Optimality conditions in control problems with random state constraints in probabilistic or almost-sure form

Caroline Geiersbach, René Henrion

submitted: June 16, 2023

Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: caroline.geiersbach@wias-berlin.de
          rene.henrion@wias-berlin.de

Optimality conditions in control problems with random state constraints in probabilistic or almost-sure form

Caroline Geiersbach, René Henrion

Abstract

In this paper, we discuss optimality conditions for optimization problems subject to random state constraints, which are modeled in probabilistic or almost sure form. While the latter can be understood as the limiting case of the former, the derivation of optimality conditions requires substantially different approaches. We apply them to a linear elliptic partial differential equation (PDE) with random inputs. In the probabilistic case, we rely on the spherical-radial decomposition of Gaussian random vectors in order to formulate fully explicit optimality conditions involving a spherical integral. In the almost sure case, we derive optimality conditions and compare them to a model based on robust constraints with respect to the (compact) support of the given distribution.

1 Introduction

Uncertainty appears in many applications in optimization, from finance to medicine to engineering, due to unknown inputs or parameters in underlying systems. Ignoring uncertainty can lead to solutions that poorly reflect how the system behaves with high probability. This fact holds true no matter if the space of decisions is finite-dimensional as in operations research or infinite-dimensional as in PDE-constrained optimization. The present work concerns itself with the second context. For an introductory monograph on PDE-constrained optimization under uncertainty, we refer to [26]. One way to classify different approaches in this field consists in considering the degree of risk involved in decision making: on the one end of the spectrum, one deals with risk-neutral (expectation-based) models, e.g., [8, 12, 27]. On the other end, one is rather interested in absolutely safe decisions and addresses uncertainty via almost sure constraints [13, 14]. In between these two extremes, reasonable compromises can be made based on several risk-averse approaches such as conditional value-at-risk (CVaR), value-at-risk (VaR), stochastic dominance, or robust constraints [1, 9, 22, 23, 26]. Besides modeling, numerical solution, and structural analysis, there have been efforts in the past few years to derive optimality conditions for problems involving random PDEs, see [23] (without state constraints, meaning additional constraints on the solution of the PDE) or [12, 13, 14] (with almost sure state constraints).

In this paper, we consider optimality conditions for problems of the form

$$\min_{u \in U} \ F(u) \quad \text{subject to (s.t.)} \quad \mathbb{P}(g(u, \xi) \leq 0) \geq p \quad (p \in (0, 1]), \tag{1}$$

where U is a Banach space, ξ is some random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, g is a (random) constraint function, and p is some given probability level. Constraints like those appearing in (1) are referred to as probabilistic or chance constraints. They express the condition that a feasible decision u has to guarantee the satisfaction of the random inequality constraint $g \leq 0$ at least with probability p. This can be equivalently formulated by a value-at-risk constraint as mentioned above. Probabilistic constraints have been introduced in the context of operations research problems more

than 60 years ago by Charnes et al. [6]. Their algorithmical treatment and theoretical understanding have been fundamentally advanced by Prékopa, whose monograph [32] is still a fundamental reference on this topic in finite-dimensional optimization. Modern presentations can be found in [38] or [40], respectively. In recent years, there has been growing interest in considering probabilistic constraints in the framework of optimal control in general or PDE-constrained optimization in particular, e.g., [5, 10, 11, 15, 16, 24, 29, 39]. These works include proposals for numerical approaches as well as structural investigations.

The abstract problem (1) encompasses PDE-constrained optimization subject to probabilistic state constraints. The contribution of the present work will consist in deriving explicit optimality conditions for problems of this type, where the state is the solution of a random linear elliptic PDE. The challenge here lies first in providing the optimality conditions for the problem as is (i.e., without an approximation and without the use of asymptotic arguments). As a second challenge, the random state constraints involved are uniform with respect to the domain. This leads to probabilistic constraints that are joint with respect to an infinite number of random inequalities. For such constraints, the name "probust"(probabilistic/robust) was coined. Their theoretical analysis [10, 42] and algorithmic treatment [3] is still in its infancy and Section 2 of the present work makes a contribution to these investigations as well. Our workhorse for the derivation of optimality conditions will be the so-called *spherical-radial decomposition* of Gaussian random vectors (which actually generalizes to any elliptically-symmetric distribution). This tool has proven to be useful in many applications involving the numerical treatment of probabilistic constraints (e.g., [3, 11]) as well as for the analytic derivation of (sub-) gradients of the decision-dependent probability function (e.g., [18, 41]). As a typical outcome, these (sub-) gradients are represented as spherical integrals, which makes them—contrary to derivative formulae in which the domain of integration is decision-dependent—amenable to numerical approximation and theoretical manipulation.

It is noteworthy that admitting a probability level of $p = 1$ in (1) allows one (at least formally) to observe almost sure constraints of the type "$g(u, \xi) \leq 0$ $\mathbb{P}$-a.s." in the model. We shall see, however, that the derivation of optimality conditions in Section 2 requires a restriction of probability levels to values strictly smaller than one. It does not make sense either to consider optimality conditions in the limit when driving p towards one because then the Lagrange multiplier tends to infinity; see Example 3.1. Therefore, in order to arrive at optimality conditions for almost sure constraints, one has to rather use functional-analytic approaches originally developed for two-stage stochastic programming [33, 34, 35, 36] and further for problems in PDE-constrained optimization in [13, 14]. On the other hand, almost sure constraints are closely tied with robust constraints of the type "$g(u, z) \leq 0$" for all z belonging to the support of the random vector. This opens yet another perspective to the derivation of optimality conditions in case one is interested in absolute safety. In Section 3, we will shed some light on various approaches to optimality conditions in such models and their comparison in the context of the PDE mentioned above.

1.1 Notation

The dual space of a Banach space X is denoted by X^*. Throughout, the duality pairing $\langle \cdot, \cdot \rangle_{X^*, X}$ will be written as $\langle \cdot, \cdot \rangle$ since the underlying spaces can be easily guessed by the elements used in the pairing. Weak convergence is denoted by $\rightharpoonup$. A norm on a Banach space X is denoted by $\|\cdot\|_X$ and the Euclidean norm is denoted by $\|\cdot\|_2$.

Given a set $D \subset \mathbb{R}^d$, the space $L^p(D)$ denotes the usual Lebesgue space with the Lebesgue measure on D. For $q \in [1, \infty]$, the Sobolev space $W^{1,q}(D)$ is given by the set of $L^q(D)$ functions

with weak derivatives in $L^q(D)$; the set $W_0^{1,q}(D)$ is the set of $W^{1,q}(D)$ functions that vanish on the boundary ∂D. We have $H^1(D) = W^{1,2}(D)$ and the dual space $W^{-1,q}(D) := (W^{1,q'}(D))^*$ with $\frac{1}{q} + \frac{1}{q'} = 1$.

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Banach space X, the Bochner space $L_\mathbb{P}^r(\Omega, X) := L^r(\Omega, \mathcal{F}, \mathbb{P}; X)$ is the set of all (equivalence classes of) strongly measurable functions $y \colon \Omega \to X$ having finite norm, where the norm is given by

$$\|y\|_{L_\mathbb{P}^r(\Omega, X)} := \begin{cases} (\int_\Omega \|y(\omega)\|_X^r \, \mathrm{d}\mathbb{P}(\omega))^{1/r}, & r < \infty \\ \operatorname{ess\,sup}_{\omega \in \Omega} \|y(\omega)\|_X, & r = \infty \end{cases}.$$

Given a compact topological space K and a Banach space X, the set $\mathcal{C}(K, X)$ is the set of continuous functions from K into X, where the norm is given by $\|f\|_{\mathcal{C}(K,X)} := \sup_{z \in K} \|f(z)\|_X$. If $X = \mathbb{R}$, we write $\mathcal{C}(K)$ as a shorthand for $\mathcal{C}(K, \mathbb{R})$. The set $\mathcal{C}^{0,\alpha}(K)$ denotes the space of α-Hölder continuous functions on D for $\alpha \in (0, 1)$.

