary measures and the differentials involving the weak operator topologies

As already mentioned in Remark 3.2, the generalized differentials using the weak operator topology are supersets of the differentials characterized in Theorem 6.1. In this subsection, we will see that they are in fact larger and get to know the objects they contain in addition.

Definition 6.3. We denote by $\mathcal{M}_0(\Omega)$ the set of all regular Borel measures μ on Ω with the property that $\mu(B) = 0$ holds for every Borel set $B \subseteq \Omega$ with $\operatorname{cap}(B) = 0$. Here, regularity of μ means that $\mu(B) = \inf\{\mu(O) : O \text{ quasi-open}, B \subseteq O\}$ holds for every Borel set $B \subseteq \Omega$. The set $\mathcal{M}_0(\Omega)$ is called the set of capacitary measures on Ω.

The convergence of solution operators of

$$\text{Find } \eta \in H_0^1(O), \quad \langle L\eta - h, z \rangle = 0 \quad \text{for all } z \in H_0^1(O) \tag{6.1}$$

for quasi-open sets $O \subseteq \Omega$ in the weak operator topology is metrizable, see [12, Prop. 4.9], but the resulting metric space is not a complete space. Recall that the Gâteaux derivative operators of S in points of differentiability are exactly of this form with $I(u) \subseteq O \subseteq \Omega \setminus A_s(u)$, see Lemma 4.2. For $\mu \in \mathcal{M}_0(\Omega)$, denote by $X_\mu(\Omega)$ the space $H_0^1(\Omega) \cap L_\mu^2(\Omega)$, where $L_\mu^2(\Omega)$ is the Lebesgue space of square integrable functions on Ω w.r.t. the measure μ. As shown in, e.g., [12], the completion contains exactly the solution operators of

$$\text{Find } \eta \in X_\mu(\Omega), \quad \int_\Omega \nabla\eta \cdot \nabla z \, dx + \int_\Omega \eta \, z \, d\mu = \langle h, z \rangle \quad \text{for all } z \in X_\mu(\Omega) \tag{6.2}$$

for all capacitary measures μ on Ω. Thus, it is not surprising, that a subset of these solution operators of (6.2) enter the set $\partial_B^{sw} S(u)$. For the details, see [36].

Theorem 6.4. *Under some regularity assumptions on the obstacle and on $S(u)$, the Bouligand differential $\partial_B^{sw} S(u)$ in $u \in H^{-1}(\Omega)$ contains exactly all solution operators of*

(6.2) *for any capacitary measure* μ *fulfilling* $\mu(I(u)) = 0$ *and* $\mu = +\infty$ *on* $A_s(u)$. *Here,* $\mu = +\infty$ *on* $A_s(u)$ *means that* $v = 0$ *q.e. on* $A_s(u)$ *holds for all* $v \in H_0^1(\Omega) \cap L_\mu^2(\Omega)$.

Remark 6.5.

1. For a quasi-open set $O \subset \Omega$, we can define for each Borel set $B \subset \Omega$

$$\infty_{\Omega\setminus O}(B) := \begin{cases} 0, & \text{if } \operatorname{cap}(B \setminus O) = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

   With this definition, $\infty_{\Omega\setminus O}$ is a capacitary measure and L_O is the solution operator of (6.1) if and only if L_O is the solution operator of (6.2) with $\mu = \infty_{\Omega\setminus O}$. The condition $I(u) \subseteq O \subseteq A_s(u)$ can be expressed as $\infty_{\Omega\setminus O}(I(u)) = 0$ and $\infty_{\Omega\setminus O} = +\infty$ on $A_s(u)$. In this sense, $\partial_B^{ss} S(u) = \partial_B^{ws} S(u) \subseteq \partial_B^{sw} S(u)$.
2. The notion of convergence for the solution operators of (6.2) based on the weak operator topology is also called γ-convergence of the respective measures. The study of this convergence and the so called relaxed Dirichlet problems, such as (6.2), is interesting also in shape optimization, see [9] or [2].
3. Mosco convergence of sets $H_0^1(O_n)$ to $H_0^1(O)$ for quasi-open sets $O_n, O \subseteq \Omega$ is equivalent to the γ-convergence of the measures $\infty_{\Omega\setminus O_n}$ to $\infty_{\Omega\setminus O}$. This gives the link to the approach in Section 5.

In [36], an example is given which illustrates that the generalized differential $\partial_B^{ww} S(u)$ is very large, even when S is Gâteaux differentiable in u and $A(u) = A_s(u)$.

Based on the characterization of the generalized differentials for S, necessary optimality conditions for the optimal control of the obstacle problem with control constraints can be obtained, see [36].

7. The stochastic obstacle problem

The subject of this section is the optimal control problem ($\mathbf{P}_\mathrm{s}$) governed by the stochastic obstacle problem ($\mathbf{VI}_\mathrm{s}$). We want to find stationary points of the reduced objective function by applying the bundle method, developed in Section 2. To do so, we need to calculate a subgradient of the reduced objective function or an approximate subgradient in the sense of Assumption 2.1.

7.1. Problem Setting

Let $(\Xi, \mathcal{A}, P)$ be a measure space and set $Y := H_0^1(\Omega)$. For $\xi \in \Xi$, we consider a variational inequality of type (VI). In particular, let $L_\xi \in \mathcal{L}(Y, Y^*)$ be an operator, $\psi_\xi \in H^1(\Omega)$ an obstacle, $b \in Y^*$, define the set $K_{\psi_\xi} := \{y \in H_0^1(\Omega) : y \geq \psi_\xi\}$ and the parametric obstacle problem

$$\text{Find } y \in K_{\psi_\xi}, \quad \langle L_\xi y - b, z - y\rangle_{Y^*,Y} \geq 0 \quad \text{for all } z \in K_{\psi_\xi}. \tag{VI$_\xi$}$$

We want to relate the solutions to (VI$_\xi$) and ($\mathbf{VI}_\mathrm{s}$), see [16, 17] for related results. Using standard techniques, one can show that the projection onto the set $\mathbf{K}_{\boldsymbol{\psi}}$, defined in (1.2), agrees pointwise P-a.e. with the projection onto K_{ψ_ξ}:

Lemma 7.1. *If $\boldsymbol{\psi} \in \bar{\mathbf{Y}}$ such that $K_{\psi_\xi} \neq \emptyset$ for P-a.a. $\xi \in \Xi$ then $\mathbf{K}_{\boldsymbol{\psi}}$ is a nonempty closed convex subset of $\mathbf{Y}$ and $P_{\mathbf{K}_{\boldsymbol{\psi}}}(\mathbf{y})(\xi) = P_{K_{\psi_\xi}}(\mathbf{y}(\xi))$ for P-a.a. $\xi \in \Xi$ and for all $\mathbf{y} \in \bar{\mathbf{Y}}$.*

Using this result, we can show that the solution operator of ($\mathbf{VI}_\mathrm{s}$) agrees pointwise P-a.e. with the solution operator of (VI_ξ). We need the following definition:

Definition 7.2. A family of operators $(L_\xi)_{\xi\in\Xi} \subset \mathcal{L}(Y, Y^*)$ is called uniformly coercive, if there exists a parameter $C_L > 0$ such that $\langle L_\xi x, x\rangle_{Y^*,Y} \geq C_L \|x\|_Y^2$ for P-a.a. $\xi \in \Xi$.

Theorem 7.3. *Assume that $\xi \mapsto L_\xi y$ is measurable for every $y \in Y$, $\xi \mapsto \|L_\xi\|_{\mathcal{L}(Y,Y^*)}$ is in $L^\infty(\Xi)$, and $\boldsymbol{\psi} : \xi \mapsto \psi_\xi$ is in $\bar{\mathbf{Y}}$. Then, for every $\mathbf{y} \in \mathbf{Y}$, the map $\mathbf{L}_\mathbf{y} : \xi \mapsto L_\xi(\mathbf{y}(\xi))$ is in $\mathbf{Y}^*$ and $\mathbf{L} : \mathbf{y} \mapsto \mathbf{L}_\mathbf{y}$ is in $\mathcal{L}(\mathbf{Y}, \mathbf{Y}^*)$. Moreover, suppose that $(L_\xi)_{\xi\in\Xi}$ is uniformly coercive and that $K_{\psi_\xi} \neq \emptyset$ for P-a.a. $\xi \in \Xi$. Then, for P-a.a. $\xi \in \Xi$ and all $b \in Y^*$, (VI_ξ) has a unique solution $y_{\xi,b}$ and the solution operator $S_\xi : Y^* \to Y$, $S_\xi(b) := y_{\xi,b}$, is Lipschitz with modulus $1/C_L$. Furthermore, for all $\mathbf{b} \in \mathbf{Y}^*$, ($\mathbf{VI}_\mathrm{s}$) has a unique solution $\mathbf{y}_\mathbf{b}$, the solution operator $\mathbf{S} : \mathbf{Y}^* \to \mathbf{Y}$, $\mathbf{S}(\mathbf{b}) := \mathbf{y}_\mathbf{b}$ is Lipschitz with modulus $1/C_L$, and $(\mathbf{S}(\mathbf{b}))(\xi) = S_\xi(\mathbf{b}(\xi))$ for P-a.a. $\xi \in \Xi$.*

Proof. Under the given integrability assumptions one can show that $\mathbf{L} \in \mathcal{L}(\mathbf{Y}, \mathbf{Y}^*)$ is well defined and coercive with constant C_L. The Lions-Stampacchia theorem, cf. [26, Thm. 2.1] implies that both problems are uniquely solvable. Since $\mathbf{y} \in \mathbf{Y}$, defined by $\mathbf{y}(\xi) := S_\xi(\mathbf{b}(\xi)) \in K_{\psi_\xi}$, fulfills ($\mathbf{VI}_\mathrm{s}$) and the solution of ($\mathbf{VI}_\mathrm{s}$) is unique, we deduce $\mathbf{y} = \mathbf{S}(\mathbf{b})$. □

7.2. Approximate subgradients of the stochastic reduced objective function

In this section we show that the weak subgradients (1.4) can be used in the bundle method since they fulfill Assumption 2.1. We work in the following setting:

Assumption 7.4. Let $\mathcal{F}_Z$ be an open subset of a separable reflexive Banach space Z. Suppose that for all $\xi \in \Xi$ the functions $p_\xi : \mathcal{F}_Z \to \mathbb{R}$ satisfy the following conditions:
1. For all $z \in \mathcal{F}_Z$, the map $\xi \mapsto p_\xi(z)$ is measurable.
2. There exists a $z \in \mathcal{F}_Z$ such that $\int_\Xi |p_\xi(z)|\, dP(\xi) < \infty$.
3. For all bounded sets $B \subset Z$ there exists a function $L_B \in L^1(\Xi)$ such that

$$|p_\xi(z_1) - p_\xi(z_2)| \leq L_B(\xi)\|z_1 - z_2\|_Z \quad \text{for all } z_1, z_2 \in B \cap \mathcal{F}_Z \text{ and for } P\text{-a.a. } \xi \in \Xi.$$

Let $\bar{\mathcal{F}}$ be a closed subset of $\mathcal{F}_Z$ and consider the map $G : \bar{\mathcal{F}} \rightrightarrows Z^*$ defined by

$$G(z) := \left\{ \int_\Xi g(\xi)\, dP(\xi) : g \in L^1(\Xi, Z^*), g(\xi) \in \partial_C p_\xi(z)\ P\text{-a.e.} \right\}. \tag{7.1}$$

Theorem 7.5. *Under Assumption 7.4, the multifunction $G : \bar{\mathcal{F}} \rightrightarrows Z^*$, defined in (7.1), fulfills Assumption 2.1 and it holds $\partial_C p(z) \subset G(z)$ for all $z \in \bar{\mathcal{F}}$, where $p : \bar{\mathcal{F}} \to \mathbb{R}$ is defined by $p(z) := \int_\Xi p_\xi(z)\, dP(\xi)$.*

Proof. $\boldsymbol{\partial_C p(z) \subset G(z)}$. Let $z \in \bar{\mathcal{F}}$ be arbitrary. By [11, Thm. 2.7.2], p is well defined, locally Lipschitz and for every $g \in \partial_C p(z)$ there is a corresponding mapping $\xi \mapsto g_\xi$ from Ξ to Z^* with $g_\xi \in \partial_C p_\xi(z)$ P-a.e. and such that for every $v \in Z$, the function

$\xi \mapsto \langle g_\xi, v\rangle_{Z^*,Z}$ belongs to $L^1(\Xi)$ and one has $\langle g, v\rangle_{Z^*,Z} = \int_\Xi \langle g_\xi, v\rangle_{Z^*,Z}\, dP(\xi)$. Consequently, by [24, Cor. 1.1.2], the map $\xi \mapsto g_\xi$ is measurable. Denote by $L_B \in L^1(\Xi)$ the function according to property 3 of Assumption 7.4 for $B := \bar{B}_X(z, 1)$. From $\int_\Xi \|g_\xi\|_{X^*}\, dP(\xi) \le \int_\Xi L_B(\xi)\, dP(\xi) < \infty$ we deduce that $\xi \mapsto g_\xi$ is in $L^1(\Xi, Z^*)$ which shows $\partial_C p(z) \subset G(z)$.

1. For arbitrary $z \in \bar{\mathcal{F}}$ it holds $\emptyset \neq \partial_C p(z) \subset G(z)$. Therefore $G(z)$ is nonempty. Since $\partial_C p_\xi(z)$ is convex P-a.e. , the set $G(z)$ is convex.

2. Let $B \subset Z$ be a bounded set and denote

$$\hat{G} := \{\hat{g} \in L^1(\Xi, Z^*) : z \in B \cap \bar{\mathcal{F}}, \hat{g}(\xi) \in \partial_C p_\xi(z)\ P\text{-a.e. }\}. \tag{7.2}$$

Choose a neighborhood $\hat{B} \subset Z$ of $B \cap \bar{\mathcal{F}}$ and denote by $L_{\hat{B}} \in L^1(\Xi)$ the function which fulfills property 3 of Assumption 7.4. By [11, Prop. 2.1.2], there holds $\partial_C p_\xi(z) \subset \bar{B}_{Z^*}(0, L_{\hat{B}}(\xi))$ for all $z \in B \cap \bar{\mathcal{F}}$. Consequently, $\hat{G}$ is bounded in $L^1(\Xi, Z^*)$ by the constant $\int_\Xi L_{\hat{B}}(\xi)\, dP(\xi) < \infty$ and we find for arbitrary $g \in G(B \cap \bar{\mathcal{F}})$ that there exists a $\hat{g} \in \hat{G}$ such that $g = \int_\Xi \hat{g}(\xi)\, dP(\xi)$ and it holds

$$\|g\|_{Z^*} = \| \int_\Xi \hat{g}(\xi)\, dP(\xi)\|_{Z^*} \le \int_\Xi \|\hat{g}(\xi)\|_{Z^*}\, dP(\xi) \le \int_\Xi L_{\hat{B}}(\xi)\, dP(\xi).$$