Given Banach spaces Y and Z, the set of bounded linear operators from Y to Z is denoted by $\mathcal{L}(Y, Z)$. The Fréchet derivative of $F \colon X \to \mathbb{R}$ is denoted by $DF \colon X \to \mathcal{L}(X, \mathbb{R}) = X^*$. For appropriate functions f defined on X, we denote by $\partial f, \partial^C f, \partial^F f, \partial^M f$ the classical subdifferential of convex analysis, the Clarke subdifferential, the Fréchet subdifferential, and the Mordukhovich subdifferential, respectively. For the definition of these objects, we refer the reader to the standard monographs [7, 28, 30]. Partial subdifferentials are understood as subdifferentials of partial functions. For a locally Lipschitzian function $f : X \to \mathbb{R}$, we denote by

$$f^0(x; h) := \limsup_{x' \to x, t \downarrow 0} \frac{f(x' + th) - f(x')}{h}$$

its Clarke directional derivative at $x \in X$ in direction $h \in X$. We recall that a locally Lipschitzian function is called Clarke regular at some point if for all directions its ordinary directional derivative exists at that point and coincides with its Clarke directional derivative.

1.2 A model PDE

As mentioned in the introduction, we will focus on the derivation of optimality conditions for a problem involving a random linear elliptic PDE. In this section, we will introduce this PDE and provide regularity results that will be of use later. We first consider the following parametrized elliptic PDE

$$\begin{aligned} -\Delta \hat{y}(x, z) &= u(x) + f(x, z), & (x, z) \in D \times \Xi, \\ \hat{y}(x, z) &= 0, & (x, z) \in \partial D \times \Xi. \end{aligned} \tag{2}$$

Here, D is an open bounded subset of $\mathbb{R}^d$ with the boundary ∂D and Ξ is a (possibly unbounded) subset of $\mathbb{R}^m$. The Laplacian $\Delta y = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} y$ acts only on the variable x. For the first result, we will require the domain to be of *class S*, meaning that there exist constants $\gamma \in (0, 1)$ and $r_0 > 0$ such that $\operatorname{meas}(B_r(x) \backslash D) \geq \gamma \operatorname{meas}(B_r(x))$ for all $x \in \partial D$ and for all $r < r_0$. This classical definition from [21] prevents inward cusps of D, and is a mild requirement that covers many domains of interest from the literature, such as Lipschitz and convex domains. We impose the following assumptions on problem (2).

Assumption 1.1. *The open and bounded set $D \subseteq \mathbb{R}^d$ ($d = 1, 2, 3$) is of class S. Additionally, $u \in L^2(D)$ and the function $f\colon \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ is defined by*

$$f(x, z) := f_0(x) + \sum_{i=1}^{m} z_i \phi_i(x) \tag{3}$$

for some given $f_0, \phi_i \in L^2(D)$.

The choice (3) will be motivated later in Remark 1.3. First, we establish well-posedness of (2) under this assumption.

Lemma 1.2. *Suppose Assumption 1.1 holds. Then for any $u, f_0, \phi_i \in L^2(D)$ ($i = 1, \dots, m$) and every $z \in \Xi$, there exists a unique solution $\hat{y}(\cdot, z) \in H_0^1(D) \cap \mathcal{C}(\bar{D})$ of (2). Moreover, there exists a constant $C > 0$ such that*

$$\|\hat{y}(\cdot, z)\|_{\mathcal{C}(\bar{D})} \le C(\|z\|_2 + \|u + f_0\|_{L^2(D)}). \tag{4}$$

Proof. By the Lax–Milgram lemma, the operator $-\Delta$ defines an isomorphism between $H_0^1(D)$ and $H^{-1}(D)$. In particular, problem (2) has a unique solution $\hat{y}(\cdot, z) \in H_0^1(D)$ for every $z \in \Xi$ and every $u, f_0, \phi_i \in L^2(D)$ ($i = 1, \dots, m$). For $d = 1$, (4) follows trivially. For $d = 2, 3$, the result follows by [21, Chapter II, Appendix B and C]. $\square$

It is worth mentioning that [21] gives an even stronger result than we need here, namely Hölder continuity of solutions as well as continuity of the operator $A_q^{-1}\colon W^{-1,q}(D) \to \mathcal{C}^{0,\alpha}(\bar{D})$ for some $\alpha \in (0, 1)$, A_q being the part of $-\Delta$ in $W^{-1,q}(D)$ ($q = 4$). Such regularity can be obtained for more general elliptic PDEs than presented here; see [17] and the references therein. Due to the continuity of the embeddings $\iota_1\colon L^2(D) \to W^{-1,q}(D)$ (for $d = 1, 2, 3$) and $\iota_2\colon \mathcal{C}^{0,\alpha}(\bar{D}) \to \mathcal{C}(\bar{D})$, the operator $A\colon L^2(D) \to \mathcal{C}(\bar{D})$ defined by $A := \iota_2 \circ A_q \circ \iota_1$ has a continuous inverse. With this operator, we can define the parametrized control-to-state operator $S\colon L^2(D) \times \mathbb{R}^m \to \mathcal{C}(\bar{D})$ by

$$S(u, z) := A^{-1}(u + f(\cdot, z)) =: P(u, z) + y_0. \tag{5}$$

Here, $P(u, z) = A^{-1}u + \sum_{i=1}^{m} z_i A^{-1}\phi_i$ and $y_0 = A^{-1}f_0$. Due to (4), we have

$$\|S(u, z)\|_{\mathcal{C}(\bar{D})} \le C(\|z\|_2 + \|u + f_0\|_{L^2(D)}), \tag{6}$$

making S a continuous operator. Applying (6) to the special case $f_0 = 0$, we also have

$$\|P(u, z)\|_{\mathcal{C}(\bar{D})} \le C(\|z\|_2 + \|u\|_{L^2(D)}) \tag{7}$$

so that P is a continuous linear operator.

Now, suppose $\xi = (\xi_1, \dots, \xi_m)\colon \Omega \to \mathbb{R}^m$ is a random vector. We will consider the following elliptic PDE with a random right-hand side generated by ξ:

$$\begin{aligned} -\Delta y(x, \omega) &= u(x) + f(x, \xi(\omega)), \quad &x \in D \quad &\mathbb{P}\text{-a.s.}, \\ y(x, \omega) &= 0, \quad &x \in \partial D \quad &\mathbb{P}\text{-a.s.} \end{aligned} \tag{8}$$

Remark 1.3. *The random source term from (8) in combination with the structure given by (3) might result, for instance, from the truncation of a Karhunen–Loève expansion of a random field on D. Since the random vector ξ has images in $\mathbb{R}^m$, the source term also satisfies the finite-dimensional noise assumption frequently employed in numerical simulations. It is worth mentioning that, for the purpose of this work, we exclude the case of random diffusion terms and that this, in combination with the structure afforded by (3), will be needed to prove convexity of the probability function in Section 2.2.*

Finally, the following measurability statement holds true and thus justifies the setting of (8).

Lemma 1.4. *For each $x \in \bar{D}$ the mapping $\omega \mapsto y(x, \omega)$ is measurable.*

Proof. Thanks to $y(x, \omega) = [S(u, \xi(\omega))](x)$ and to ξ being measurable as a random vector, the claim follows by the continuity of the mappings $z \mapsto [S(u, z)](x)$ for each $x \in \bar{D}$. To see this, let $z_n \to z$ and observe that, by linearity of P and by (7),

$$|[S(u, z)](x) - [S(u, z_n)](x)| = |[P(0, z - z_n)](x)| \leq C \|z - z_n\|_2 \to 0 \quad \forall x \in \bar{D}.$$

$\square$

2 Probabilistic constraints in optimal control

Before dealing with probabilistic state constraints in the application, we collect some general results on probabilistic constraints.