3. We verify the assumptions of [34, Thm. 4.2]. Since $\bar{\mathcal{F}}$ is a closed subset of a complete metric space, $(\bar{\mathcal{F}}, \|\cdot\|_Z)$ is a complete metric space. By [11, Prop. 2.1.2] the map $(\xi, z) \mapsto \partial_C p_\xi(z)$ is nonempty, closed and convex valued. Using [11, Lem. 2.7.2], [3, Thm. 8.2.11 and Thm. 8.2.9] one sees that the multifunction $\xi \mapsto \partial_C p_\xi(z)$ is measurable for all $z \in \mathcal{F}_Z$, i.e. for every open set $\mathcal{O}$ the inverse image $\{\xi \in \Xi : \partial_C p_\xi(z) \cap \mathcal{O} \neq \emptyset\}$ is measurable. By [11, Prop. 2.1.5], for all $\xi \in \Xi$, $\partial_C p_\xi$ has a weakly closed graph. Now let $B \subset Z$ be a compact set and denote by $L_{\hat{B}} \in L^1(\Xi)$ a function which fulfills property 3 of Assumption 7.4 for a neighborhood $\hat{B}$ of B. Define $G_B : \Xi \rightrightarrows Z^*$ to be the multifunction $G_B(\xi) := \text{w-cl co } \cup_{z\in B}\, \partial_C p_\xi(z)$. First note that, since B is bounded, [11, Prop. 2.1.2] implies that $\cup_{z\in B} \partial_C p_\xi(z)$ is bounded by $L_{\hat{B}}(\xi)$ P-almost everywhere, i.e. G_B is integrably bounded. Also, for fixed $\xi \in \Xi$, the set $G_B(\xi)$ is bounded. Consequently, by Alaoglu's theorem, $G_B(\xi)$ is weakly compact, and obviously nonempty and convex. As $\partial_C p_\xi(z) \subset G_B(\xi)$ P-a.e. and for all $z \in B$, [34, Thm. 4.2] yields that $z \mapsto \tilde{G}(z)$ is weakly upper semicontinuous, i.e. for every weakly closed set $C \subset Y$ the set $G^-(C) := \{x \in \mathcal{F} : G(x) \cap C \neq \emptyset\}$ is closed in $\mathcal{F}$. By [33, Cor. 3.1], the multifunction $\tilde{G}$ is weakly closed valued. Therefore, [23, Thm. 2.5] implies that $G : \bar{\mathcal{F}} \rightrightarrows Z^*$ has a weakly closed graph. □

Example. For all $\xi \in \Xi$, let $J_\xi : Y \to \mathbb{R}$ be given via $J_\xi(\cdot) := \frac{1}{2}\|O_\xi(\cdot) - y_\xi^d\|_H^2$, where $O_\xi \in \mathcal{L}(Y, H)$ is the stochastic observation operator and $y_\xi^d \in H$ is the stochastic desired state. Under the assumptions of Theorem 7.3 and if additionally $(\Xi, \mathcal{A}, P)$ is a probability space, $\xi \mapsto O_\xi y$ is measurable for all $y \in Y$, $\xi \mapsto \|O_\xi\|_{\mathcal{L}(Y,H)}$ is in $L^\infty(\Xi)$ and $\xi \mapsto y_\xi^d$ is in $L^2(\Xi, H)$, then the functions $p_\xi := J_\xi(S_\xi(\cdot))$ satisfy Assumption 7.4. Consequently, $G(z) := \mathbb{E}[\partial_C(J_\xi(S_\xi(\cdot)))(z)]$ fulfills Assumption 2.1 and can be used as a subdifferential for the bundle method.

7.3. Computation of exact subgradients

Although the approach in the previous section is very versatile, it uses an approximate subdifferential G that is possibly larger than the Clarke differential of the cost function. The resulting (ϵ-)stationarity, cf. Theorem 2.6, then corresponds to weak (ϵ-)stationarity (cf. (1.5) for weak stationarity). If possible, it would be favorable to search for Clarke stationary points (1.3). To do so, the bundle method requires elements of the Clarke subdifferential $\partial_C(\mathbb{E}\left[J_\xi(S_\xi(\iota(\cdot)))\right])(u)$ (or approximations thereof). In this section we derive a formula to compute such subgradients under additional assumptions on the regularity of the problem data. We do not make use of chain rules for Clarke's subdifferential since they require a certain regularity, see [11].

Assume that Ξ is a separable Banach space, and let $(\Xi, \mathcal{A}, P)$ be a finite measure space. We assume that there is a nondegenerate Gaussian measure $\mathbb{P}$ on Ξ, such that P is absolutely continuous w.r.t. $\mathbb{P}$. For the notion of nondegenerate Gaussian measures we refer to [6] and Definition 7.7. For $\xi \in \Xi$, let $L_\xi \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$ be a T-monotone operator. We assume that the family of operators $(L_\xi)_{\xi\in\Xi}$ is uniformly coercive in the sense of Definition 7.2. Furthermore, let $\Xi \ni \xi \mapsto L_\xi \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$ be Lipschitz continuous on bounded sets. Then the maps $\Xi \ni \xi \mapsto L_\xi y \in H^{-1}(\Omega)$ are measurable for all $y \in H_0^1(\Omega)$. We also assume that $\xi \mapsto \|L_\xi\|_{\mathcal{L}(H_0^1(\Omega),H^{-1}(\Omega))}$ is in $L^\infty(\Xi)$. Let $F\colon \Xi \times U \to H^{-1}(\Omega)$ be an operator that is Lipschitz continuous on bounded sets and satisfies $F(\cdot, u) \in L^2(\Xi, H^{-1}(\Omega))$ for all $u \in U$. As before, assume that $F(\xi, \cdot)$ is monotone and continuously differentiable for almost all $\xi \in \Xi$. We keep the assumptions on U previously considered in Section 5, see Assumption 5.5. We consider the following subclass of parametric obstacle problems (VI$_\xi$)

$$\text{Find } y \in K_\psi, \quad \langle L_\xi y - F(\xi, u), z - y\rangle \geq 0 \quad \text{for all } z \in K_\psi. \qquad (\mathrm{VI}'_\xi)$$