2.1 General framework

We consider the following general optimization problem with probabilistic constraints:

$$\min_{u \in U} \; F(u) \text{ s.t. } \varphi(u) \geq p \quad (p \in (0, 1]). \tag{9}$$

Here, U is a reflexive and separable Banach space, $F : U \to \mathbb{R}$ is some convex, Fréchet differentiable cost function, and $\varphi : U \to \mathbb{R}$ denotes a probability function defined by

$$\varphi(u) := \mathbb{P}(\omega \mid g(u, \xi(\omega)) \leq 0).$$

In the definition above, ξ is an m-dimensional Gaussian random vector defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having a centered Gaussian distribution $\mathcal{N}(0, \Sigma)$ with covariance matrix Σ, and $g : U \times \mathbb{R}^m \to \mathbb{R}$ is some constraint function. As φ is not analytically given, it is important to derive some analytical properties (Lipschitz continuity, convexity) from the given data (g, ξ). To this end, we make the following general assumption on g:

$$g \text{ is locally Lipschitzian and } g(u, \cdot) \text{ is convex for all } u \in U. \tag{GA}$$

In the following, we will fix a point of interest $\bar{u} \in U$ and require two additional assumptions at that point. The first one requires that the mean zero of the distribution of ξ is a Slater point of the convex inequality $g(\bar{u}, z) \leq 0$:

$$g(\bar{u}, 0) < 0. \tag{10}$$

We note that this assumption is not restrictive. Indeed, requiring a Slater point is necessary in order to guarantee that the probability function φ is at least continuous. But, if some Slater point exists at all, then the mean zero will be a Slater point too whenever $\varphi(\bar{u}) \geq 0.5$, see [41, Proposition 3.11]. On the other hand, $\varphi(\bar{u})$ is typically close to one in probabilistic programming.

Our general assumptions (GA) along with (10) guarantee that φ is (strongly) continuous at $\bar{u}$ (see, e.g., [18, Theorem 1]). However, the fact that g is even locally Lipschitzian does not yet imply the same property for φ despite the nice probability distribution (multivariate Gaussian), see [18, Example

1]. Similarly, Fréchet differentiability of g does not imply the same property of φ, see [41, Proposition 2.2]. These counterexamples are based on the fact that at the point of interest $\bar{u} \in U$, the set

$$M := \{z \in \mathbb{R}^m \mid g(\bar{u}, z) \leq 0\} \tag{11}$$

of feasible realizations of the random vector may be unbounded. In order to derive the local Lipschitz continuity of φ in our setting and then to estimate its Clarke subdifferential, we need the following condition of moderate growth to be satisfied at $\bar{u}$:

$$\exists l > 0 \ \forall h \in U : g^\circ(\cdot, z)(u; h) \leq l \, \|z\|_2^{-m} \exp\left(\frac{\|z\|_2^2}{2 \, \|\Sigma^{1/2}\|^2}\right) \|h\|_U \tag{12}$$
$$\forall u \in \mathbb{B}_{1/l}(\bar{u}), \forall z : \|z\|_2 \geq l.$$

Here, $g^\circ(\cdot, z)(u; h)$ refers to the Clarke directional derivative of the locally Lipschitzian (by (GA)) partial function $g(\cdot, z)$ at the argument u in direction h. Moreover, $\Sigma^{1/2}$ denotes a root of Σ.

Next, we introduce a radial probability function $e \colon U \times \mathbb{S}^{m-1} \to \mathbb{R}$ defined by

$$e(u, v) := \mu_\eta \{r \geq 0 \mid g(u, r\Sigma^{1/2}v) \leq 0\}, \tag{13}$$

where μ_η is the one-dimensional chi distribution with m degrees of freedom. The following representation of the total probability function φ as a spherical integral over the radial probability function e is a consequence of the well-known *spherical radial decomposition* of Gaussian random vectors:

$$\varphi(u) = \int_{\mathbb{S}^{m-1}} e(u, v) \, \mathsf{d}\mu_\zeta(v) \quad (u \in U). \tag{14}$$

The following result on subdifferentiation under the integral sign holds true:

Theorem 2.1 ([18], Theorem 5, Corollary 2 and Proposition 6)**.** *In addition to the basic assumptions (GA) let (10) and (12) be satisfied at some $\bar{u} \in U$ (instead of (12), we may alternatively assume that the set M in (11) is bounded). Then, the partial radial probability functions $e(\cdot, v)$, $v \in \mathbb{S}^{m-1}$, are uniformly locally Lipschitzian around $\bar{u}$ with a Lipschitz constant independent of v. Moreover, the (total) probability function φ is locally Lipschitzian around $\bar{u}$ and its Clarke subdifferential at $\bar{u}$ can be estimated from above by*

$$\partial^C \varphi(\bar{u}) \subseteq \int_{\mathbb{S}^{m-1}} \partial_u^C e(\bar{u}, v) \, \mathsf{d}\mu_\zeta(v). \tag{15}$$

Here, μ_ζ refers to the uniform distribution on $\mathbb{S}^{m-1}$, ∂^C is the Clarke subdifferential, and ∂_u^C to the partial Clarke subdifferential with respect to u. If, in addition, the condition

$$\mu_\zeta \{v \in \mathbb{S}^{m-1} \mid \#\partial_u^C e(\bar{u}, v) \geq 2\} = 0 \tag{16}$$

is satisfied, then the Clarke subdifferential of φ at $\bar{u}$ reduces to a singleton and equality holds in (15). As a consequence, φ is strictly differentiable at $\bar{u}$ in the Hadamard sense [7, p. 30].

The integral above is to be understood in the sense of Aumann. In particular, the asserted inclusion means that for each $x^* \in \partial^C \varphi(\bar{u})$ there exists a measurable selection $\hat{x}_v^* \in \partial_u^C e(\bar{u}, v)$ (for μ_ζ-a.e. $v \in \mathbb{S}^{m-1}$) such that

$$x^*(h) = \int_{\mathbb{S}^{m-1}} \hat{x}_v^*(h) \, \mathsf{d}\mu_\zeta(v) \quad \forall h \in U.$$

At this point one might wonder why the inclusion in Theorem 2.1 is not sharp apart from the differentiable situation under condition (16). The reason lies in Clarke's theorem on subdifferentiation of integral functionals [7, Theorem 2.7.2], which provides a sharp estimate only in case that the integrand is Clarke regular. It is easily seen that the radial probability function fails to be Clarke regular in general. However, at this point one may pick up a remark by Clarke [7, Remark 2.3.5] pointing to the fact that sharp results may be shown also for functions being the negative of a Clarke regular one. This observation allows us in the following, by slightly strengthening our assumptions, to derive an identity rather than upper estimate in the formula of Theorem 2.1. In order to prepare this argument, we observe that by continuity, (10) holds locally, i.e., there exists some neighborhood $\mathcal{N}(\bar{u})$ of $\bar{u}$ with $g(u, 0) < 0$ for all $u \in \mathcal{N}(\bar{u})$. Then, thanks to the convexity assumption in (GA), for each $u \in \mathcal{N}(\bar{u})$ and each $v \in \mathbb{S}^{m-1}$, there exists a unique (possibly infinite) $\rho(u, v)$ such that

$$\{r \geq 0 \mid g(u, r\Sigma^{1/2}v) \leq 0\} = [0, \rho(u, v)]. \tag{17}$$

If, in particular, $\rho(u, v) < \infty$, then this value is the unique solution in r of the equation

$$g(u, r\Sigma^{1/2}v) = 0. \tag{18}$$

In other words, we have defined a "radius function" $\rho \colon \mathcal{N}(\bar{u}) \times \mathbb{S}^{m-1} \to \mathbb{R} \cup \{\infty\}$ such that the radial probability function e from (13) may be represented locally around $\bar{u}$ by

$$e(u, v) = \begin{cases} F_\eta(\rho(u, v)) & \text{if } \rho(u, v) < \infty \\ 1 & \text{if } \rho(u, v) = \infty \end{cases} \quad \forall u \in \mathcal{N}(\bar{u}) \; \forall v \in \mathbb{S}^{m-1}, \tag{19}$$

where F_η refers to the cumulative distribution function of the (one-dimensional) chi distribution with m degrees of freedom. We are now in a position to establish the Clarke regularity of the negative radial probability function.

Lemma 2.2. *In addition to the general assumptions (GA), suppose that the constraint function g is jointly convex in both variables. Consider some $\bar{u} \in U$ with (10) and assume that the set M in (11) is bounded. Then, for each $v \in \mathbb{S}^{m-1}$, the function $-e(\cdot, v)$ is Clarke regular on the open neighborhood $\mathcal{N}(\bar{u})$ introduced above.*

Proof. By our assumptions and by (19), the radial probability function has the representation

$$e(u, v) = F_\eta(\rho(u, v)) \quad \forall u \in \mathcal{N}(\bar{u}) \; \forall v \in \mathbb{S}^{m-1}.$$

We observe first that for each fixed $v \in \mathbb{S}^{m-1}$, the partial radius function $\rho(\cdot, v) \colon \mathcal{N}(\bar{u}) \to \mathbb{R}$ is concave. Indeed, let $u^1, u^2 \in \mathcal{N}(\bar{u})$ and $\lambda \in [0, 1]$ be arbitrary. Then, by (18),

$$g(u^1, \rho(u^1, v)\Sigma^{1/2}v) = g(u^2, \rho(u^2, v)\Sigma^{1/2}v) = 0$$

and thus, by convexity of g,

$$g(\lambda u^1 + (1 - \lambda)u^2, \lambda \rho(u^1, v)\Sigma^{1/2}v + (1 - \lambda)\rho(u^2, v)\Sigma^{1/2}v) \leq 0.$$

Now, (17) yields that

$$\lambda \rho(u^1, v) + (1 - \lambda)\rho(u^2, v) \leq \rho(\lambda u^1 + (1 - \lambda)u^2, v).$$

This shows the asserted concavity statement. Next, we extend the cumulative distribution function for the chi distribution from $\mathbb{R}_+$ to $\mathbb{R}$ via

$$F_\eta^*(t) := \begin{cases} F_\eta(t) & t \geq 0, \\ -F_\eta(-t) & t < 0. \end{cases}$$

Since for $t > 0$, the derivative of F_η equals the probability density function for the chi distribution, which for arbitrary degrees of freedom is a continuous function, it follows readily by construction that F_η^* is continuously differentiable. Moreover, $F_\eta'(t) \geq 0$ for all $t \geq 0$ because F_η is a distribution function. This implies that $(F_\eta^*)'(t) \geq 0$ for all $t \in \mathbb{R}$. We now have that

$$-e(u, v) = -F_\eta(\rho(u, v)) = F_\eta^*(-\rho(u, v)) \quad \forall u \in \mathcal{N}(\bar{u}) \; \forall v \in \mathbb{S}^{m-1}.$$

Let us fix an arbitrary $v \in \mathbb{S}^{m-1}$. As we have shown that $-\rho(\cdot, v) \colon \mathcal{N}(\bar{u}) \to \mathbb{R}$ is convex, the same function is Clarke regular [7, Proposition 2.3.6]. Therefore, $-e(u, v)$ is the composition of a continuously differentiable, hence Clarke regular (see [7, Proposition 2.3.6]) function having non-negative derivative everywhere with another Clarke regular function and is therefore itself Clarke regular [7, Theorem 2.3.9 (i)]. □

Theorem 2.3. *Under the assumptions of Lemma 2.2, one has the exact formula*

$$\partial^C \varphi(\bar{u}) = \int\limits_{\mathbb{S}^{m-1}} \partial_u^C e(\bar{u}, v) \, d\mu_\zeta(v).$$

Proof. Under the assumptions we made, the function e is continuous on $\mathcal{N}(\bar{u}) \times \mathbb{S}^{m-1}$ [18, Proposition 1]. In particular, $-e(u, \cdot) \colon \mathbb{S}^{m-1} \to \mathbb{R}$ is measurable for each $u \in \mathcal{N}(\bar{u})$. Moreover, according to Theorem 2.1, the functions $e(\cdot, v)$ and, hence, $-e(\cdot, v)$ with $v \in \mathbb{S}^{m-1}$, are uniformly locally Lipschitzian around $\bar{u}$ with a Lipschitz constant independent of v. Since constants are integrable with respect to the uniform measure on the sphere, all assumptions of Theorem 2.72 in [7] are satisfied for the relation

$$-\varphi(u) = \int\limits_{\mathbb{S}^{m-1}} -e(u, v) \, d\mu_\zeta(v) \quad (u \in U),$$

which is equivalent to (14). Accordingly, the mentioned theorem ensures the inclusion

$$\partial^C(-\varphi)(\bar{u}) \subseteq \int\limits_{\mathbb{S}^{m-1}} \partial_u^C(-e)(\bar{u}, v) \, d\mu_\zeta(v).$$

By the Clarke regularity of the functions $-e(\cdot, v)$ on $\mathcal{N}$ for each $v \in \mathbb{S}^{m-1}$ (shown in Lemma 2.2), the same theorem even guarantees that this inclusion reduces to an equality. Then, however, we may exploit the linearity of Clarke's subdifferential in order to show that

$$\partial^C \varphi(\bar{u}) = -\partial^C(-\varphi)(\bar{u}) = -\int\limits_{\mathbb{S}^{m-1}} \partial_u^C(-e)(\bar{u}, v) \, d\mu_\zeta(v) = \int\limits_{\mathbb{S}^{m-1}} \partial_u^C e(\bar{u}, v) \, d\mu_\zeta(v).$$

□

Apart from Lipschitz continuity or even differentiability of the constraint, it would be interesting to know what convexity properties are inherent to problem (9). Given the assumed convexity of the objective, it is tempting to verify the concavity of φ in order to arrive at a fully convex problem. Unfortunately, being bounded by zero and one, probability functions are hardly ever concave (they would have to be constant then which is extremely untypical). Nonetheless, some generalized concavity, so-called log-concavity, may hold true, which allows one to reformulate the problem as a convex one. Indeed, one may derive (e.g., [10, Proposition 4]) from well-known results by Prékopa (see [32, Theorem 10.2.1]), the following statement Hent:

If g is quasi-convex (jointly) in both variables and if the density of ξ is log-concave, then the probability function φ is log-concave too, i.e., $\log \varphi$ is concave under the convention $\log 0 := \infty$.

Lemma 2.4. *If g is quasi-convex (jointly) in both variables, then the optimization problem* (9) *can be reformulated as the convex optimization problem*

$$\min_{u \in U}\{F(u) \mid \tilde{\varphi}(u) \leq 0\} \quad (p \in (0,1]),\tag{20}$$

where $\tilde{\varphi}\colon U \to \mathbb{R} \cup \{\infty\}$ is an extended-valued convex function defined by $\tilde{\varphi}(u) := -\log\varphi(u) + \log p$ for $u \in U$ under the convention $-\log 0 := \infty$. Moreover, the function $\tilde{\varphi}$ is lower semicontinuous.

Proof. Our conventions on the $\log$ along with the assumption that $p > 0$ yield the equivalence

$$\varphi(u) \geq p \iff -\log\varphi(u) + \log p \leq 0 \quad \forall u \in U.$$

Consequently, (9) and (20) are equivalent optimization problems. Since the density of our Gaussian random vector ξ is log-concave, the statement preceding this lemma yields that $\tilde{\varphi}$ is (possibly extended-valued) convex and, hence, (20) is a convex optimization problem. As mentioned above, our general assumptions (GA) along with (10) guarantee that the probability function φ is strongly continuous. As a consequence, for each $t \in \mathbb{R}$, the upper level sets

$$\{u \in U \mid \varphi(u) \geq t\}$$

are (strongly) closed in U. Now, let $\tau \in \mathbb{R}$ be arbitrarily given. Then, the identity (again thanks to $p > 0$)

$$\{u \in U \mid \tilde{\varphi}(u) \leq \tau\} = \{u \in U \mid \varphi(u) \geq pe^{-\tau}\}$$

yields the closedness of the left-hand side set thanks to the closedness of the right-hand side upper level set (with respect to the level $t := pe^{-\tau}$). Since τ was arbitrary, we have shown that all lower level sets of $\tilde{\varphi}$ are (strongly) closed, which is the same as saying that $\tilde{\varphi}$ is (strongly) lower semicontinuous on U. $\qquad\square$

Now, we are in a position to formulate necessary and sufficient optimality conditions for the originally nonconvex problem (9).