Here, the obstacle $\psi \in H^1(\Omega)$ is chosen such that $K_\psi \neq \emptyset$. In contrast to (VI$_\xi$), the obstacle does not depend on the parameter ξ. We denote the solution operator of the family in (VI$'_\xi$) by $S_{F,\xi}\colon U \to H_0^1(\Omega)$. Note that $S_{F,\xi}$ is defined only for almost all $\xi \in \Xi$ and that Theorem 7.3 implies that $\xi \mapsto S_{F,\xi}(u)$ is in $L^2(\Xi, H_0^1(\Omega))$ for all $u \in U$. Let $(J_\xi\colon H_0^1(\Omega) \times U \to \mathbb{R})_{\xi\in\Xi}$ be a family of parametrized objective functions, such that almost all J_ξ are continuously differentiable and such that $(J_\xi)_{\xi\in\Xi}$ is uniformly Lipschitz continuous on bounded sets for almost all $\xi \in \Xi$. Furthermore, let $\xi \mapsto J_\xi(S_{F,\xi}(u), u)$ be integrable for all $u \in U$. For example, consider the parameter dependent objective functions $J_\xi(y, u) = \frac{1}{2}\|y - y_\xi^d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|_U^2$, for a family $(y_\xi^d)_{\xi\in\Xi} \subseteq H_0^1(\Omega)$ such that $\xi \mapsto y_\xi^d$ is in $L^2(\Xi, H_0^1(\Omega))$. Now, we are interested in the optimal control of the stochastic obstacle problem of the form

$$\min_{u\in U_{ad}} \hat{J}(u) = \int_\Xi J_\xi(S_{F,\xi}(u), u)\, dP(\xi). \qquad (\mathbf{P'})$$

The set $U_{\text{ad}} \subset U$ is a closed convex subset of U.

We verify the following Lipschitz continuity of $S_{F,\xi}$.

Lemma 7.6. *Under the above assumptions, the mapping $T\colon \Xi \times U \ni (\xi, u) \mapsto S_{F,\xi}(u) \in H_0^1(\Omega)$ is Lipschitz continuous on bounded subsets of $\Xi \times U$.*

Proof. Let $B_\Xi \times B_U$ be a bounded subset. By assumption $\xi \mapsto L_\xi$ is Lipschitz continuous on B_Ξ and F is Lipschitz continuous on $B_\Xi \times B_U$ with Lipschitz constants c_1 and c_2, respectively. Moreover, L_ξ is coercive with a common coercivity constant C_L for almost all ξ and by continuity for all ξ.

For $i = 1, 2$, let $(\xi_i, u_i) \in B_\Xi \times B_U$ and denote $F_i := F(\xi_i, u_i)$, $L_i := L_{\xi_i}$ and $y_i := S_{F,\xi_i}(u_i)$. Now, we estimate

$$\begin{aligned}
&C_L \|y_1 - y_2\|^2_{H^1_0(\Omega)} \le \langle L_1 y_1 - L_2 y_2 + (L_2 - L_1) y_2, y_1 - y_2 \rangle \\
&\le \langle F_1 - F_2, y_1 - y_2 \rangle + \langle (L_2 - L_1) y_2, y_1 - y_2 \rangle \\
&\le (\|F_1 - F_2\|_{H^{-1}(\Omega)} + \|(L_2 - L_1) y_2\|_{H^{-1}(\Omega)}) \|y_1 - y_2\|_{H^1_0(\Omega)} \\
&\le \left(c_1 \|u_1 - u_2\|_U + (c_1 + c_2 \|y_2\|_{H^1_0(\Omega)}) \|\xi_1 - \xi_2\|_\Xi \right) \|y_1 - y_2\|_{H^1_0(\Omega)}.
\end{aligned}$$

We obtain

$$\|y_1 - y_2\|_{H^1_0(\Omega)} \le \frac{1}{C_L} \left(c_1 \|u_1 - u_2\|_U + (c_1 + c_2 \|y_2\|_{H^1_0(\Omega)}) \|\xi_1 - \xi_2\|_\Xi \right). \tag{7.3}$$

Since this inequality holds for fixed $y_2 = S_{F,\xi_2}(u_2)$ and arbitrary $y_1 = S_{F,\xi_1}(u_1)$ with $(\xi_1, u_1) \in B_\Xi \times B_u$, this shows that $\|y\| \le c$ holds for some constant $c > 0$ and for all $y \in S_{F,B_\Xi}(B_U)$. Inequality (7.3) also shows the desired Lipschitz continuity on $B_\Xi \times B_U$. □

7.4. Construction of a subgradient for $\hat{J}$

Similar to the argument in Section 5.2, using the Lipschitz continuity of T established in Lemma 7.6, we can argue that T is Gâteaux differentiable on a large set. For our analysis, we need the notion of Gaussian measures on separable Banach spaces, see also [6], [7].

Definition 7.7.