Proposition 2.5. *We strengthen the general assumptions (GA) by assuming that the constraint function g is jointly convex in <u>both</u> variables. Let there exist some $\hat{u} \in U$ with $\varphi(\hat{u}) > p$ (generalized Slater point). Consider some $u^* \in U$ that is feasible in* (9) *and satisfies assumptions* (10) *and* (12) *(instead of* (12)*, we may alternatively assume that the set M in* (11) *is bounded). Then, for u^* to be a solution to* (9)*, it is necessary and sufficient that there exists some $\lambda \geq 0$ with*

$$DF(u^*) \in \lambda\partial^C\varphi(u^*), \quad \lambda(\varphi(u^*) - p) = 0.\tag{21}$$

In particular, for u^* to be a solution to (9)*, it is necessary that there exists some $\lambda \geq 0$ such that*

$$DF(u^*) \in \lambda \int_{\mathbb{S}^{m-1}} \partial_u^C e(u^*, v)\, d\mu_\zeta(v), \quad \lambda(\varphi(u^*) - p) = 0,\tag{22}$$

where e refers to the radial probability function introduced in Theorem 2.1. If, moreover, (16) *or the conditions of Lemma 2.2 are satisfied, then* (22) *is also sufficient for u^* to be a solution to* (9)*.*

Proof. The point u^* being a solution to (9) is equivalent to it being a solution to the convex problem (20). Due to $\tilde{\varphi}(\hat{u}) = -\log\varphi(\hat{u}) + \log p < 0$, $\hat{u}$ is a Slater point for (20). Hence, by Lemma 2.4, $\tilde{\varphi}$ is a lower semicontinuous, convex and proper function. Now, it follows from classical convex analysis (admitting extended-valued inequality constraints, e.g., [30, Theorem 3.66, Remark 3.67]) that u^* being a solution to (20) (and (9)) is equivalent to the existence of some $\lambda \geq 0$ such that $-DF(u^*) \in \lambda\partial\tilde{\varphi}(u^*)$ and $\lambda\tilde{\varphi}(u^*) = 0$, where "$\partial$" refers to the subdifferential in the sense of convex analysis. Now, the feasibility of u^* implies that $\varphi(u^*) \geq p > 0$. With φ being locally Lipschitzian around u^* (by Theorem 2.1)) and $\log$ being locally Lipschitzian around the positive number $\varphi(u^*)$, the function $\tilde{\varphi}$ is locally Lipschitzian around u^*. Since $\tilde{\varphi}$ is also convex, it follows that $\partial\tilde{\varphi}(u^*) = \partial^C\tilde{\varphi}(u^*)$, see [7, Proposition 2.2.7]. On the other hand, the chain rule for Clarke's subdifferential from [7, Theorem 2.3.9 (ii)] yields that

$$\partial\tilde{\varphi}(u^*) = \partial^C\tilde{\varphi}(u^*) = -\frac{1}{\varphi(u^*)}\partial^C\varphi(u^*).$$

Owing to $\varphi(u^*) > 0$, we have shown that the existence of some $\lambda \geq 0$ with $-DF(u^*) \in \lambda\partial\tilde{\varphi}(u^*)$ and $\lambda\tilde{\varphi}(u^*) = 0$ is equivalent with the existence of some $\tilde{\lambda} \geq 0$ with $DF(u^*) \in \tilde{\lambda}\partial^C\varphi(u^*)$ and $\tilde{\lambda}(\varphi(u^*) - p) = 0$. Altogether, this yields the first statement of this Proposition. The remaining statements on necessity and sufficiency of (22) are immediate consequences of (21) and of Theorems 2.1 and 2.3, respectively. □

Remark 2.6. *The existence of some $\hat{u}$ with $\varphi(\hat{u}) > p$ as required in Proposition 2.5 implicitly necessitates that $p < 1$, which means that the "almost sure" case is excluded. Without this requirement, the lack of a Slater point in the convex problem (20) would not allow the derivation of optimality conditions and simple examples show that they do not apply then, indeed. The case $p = 1$ will be addressed in Section 3 in an independent manner.*

2.2   Probabilistic state constraints in the concrete PDE

In this section, we are going to apply the general results obtained before to the specific optimization problem

$$\min_{u \in L^2(D)} F(u) \tag{23a}$$

$$\text{s.t.} \quad -\Delta y(x,\omega) = u(x) + f(x,\xi(\omega)), \quad x \in D \quad \mathbb{P}\text{-a.s.,} \tag{23b}$$

$$y(x,\omega) = 0, \qquad\qquad\qquad x \in \partial D \quad \mathbb{P}\text{-a.s.,} \tag{23c}$$

$$\mathbb{P}(y(x,\omega) \leq \alpha \quad \forall x \in D) \geq p. \tag{23d}$$

We recall the setting introduced in Section 1.2 under Assumption 1.1. This setting provides the continuity of the solution operator and the measurability of the functions $\omega \mapsto y(x,\omega)$ for each $x \in \bar{D}$ via Lemma 1.4. This justifies relations (23b) and (23c). As for (23d), a corresponding measurability statement will be made in Lemma 2.7 below. Here, $U = L^2(D)$ and $F: U \to \mathbb{R}$ is a convex, Fréchet differentiable cost function, $\xi \sim \mathcal{N}(0, \Sigma)$ is a centered m-dimensional Gaussian random vector defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $p \in (0, 1]$ is some given probability level, and $\alpha \in \mathbb{R}$ is some upper threshold for the random state $y(\cdot, \omega)$. Note that $\mathbb{E}f(\cdot, \xi) = f_0$ on account of ξ being centered (see (3)).

Recall the operators defined in (5), allowing us to reformulate problem (23) as

$$\min_{u \in L^2(D)} F(u) \quad \text{s.t.} \quad \varphi(u) \geq p, \tag{24}$$

where

$$\varphi(u) := \mathbb{P}(\omega \mid g(u, \xi(\omega)) \leq 0) \quad (u \in L^2(D));$$
$$g(u, z) := \sup_{x \in D} [S(u, z)](x) - \alpha \quad ((u, z) \in L^2(D) \times \mathbb{R}^m).$$

Problem (24) falls into the setting of the general problem (9) with introduced before. In order to apply the results from Section 2.1 to problem (24), one has to verify first that the assumptions made there are satisfied. Since by Lemma 1.2, $S(u, z) \in H_0^1(D) \cap \mathcal{C}(\bar{D})$ for all $(u, z) \in L^2(D) \times \mathbb{R}^m$, we may write g as a maximum over $\bar{D}$ rather than a supremum over D:

$$g(u, z) = \max_{x \in \bar{D}} [S(u, z)](x) - \alpha \leq 0 \quad ((u, z) \in L^2(D) \times \mathbb{R}^m).$$

Next, we verify that g satisfies the general assumption (GA) of Section 2.1:

Lemma 2.7. *The function* $g \colon L^2(D) \times \mathbb{R}^m \to \mathbb{R}$ *is convex and (globally) Lipschitzian. As a consequence, the (coinciding) events occurring in the probabilities of* (23d) *and in the definition of* φ *above are measurable.*

Proof. The convexity property follows directly from the fact that g is the maximum of affine linear functions

$$(u, z) \mapsto [S(u, z)](x) = [P(u, z)](x) + y_0(x) - \alpha \quad (x \in \bar{D}).$$

As for Lipschitz continuity, let $(u_1, z_1), (u_2, z_2) \in L^2(D) \times \mathbb{R}^m$ be arbitrary. Then,

$$|g(u_1, z_1) - g(u_2, z_2)| = \left| \max_{x \in \bar{D}} [S(u_1, z_1)](x) - \max_{x \in \bar{D}} [S(u_2, z_2)](x) \right|$$
$$= |[S(u_1, z_1)](x_1^*) - [S(u_2, z_2)](x_2^*)|,$$

where $x_1^*, x_2^* \in \bar{D}$ are arguments realizing the respective maxima. Assuming w.l.o.g. $S(u_1, z_1)](x_1^*) \geq S(u_2, z_2)](x_2^*)$, we may exploit the linearity of P and refer to (7) in order to derive that

$$|g(u_1, z_1) - g(u_2, z_2)| \leq [S(u_1, z_1)](x_1^*) - [S(u_2, z_2)](x_1^*)$$
$$= [P(u_1, z_1)](x_1^*) - [P(u_2, z_2)](x_1^*)$$
$$= [P(u_1 - u_2, z_1 - z_2)](x_1^*) \leq \|P(u_1 - u_2, z_1 - z_2)\|_{\mathcal{C}(\bar{D})}$$
$$\leq C(\|u_1 - u_2\|_{L^2(D)} + \|z_1 - z_2\|_2).$$

This proves the (global) Lipschitz continuity of g and in particular the remaining measurability statement. $\square$

The application of Theorem 2.1 requires, in particular, that the growth condition (12) is satisfied or, alternatively, that the set M in (11) is bounded. The latter condition may not hold in general as can be seen from Examples 2.15 and 2.16 below, where the radius function ρ from (17) becomes infinite for certain directions. This observation urges us to verify the general validity of the growth condition (12) before applying Theorem 2.1.