1. Let X be a separable Banach space. A Borel probability measure $\mathbb{P}$ on X is called a Gaussian measure if for every $x^* \in X^*$ the pushforward measure $\mathbb{P}_{x^*}$ of $\mathbb{P}$ with respect to x^*, i.e., $\mathbb{P}_{x^*}(A) = \mathbb{P}((x^*)^{-1}(A))$ for any measurable set $A \subset \mathbb{R}$, has a Gaussian distribution.
2. The Gaussian measure $\mathbb{P}$ is called nondegenerate, if $\mathbb{P}_{x^*}$ is nondegenerate for every $X^* \ni x^* \neq 0$, i.e., if it is not a Dirac measure.

Lemma 7.8. *Let U fulfill the conditions of Assumption 5.5 and let $\mathbb{V}$ be an arbitrary nondegenerate Gaussian measure on V. As in Lemma 7.6, consider the operator T with $T(\xi, u) = S_{F,\xi}(u)$. For an arbitrary $u \in U$ let $\bar{T} \colon \Xi \times V \to H^1_0(\Omega)$ be defined by $\bar{T}(\xi, v) = T(\xi, v + u)$. Then $\bar{T}$ is Gâteaux differentiable except on a $P \otimes \mathbb{V}$-null set in $\Xi \times V$.*

Proof. By the properties of V, the operator $\bar{T}$ is Lipschitz continuous on bounded subsets of $\Xi \times V$. [6, Theorem 6.42] and the equivalence of notions of neglible sets developed in [6, Chap. 6.3], imply that $\bar{T}$ is Gâteaux differentiable on all points of its domain $\Xi \times V$ except on a Gauss null set, i.e., all nondegenerate Gaussian measures on $\Xi \times V$ vanish on this set. Note that the results from [6] easily carry over to operators which are Lipschitz continuous only on bounded subsets of their domain.

Since $\mathbb{P}, \mathbb{V}$ are nondegenerate Gaussian measures on Ξ, respectively V, the measure $\mathbb{P} \otimes \mathbb{V}$ is a Gaussian measure on $\Xi \times V$, see [7, Cor. 2.2.6]. It is also nondegenerate. To see this, let $(\xi^*, v^*) \in (\Xi \times V)^*$ be an arbitrary element of the dual space. Denote the density of $\mathbb{P}_{\xi^*}$, respectively $\mathbb{V}_{v^*}$, w.r.t. the Lebesgue measure by ρ_{ξ^*}, respectively ρ_{v^*}. Then, for all measurable sets A it holds

$$
\begin{aligned}
&(\mathbb{P} \otimes \mathbb{V})_{(\xi^*, v^*)}(A) = \int_\Xi \mathbb{V}(\{v \in V : v^*(v) \in (A - \xi^*(\xi))\})\, d\mathbb{P}(\xi) \\
&= \int_\Xi \mathbb{V}_{v^*}(A - \langle \xi^*, \xi \rangle)\, d\mathbb{P}(\xi) = \int_\mathbb{R} \mathbb{V}_{v^*}(A - t)\, d\mathbb{P}_{\xi^*}(t) \\
&= \int_\mathbb{R} \int_\mathbb{R} \chi_{A-t}(s)\, \rho_{v^*}(s)\, ds\, \rho_{\xi^*}(t)\, dt = \int_\mathbb{R} \int_\mathbb{R} \chi_A(r) \rho_{v^*}(r - t)\, dr\, \rho_{\xi^*}(t)\, dt \\
&= \int_\mathbb{R} \chi_A(r) \int_\mathbb{R} \rho_{v^*}(r - t) \rho_{\xi^*}(t)\, dt\, dr = \int_\mathbb{R} \chi_A(r) (\rho_{v^*} * \rho_{\xi^*})(r)\, dr.
\end{aligned}
$$

Since the convolution of two normal distributions is again a normal distribution, the conclusion follows. Thus, the set of points where $\bar{T}$ is not Gâteaux differentiable is a $\mathbb{P} \otimes \mathbb{V}$-null set. Since P is absolutely continuous w.r.t. $\mathbb{P}$, the lemma is proved. ☐

Remark 7.9. By [6, Prop. 6.18, 6.20] there is a nondegenerate Gaussian measure on V.

Lemma 7.10. *Let U fulfill the conditions of Assumption 5.5 and let $u \in U$ be arbitrary. Then there is an increasing sequence $(u_n)_{n \in \mathbb{N}} \subseteq U$ converging to u where $T(\xi, \cdot) = S_{F,\xi}$ is differentiable for P-almost all $\xi \in \Xi$.*

Proof. Let $\mathbb{V}$ be an arbitrary nondegenerate Gaussian measure on V and let N be the set of points in $\Xi \times V$ where $\bar{T}$, defined as in Lemma 7.8, is not differentiable. Then, Lemma 7.8 implies

$$0 = (P \otimes \mathbb{V})(N).$$

We want to proceed as in the proof of Proposition 5.6, see also [35, Prop. 5.1] for the proof, and construct a sequence $(u_n)_{n \in \mathbb{N}} \subseteq V$, such that each u_n is taken from a specified set with interior points, to ensure the monotonicity of the sequence. Thus, we have to ensure that sets with interior points contain common points of Gâteaux differentiability of the family $(\bar{T}(\xi, \cdot))_{\xi \in \Xi}$ for P-almost all $\xi \in \Xi$.

Therefore, let $n \in \mathbb{N}$ and let O_n be a set in V with interior points. In [40], the support of Gaussian measures is discussed. Nondegenerate Gaussian measures on separable spaces have full support, i.e., any nondegenerate Gaussian measure has a positive measure on any measurable set with interior points. This implies $\mathbb{V}(O_n) > 0$. For each $v \in V$ define $N_v := \{\xi \in \Xi : (\xi, v) \in N\}$ and consider $V_0 := \{v \in V : P(N_v) = 0\}$. Then, we have

$$0 = (P \otimes \mathbb{V})(N) = \int_V P(N_v)\, d\mathbb{V}(v),$$

i.e., $\mathbb{V}(V \setminus V_0) = 0$. Let us now consider the set $\tilde{O}_n := O_n \cap V_0 = \{v \in O_n : P(N_v) = 0\}$. Then, it holds $\mathbb{V}(O_n) = \mathbb{V}(\tilde{O}_n) > 0$. Choose an arbitrary $v_n \in \tilde{O}_n$. By definition, $\bar{T}$ is Gâteaux differentiable in (ξ, v_n) for P-almost all $\xi \in \Xi$. In particular, $\bar{T}(\xi, \cdot)$ is Gâteaux differentiable in v_n for P-almost all $\xi \in \Xi$. We can thus find a sequence $(v_n)_{n\in\mathbb{N}} \subseteq V$, where $\bar{T}(\xi, \cdot)$ is Gâteaux differentiable for almost all $\xi \in \Xi$.

We can again argue as in the proof of Proposition 5.6 to conclude that also $T(\xi, \cdot) = S_{F,\xi}$ defined on the whole space U is Gâteaux differentiable in $v_n + u$ for P-almost all $\xi \in \Xi$.

□

Theorem 7.11. *Let U fulfill the conditions of Assumption 5.5 and let $u \in U$ be arbitrary. Then, under the assumptions on the data specified in Section 7.3, a subgradient for $\hat{J}$ as defined in* (**P′**) *is given by $\int_\Xi \Sigma_\xi(u)\, dP(\xi)$, where $\Sigma_\xi(u)$ is the subgradient of the parameter dependent reduced objective function $\hat{J}_\xi$ in u constructed in Theorem 5.8.*

Proof. We consider the reduced objective function

$$\hat{J}(u) = \int_\Xi \hat{J}_\xi(u)\, dP(\xi) = \int_\Xi J_\xi(S_{F,\xi}(u), u)\, dP(\xi) = \int_\Xi J_\xi(T(\xi, u), u)\, dP(\xi).$$

Since $\hat{J}_\xi$ is Lipschitz continuous on bounded sets with a common Lipschitz constant for almost all $\xi \in \Xi$, we can exchange integration and differentiation and obtain the differentiability of $\hat{J}$ in each u_n and it holds $\hat{J}_u(u_n) = \int_\Xi (\hat{J}_\xi)_u(u_n)\, dP(\xi)$. Since $(\hat{J}_\xi)_u(u_n)$ is bounded by the common Lipschitz constant of $\hat{J}_\xi$ on a bounded set containing $(u_n)_{n\in\mathbb{N}}$, we can again exchange limits and obtain

$$\lim_{n\to\infty} \hat{J}_u(u_n) = \int_\Xi \lim_{n\to\infty} (\hat{J}_\xi)_u(u_n)\, dP(\xi) = \int_\Xi \Sigma_\xi(u)\, dP(\xi),$$

where for each $\xi \in \Xi$, the integrand $\Sigma_\xi(u)$ is the subgradient of $\hat{J}_\xi$ given in Theorem 5.8.

□

Acknowledgments

The authors would like to acknowledge the support by the DFG through the Priority Programme SPP 1962 within Project 23.

References

[1] D. Adams, L. Hedberg, *Function spaces and potential theory.* Springer, 1996.

[2] H. Attouch, G. Buttazzo, G. Michaille, *Variational analysis in Sobolev and BV spaces.* SIAM, 2014.

[3] J.-P. Aubin, H. Frankowska, *Set-valued analysis.* Birkhäuser, 2009.

[4] V. Barbu, *Optimal control of variational inequalities.* Pitman, 1984.

[5] H.H. Bauschke, P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces.* Springer, 2011.

[6] Y. Benyamini, J. Lindenstrauss, *Geometric nonlinear functional analysis.* AMS, 2000.

[7] V. Bogachev, *Gaussian measures.* AMS, 1998.

[8] J.F. Bonnans, A. Shapiro, *Perturbation analysis of optimization problems.* Springer, 2000.

[9] D. Bucur, G. Buttazzo, *Variational methods in shape optimization problems.* Birkhäuser, 2005.

[10] C. Christof, C. Meyer, S. Walther, C. Clason, *Optimal control of a non-smooth semilinear elliptic equation.* Math. Contr. Rel. Fields **8** (2018), 247–276.

[11] F.H. Clarke, *Optimization and nonsmooth analysis.* SIAM, 1998.

[12] G. Dal Maso, U. Mosco, *Wiener's criterion and* Γ*-convergence.* Appl. Math. Optim. **15** (1987), 15–63.

[13] A. Daniilidis, P. Georgiev, *Approximate convexity and submonotonicity.* J. Math. Anal. Appl. **291** (2004), 292–301.