Lemma 2.8. *Let* $\bar{u} \in L^2(D)$ *be arbitrary. Then,* g *satisfies the condition of moderate growth* (12).

Proof. By definition of Clarke's directional derivative, we have that

$$g^\circ(\cdot, z)(u; h) = \limsup_{t \downarrow 0, u' \to u} \frac{g(u' + th, z) - g(u', z)}{t} \quad \forall u, h \in L^2(D) \, \forall z \in \mathbb{R}^m.$$

Exploiting Lemma 2.7 with the Lipschitz constant C derived there, we may continue as

$$g^\circ(\cdot, z)(u; h) \leq C \, \|h\|_{L^2(D)} \quad \forall u, h \in L^2(D) \, \forall z \in \mathbb{R}^m. \tag{25}$$

Define $L > 0$ such that

$$\|z\|_2^{-m} \exp\left(\frac{\|z\|_2^2}{2 \, \|\Sigma^{1/2}\|^2} \right) \geq C \quad \forall z \in \mathbb{R}^m : \|z\|_2 \geq L$$

(note that the left-hand side expression tends to infinity if $\|z\|$ does so). Then, (12) is satisfied with $l := \max\{1, L\}$. □

Corollary 2.9. *Fix some point of interest $\bar{u} \in L^2(D)$. Denote by $\bar{y} := S(\bar{u}, 0)$ the unique solution of the PDE (8) associated with $\bar{u}$ and $z := 0$ (expectation of ξ). If for this expected state, it holds that*

$$\bar{y}(x) < \alpha \quad \forall x \in \bar{D}, \tag{26}$$

then the conclusions of Theorem 2.1 hold true.

Proof. The assumption implies that

$$g(\bar{u}, 0) = \max_{x \in \bar{D}} \, [S(\bar{u}, 0)](x) - \alpha = \max_{x \in \bar{D}} \, \bar{y}(x) - \alpha < 0.$$

This, however, is (10). Since all the remaining assumptions of Theorem 2.1 are satisfied by Lemmas 2.7 and 2.8, the result follows. □

Corollary 2.10. *In the setting of Corollary 2.9 assume in addition to (26) (a Slater condition in $\mathbb{R}^m$) the existence of some $\hat{u} \in U$ with $\varphi(\hat{u}) > p$ (generalized Slater point in U). Then, the results of Proposition 2.5 hold true in the concrete optimization problem (24).*

Proof. Since the function g associated with the probability function φ in (24) is convex (in both variables simultaneously) by Lemma 2.7 and all other needed assumptions are satisfied as already stated in the proof of Corollary 2.9, the assertion follows from Proposition 2.5. □

In Section 2.4, we will provide a more explicit representation of the optimality conditions for problem (24). First, we will work on formulas of the radial probability function.

2.3 Explicit formulae for the subdifferential of the radial probability function

The general results of Theorem 2.1 and Proposition 2.5 on the subdifferential of the (total) probability function φ and on optimality conditions for problem (9) are formulated in terms of the radial probability function e and in this sense not very explicit yet. In this section we provide two lemmas that characterize the (partial) Clarke subdifferential of e in terms of the original data of our concrete problem (23) (or (24)). In both lemmas we will impose the Slater condition (26) at the point $\bar{u}$, which implies the Slater condition (10) in the general setting. Accordingly, we will consider the neighborhood $\mathcal{N}(\bar{u})$ and the radius function $\rho \colon \mathcal{N}(\bar{u}) \times \mathbb{S}^{m-1} \to \mathbb{R} \cup \{\infty\}$ introduced at the end of section 2.1. We have to distinguish two cases, namely, whether at the fixed point $\bar{u}$ the value $\rho(\bar{u}, v)$ is finite or infinite for a given direction $v \in \mathbb{S}^{m-1}$. In the latter case, the associated ray in (17) is unbounded, hence the radial probability e realizes its maximum, which is one. Consequently, the (sub-) derivative at this ray reduces to zero. This is made more precise in the following lemma.

For the remainder of Section 2, we will identify the dual of $U = L^2(D)$ with itself and work with the inner product defined by $(\cdot, \cdot)_{L^2(D)} = \langle \cdot, \cdot \rangle_{L^2(D)^*, L^2(D)}$.

Lemma 2.11. *As in Corollary* 2.9, *fix some* $\bar{u} \in L^2(D)$ *with associated expected state* $\bar{y}$ *satisfying* (26). *If* $v \in \mathbb{S}^{m-1}$ *is such that* $\rho(\bar{u}, v) = \infty$, *then* $\partial_u^C e(\bar{u}, v) = \{0\}$ *for* e *defined in* (13).

Proof. Fix v as indicated. As observed in the proof of Corollary 2.9, relation (26) yields (10). Accordingly, we may find a neighborhood $\mathcal{N}(\bar{u})$ of $\bar{u}$ with $g(u, 0) \leq \frac{1}{2} g(\bar{u}, 0) < 0$ for all $u \in \mathcal{N}(\bar{u})$. Then, by [18, Corollary 1 (ii)], one has that

$$\partial_u^F e(u, v') \subseteq \{0\} \quad \forall u \in \mathcal{N}(\bar{u}) \; \forall v' \in \mathbb{S}^{m-1} : \rho(u, v') = \infty, \tag{27}$$

where ∂_u^F denotes the partial Fréchet subdifferential with respect to u. Now, let $u \in \mathcal{N}(\bar{u})$, $h \in L^2(D)$, $v' \in \mathbb{S}^{m-1}$ with $\rho(u, v') < \infty$ and $y^* \in \partial_u^F e(u, v')$ be arbitrarily given. By [18, Theorem 2], there exists

$$(u^*, z^*) \in \partial_u^C g(u, \rho(u, v')\Sigma^{1/2}v') \times \partial_z^C g(u, \rho(u, v')\Sigma^{1/2}v')$$

such that

$$(y^*, h)_{L^2(D)} \leq \frac{\chi(\rho(u, v'))}{\langle z^*, \Sigma^{1/2}v' \rangle} (u^*, -h)_{L^2(D)},$$

where $\chi(t) := Kt^{m-1}e^{-t^2/2}$ $(t \geq 0)$ is the density of the one-dimensional chi distribution with m degrees of freedom and K is some normalizing factor. Moreover, by [18, Lemma 3],

$$\langle z^*, \Sigma^{1/2}v' \rangle \geq \frac{-g(u, 0)}{\rho(u, v')} = \frac{|g(u, 0)|}{\rho(u, v')} > 0.$$

By virtue of the definition of Clarke's generalized (partial) directional derivative and by (25), we may continue with

$$
\begin{aligned}
(y^*, h)_{L^2(D)} &\leq \frac{\chi(\rho(u, v'))}{\langle z^*, \Sigma^{1/2}v' \rangle} g^0(\cdot, \rho(u, v')\Sigma^{1/2}v')(u; -h) \\
&\leq \frac{\chi(\rho(u, v'))}{\langle z^*, \Sigma^{1/2}v' \rangle} C \|h\|_{L^2(D)} \\
&\leq \frac{\chi(\rho(u, v'))\rho(u, v')}{|g(u, 0)|} C \|h\|_{L^2(D)} \\
&\leq \frac{2\chi(\rho(u, v'))\rho(u, v')}{|g(\bar{u}, 0)|} C \|h\|_{L^2(D)}.
\end{aligned}
$$

Since $h \in L^2(D)$ was arbitrary, it follows that

$$
\begin{aligned}
\|y^*\|_{L^2(D)} \leq &\frac{2\chi(\rho(u, v'))\rho(u, v')C}{|g(\bar{u}, 0)|} \\
&\forall u \in \mathcal{N}(\bar{u}) \; \forall v' \in \mathbb{S}^{m-1} : \rho(u, v') < \infty \; \forall y^* \in \partial_u^F e(u, v').
\end{aligned} \tag{28}
$$