[14] F. Facchinei, J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems.* Springer, 2007.

[15] A. Friedman, *Variational principles and free-boundary problems.* Courier Corporation, 2010.

[16] J. Gwinner, *A note on random variational inequalities and simple random unilateral boundary value problems: well-posedness and stability results.* In: Advances in convex analysis and global optimization, N. Hadjisavvas, P.M. Pardalos, eds., Springer (2001), 531–543. Springer US, Boston, MA, 2001.

[17] J. Gwinner, F. Raciti, *On a class of random variational inequalities on random sets.* Numer. Funct. Anal. Optim. **27** (2006), 619–636.

[18] F. Harder, G. Wachsmuth, *Comparison of optimality systems for the optimal control of the obstacle problem.* GAMM-Mitt. **40** (2018), 312–338.

[19] W. Hare, C. Sagastizábal and M. Solodov, *A proximal bundle method for nonsmooth nonconvex functions with inexact information.* Comput. Optim. Appl. **63** (2016), 1–28.

[20] J. Haslinger, T. Roubíček, *Optimal control of variational inequalities. Approximation theory and numerical realization.* Appl. Math. Optim. **14** (1986), 187–201.

[21] J. Heinonen, T. Kilpeläinen, O. Martio, *Nonlinear potential theory of degenerate elliptic equations.* Oxford University Press, 1993.

[22] L. Hertlein, M. Ulbrich, *An inexact bundle algorithm for nonconvex nonsmooth minimization in Hilbert space.* SIAM J. Control Optim., **57** (2019), 3137-3165.

[23] S.H. Hou, *On property* (Q) *and other semicontinuity properties of multifunctions.* Pacific J. Math. **103** (1982), 39–56.

[24] T. Hytönen, J. van Neerven, M. Veraar, L. Weis, *Analysis in Banach spaces, Vol. I: Martingales and Littlewood-Paley theory.* Springer, 2016.

[25] T. Kilpelainen, J. Malý, *Supersolutions to degenerate elliptic equations on quasi open sets.* Comm. Partial Differential Equations **17** (1992), 371–405.

[26] D. Kinderlehrer, G. Stampaccia, *An introduction to variational inequalities and their applications.* SIAM, 2000.

[27] J. Lv, L.-P. Pang, F.-Y. Meng, *A proximal bundle method for constrained nonsmooth nonconvex optimization with inexact information.* J. Glob. Optim. **70** (2018), 517–549.

[28] R. Mifflin, *An algorithm for constrained optimization with semismooth functions.* Math. Oper. Res. **2** (1977), 191–207.

[29] F. Mignot, *Contrôle dans les inéquations variationelles elliptiques.* J. Funct. Anal. **22** (1976), 130–185.

[30] U. Mosco *Convergence of convex sets and of solutions of variational inequalities.* Adv. Math. **3** (1969), 510–585.

[31] D. Noll, *Bundle method for non-convex minimization with inexact subgradients and function values.* In: Computational and analytical mathematics, D. Bailey, H. Bauschke, et al., eds., Springer (2013), 555–592.

[32] J. Outrata, M. Kočvara, J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints. Theory, applications and numerical results.* Springer, 2013.

[33] N.S. Papageorgiou, *On the theory of Banach space valued multifunctions. I. Integration and conditional expectation.* J. Multivariate Anal. **17** (1985), 185–206.

[34] N.S. Papageorgiou, *On Fatou's lemma and parametric integrals for set-valued functions.* Proc. Indian Acad. Sci. Math. Sci., **103** (1993), 181–195.

[35] A.-T. Rauls, S. Ulbrich, *Computation of a Bouligand generalized derivative for the solution operator of the obstacle problem.* SIAM J. Control Optim., **57** (2019), 3223–3248.

[36] A.-T. Rauls, G. Wachsmuth *Generalized derivatives for the solution operator of the obstacle problem.* Set-Valued Var. Anal. (2018), 1–27.

[37] J.-F. Rodrigues, *Obstacle problems in mathematical physics.* Elsevier, 1987.

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

[39] T.M. Surowiec, *Numerical optimization methods for the optimal control of elliptic variational inequalities.*, In: Frontiers in PDE-constrained optimization, H. Antil, D.P. Kouri, et al., eds., Springer (2018), 123–170.

[40] N. N. Vakhania, *The topological support of Gaussian measure in Banach space.* Nagoya Math. J. **57** (1975), 59–63.

[41] B. Velichkov, *Existence and regularity results for some shape optimization problems.* Springer, 2015.

[42] G. Wachsmuth, *Strong stationarity for optimal control of the obstacle problem with control constraints.* SIAM J. Optim. **24** (2014), 1914–1932.

[43] H. Xu, D. Zhang, *Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications.* Math. Program. **119** (2009), 371–401.

Lukas Hertlein
Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany
e-mail: hertlein@ma.tum.de

Anne-Therese Rauls
Technische Universität Darmstadt, Dolivostr. 15, 64293 Darmstadt, Germany
e-mail: rauls@mathematik.tu-darmstadt.de

Michael Ulbrich
Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany
e-mail: mulbrich@ma.tum.de

Stefan Ulbrich
Technische Universität Darmstadt, Dolivostr. 15, 64293 Darmstadt, Germany
e-mail: ulbrich@mathematik.tu-darmstadt.de