In order to prove our claim, it will be sufficient to show that $\partial_u^M e(\bar{u}, v) = \{0\}$, where ∂_u^M refers to the (partial) Mordukhovich subdifferential. Indeed, recalling that $e(\cdot, v)$ is locally Lipschitzian around $\bar{u}$ by Theorem 2.1 and referring to a well-known relation between the Mordukhovich and Clarke subdifferentials [28, Theorem 3.57], we infer that

$$\partial_u^C e(\bar{u}, v) = \text{clco}\{\partial_u^M e(\bar{u}, v)\} = \{0\},$$

where "clco" denotes the weak closure in $L^2(D)$. Since the Mordukhovich subdifferential is nonempty for locally Lipschitzian functions on Hilbert spaces (in general: Asplund spaces, see [28, Corollary

1.81]), it will be sufficient to prove the inclusion $\partial_u^M e(\bar{u}, v) \subseteq \{0\}$. In order to do so, choose an arbitrary $\bar{y}^* \in \partial_u^M e(\bar{u}, v)$. We have to show that $\bar{y}^* = 0$. By definition of the Mordukhovich sub-differential, there exist a strongly convergent sequence $u_k \to \bar{u}$ and a weakly (weak-star in general Asplund spaces) convergent sequence $y_k^* \rightharpoonup \bar{y}^*$ with $y_k^* \in \partial_u^F e(u_k, v)$. Now, define the subsequence u_{k_l} satisfying the condition $\rho(u_{k_l}, v) < \infty$. Then, it follows from [18, Lemma 1] that, $\rho(u_{k_l}, v) \to \infty$. On the other hand, $\lim\limits_{t \to \infty} t \cdot \chi(t) = 0$. Along with (28), this yields that $\left\| y_{k_l}^* \right\|_{L^2(D)} \to 0$ for this sub-sequence. The remaining elements of the original sequence, however, satisfy $\rho(u_k, v) = \infty$ which implies by (27) that $y_k^* = 0$ for those k. Altogether we have shown that the original sequence satisfies $\| y_k^* \|_{L^2(D)} \to 0$. Then, the weak sequential lower semicontinuity of the norm provides the desired result $\bar{y}^* = 0$. $\hfill\square$

The following lemma addresses the alternative case, where the radius function takes a finite value and—contrary to the previous situation—substantial information on the subdifferential of the radial and, thus, on the total probability function is provided.

Lemma 2.12. *As in Corollary 2.9, fix some $\bar{u} \in L^2(D)$ with associated expected state $\bar{y}$ satisfying (26). If $v \in \mathbb{S}^{m-1}$ is such that $\rho(\bar{u}, v) < \infty$, then*

$$\partial_u^C e(\bar{u}, v) = -\chi(\rho(\bar{u}, v)) \cdot \operatorname{clco} \left\{ \frac{1}{[P(0, \Sigma^{1/2} v)](x)} \cdot u_x \,\middle|\, x \in M^*(\bar{u}, v) \right\}$$

$$M^*(\bar{u}, v) := \{ x \in \bar{D} \mid [P(0, \Sigma^{1/2} v)](x) > 0, \ \bar{y}(x) + \rho(\bar{u}, v)[P(0, \Sigma^{1/2} v)](x) = \alpha \},$$

where for $x \in \bar{D}$, we denote by $u_x \in L^2(D)$ the continuous linear function defined by

$$(u_x, h)_{L^2(D)} := [P(h, 0)](x) = (A^{-1} h)(x) \quad (h \in L^2(D)).$$

Proof. Fix v as indicated and let $\mathcal{N}(\bar{u})$ be the neighborhood of $\bar{u}$ as defined above (17). With the concrete form of g in our PDE-constrained problem (23) and the affine representation (5), for each $u \in \mathcal{N}(\bar{u})$, $\rho(u, v)$ is the unique (possibly infinite) value such (17) holds true, which amounts to

$$\left\{ r \geq 0 \,\middle|\, \max_{x \in \bar{D}} [P(u, r\Sigma^{1/2} v)](x) + y_0(x) \leq \alpha \right\} = [0, \rho(u, v)] \quad \forall u \in \mathcal{N}(\bar{u}). \tag{29}$$

By definition of $\mathcal{N}(\bar{u})$ it holds that $g(u, 0) < 0$ for all $u \in \mathcal{N}(\bar{u})$ or, equivalently, that

$$\max_{x \in \bar{D}} [P(u, 0)](x) + y_0(x) < \alpha \quad \forall u \in \mathcal{N}(\bar{u}). \tag{30}$$

Taking into account the linearity of P, we may rewrite (29) as

$$\left\{ r \geq 0 \,\middle|\, r \cdot [P(0, \Sigma^{1/2} v)](x) \leq \alpha - y_0(x) - [P(u, 0)](x) \quad \forall x \in \bar{D} \right\}$$
$$= [0, \rho(u, v)] \quad \forall u \in \mathcal{N}(\bar{u}). \tag{31}$$

As a consequence of (30), we have that

$$\alpha - y_0(x) - [P(u, 0)](x) > 0 \quad \forall x \in \bar{D} \ \forall u \in \mathcal{N}(\bar{u}). \tag{32}$$

Introduce an open subset of $\bar{D}$ by

$$\mathcal{O} := \{ x \in \bar{D} \mid [P(0, \Sigma^{1/2} v)](x) > 0 \}.$$

Note that $\mathcal{O} \neq \emptyset$ because otherwise the left-hand side of (31) would be equal to $\mathbb{R}_+$ thanks to (32). Then, on the right-hand side of (31) we would end up at the contradiction $\rho(u, v) = \infty$ for all $u \in \mathcal{N}(\bar{u})$ with our assumption $\rho(\bar{u}, v) < \infty$. Now, again by virtue of (32), equation (31) can be reformulated as

$$
\begin{aligned}
&\{r \geq 0 \mid r \cdot [P(0, \Sigma^{1/2} v)](x) \leq \alpha - y_0(x) - [P(u, 0)](x) \quad \forall x \in \mathcal{O}\} \\
&= [0, \rho(u, v)] \quad \forall u \in \mathcal{N}(\bar{u}).
\end{aligned}
\tag{33}
$$

This implies that

$$
\rho(u, v) = \inf_{x \in \mathcal{O}} \gamma(u, x) \quad \forall u \in \mathcal{N}(\bar{u}),
\tag{34}
$$

where $\gamma \colon L^2(D) \times \mathcal{O} \to \mathbb{R}$ is a continuous function defined by

$$
\gamma(u, x) := \frac{\alpha - y_0(x) - [P(u, 0)](x)}{[P(0, \Sigma^{1/2} v)](x)} \quad \forall x \in \mathcal{O} \; \forall u \in L^2(D).
$$

In order to verify the continuity of γ, let $(u_n, x_n) \to (u, x)$ in $L^2(D) \times \mathcal{O}$. Then, $|\gamma(u, x) - \gamma(u, x_n)| \to 0$ by definition of the operators S and P and by the fact that $S(u', z') \in \mathcal{C}(\bar{D})$ for all (u', z') (see (5)). It remains to verify that $|\gamma(u, x_n) - \gamma(u_n, x_n)| \to 0$, which amounts to showing that

$$
\left| \frac{[P(u - u_n, 0)](x_n)}{[P(0, \Sigma^{1/2} v)](x_n)} \right| \to 0.
$$

Since $[P(0, \Sigma^{1/2} v)](x_n) \to [P(0, \Sigma^{1/2} v)](x) > 0$ by definition of $\mathcal{O}$, it suffices to show that $[P(u - u_n, 0)](x_n) \to 0$. This, however, follows readily from (7).

Our aim is to represent ρ not just as an infimum over an open set but as a minimum over a compact set. We introduce the multifunction

$$
M(u) := \{x \in \mathcal{O} \mid \rho(u, v) = \gamma(u, x)\} \quad (u \in \mathcal{N}(\bar{u})).
\tag{35}
$$

We show first that M has nonempty images. For this purpose, fix an arbitrary $u \in \mathcal{N}(\bar{u})$ and let $\{x_n\} \subseteq \mathcal{O}$ be a sequence with $\gamma(u, x_n) \to \rho(u, v)$. Without loss of generality, $x_n \to \bar{x} \in \bar{D}$. Assume that $\bar{x} \notin \mathcal{O}$. Then, $[P(0, \Sigma^{1/2} v)](\bar{x}) = 0$ because of $x_n \in \mathcal{O}$. At the same time,

$$
\alpha - y_0(\bar{x}) - [P(u, 0)](\bar{x}) > 0
$$

by virtue of (32). It follows that $\gamma(u, x_n) \to \infty$, whence the contradiction $\rho(u, v) = \infty$ with the already mentioned fact that $\mathcal{O} \neq \emptyset$. Therefore, $\bar{x} \in \mathcal{O}$. Now, the continuity of γ on $L^2(D) \times \mathcal{O}$ yields that

$$
\gamma(u, x_n) \to \gamma(u, \bar{x}) = \rho(u, v).
$$

Hence, $\bar{x} \in M(u) \neq \emptyset$ as was to be shown.

Next, we see that $M(\bar{u})$ is closed, hence compact as a subset of $\bar{D}$. Indeed, let $x_n \in M(\bar{u})$ be a sequence with $x_n \to \bar{x}$. Since $x_n \in \mathcal{O}$, the same reasoning as before allows us to verify that $\bar{x} \in \mathcal{O}$. Then, again by continuity of γ,

$$
\rho(\bar{u}, v) = \gamma(\bar{u}, x_n) \to \gamma(\bar{u}, \bar{x}),
$$

providing that $\rho(\bar{u}, v) = \gamma(\bar{u}, \bar{x})$ and, thus, $\bar{x} \in M(\bar{u})$.

Knowing that $M(\bar{u})$ is a nonempty and compact subset of $\mathcal{O}$, there must exist another compact subset $\mathcal{K} \subseteq \mathcal{O}$ such that $M(\bar{u}) \subseteq \operatorname{int} \mathcal{K}$. We claim that

$$
\rho(u, v) = \min_{x \in \mathcal{K}} \gamma(u, x) \quad \forall u \in \mathcal{N}(\bar{u}),
\tag{36}
$$

which would be our desired representation of ρ as a minimum over a compact set. For (36) to hold true it would be sufficient to verify that

$$M(u) \subseteq \mathcal{K} \quad \forall u \in \mathcal{N}(\bar{u}). \tag{37}$$

Indeed, assuming this holds true, we may fix an arbitrary $u \in \mathcal{N}(\bar{u})$ and select some $\hat{x} \in M(u)$ which is possible since this latter set is nonempty, as we have shown before. Then, by (37), $\hat{x} \in \mathcal{K} \subseteq \mathcal{O}$, and so, by definition of $M(u)$,

$$\rho(u, v) \leq \min_{x \in \mathcal{K}} \gamma(u, x) \leq \gamma(u, \hat{x}) = \rho(u, v),$$

where the first inequality relies on (34). This is (36) and it remains to verify (37). If this relation was false, then there would exist sequences $u_n \to \bar{u}$ and $x_n \in M(u_n) \setminus \mathcal{K}$. Without loss of generality, $x_n \to \bar{x} \in \bar{D}$. Then $\bar{x} \notin \operatorname{int} \mathcal{K} \subseteq \mathcal{O}$ and, by definition of the set $\mathcal{K}$, $\bar{x} \notin M(\bar{u})$. Select some $\hat{x} \in M(\bar{u}) \subseteq \mathcal{O}$ (recalling that $M(\bar{u}) \neq \emptyset$). Then,

$$\gamma(\bar{u}, \bar{x}) = \lim_n \gamma(u_n, x_n) = \lim_n \rho(u_n, v) \leq \lim_n \gamma(u_n, \hat{x}) = \gamma(\bar{u}, \hat{x}) = \rho(\bar{u}, v) < \gamma(\bar{u}, \bar{x}),$$

where the strict inequality comes as a consequence of $\bar{x} \in \mathcal{O} \setminus M(\bar{u})$. This contradiction establishes (37).

Now, defining a function $\kappa : L^2(D) \to \mathbb{R}$ by

$$\kappa(u) := \max_{x \in \mathcal{K}} -\gamma(u, x) \quad u \in L^2(D),$$

we derive from (36) that

$$\rho(u, v) = -\kappa(u) \quad \forall u \in \mathcal{N}(\bar{u}) \tag{38}$$

Since κ is the maximum of continuous, affine linear functions $-\gamma(\cdot, x)$ over a compact set, it is convex and finite-valued. Therefore, and taking into account the continuity of γ, the Ioffe–Tikhomirov Theorem [20, Section 4.2, Theorem 3] allows us to represent the convex subdifferential of κ at $\bar{u}$ as

$$\partial \kappa(\bar{u}) = \operatorname{clco} \left\{ \frac{\partial(-\gamma)}{\partial u}(\bar{u}, x) \,\Big|\, x \in \tilde{M}(\bar{u}, v) \right\}, \quad \tilde{M}(\bar{u}, v) := \{x \in \mathcal{K} \mid -\gamma(\bar{u}, x) = \kappa(\bar{u})\}.$$

With $u_x \in L^2(D)$ as introduced in the statement of this lemma, we arrive at

$$\frac{\partial(-\gamma)}{\partial u}(\bar{u}, x) = \frac{1}{[P(0, \Sigma^{1/2}v)](x)} \cdot u_x \quad \forall x \in \mathcal{O}.$$

The continuity of γ along with the compactness of $\mathcal{K}$ imply via (36) that κ is continuous. Hence, being also convex, κ is even locally Lipschitz continuous (cf. [4, Prop. 2.107]) and so, its subdifferential in the sense of convex analysis coincides with its Clarke subdifferential [7, Proposition 2.2.7]. Applying now a well-known calculus rule for the latter [7, Proposition 2.3.1] and taking into account (38), we derive that

$$\partial_u^C \rho(\bar{u}, v) = \partial^C(-\kappa(\bar{u})) = -\partial^C \kappa(\bar{u}) = -\partial \kappa(\bar{u})$$
$$= -\operatorname{clco} \left\{ \frac{1}{[P(0, \Sigma^{1/2}v)](x)} \cdot u_x \,\Big|\, x \in \tilde{M}(\bar{u}, v) \right\}. \tag{39}$$

Since $\rho(\cdot, v)$ is finite-valued on $\mathcal{N}(\bar{u})$ (see (36)), the representation (19) of the radial probability function yields that $e(u, v) = F_\eta(\rho(u, v))$ locally around $\bar{u}$. With the one-dimensional function F_η being continuously differentiable and having derivative χ (the density of the chi distribution with m

degrees of freedom), the chain rule for Clarke's subdifferential [7, Theorem 2.3.9 (ii)] provides the identity

$$\partial_u^C e(\bar{u}, v) = F_\eta'(\rho(\bar{u}, v)) \cdot \partial_u^C \rho(\bar{u}, v) = \chi(\rho(\bar{u}, v)) \cdot \partial_u^C \rho(\bar{u}, v).$$

From here, the assertion of our lemma will follow along with (39), once we can show that $\tilde{M}(\bar{u}, v) = M^*(\bar{u}, v)$, where the latter set has been introduced in the statement of this lemma. Indeed, if $x \in \tilde{M}(\bar{u}, v)$, then $x \in \mathcal{K} \subseteq \mathcal{O}$ and $\gamma(\bar{u}, x) = \rho(\bar{u}, v)$ by (38). Then, $x \in M^*(\bar{u}, v)$ by definition of γ and of the expected state $\bar{y}$. Conversely, let $x \in M^*(\bar{u}, v)$. Then, $x \in \mathcal{O}$ and $\rho(\bar{u}, v) = \gamma(\bar{u}, x)$ again by definition of γ. Hence, $x \in M(\bar{u})$ by (35). Now, (37) yields that $x \in \mathcal{K}$ and, hence, $x \in \tilde{M}(\bar{u}, v)$ by (38). $\qquad\square$

2.4 Subdifferential of the probability function and optimality conditions in the concrete PDE setting

The technical results of the previous section allows us now to formulate a fully explicit (in terms of the problem data) upper estimate for the subdifferential of the probability function in problem (23) and explicit necessary optimality conditions for that same problem. Combining Corollary 2.9 with Lemmas 2.11 and 2.12, we end up with

Theorem 2.13. *Fix some $\bar{u} \in L^2(D)$ with associated expected state $\bar{y}$ satisfying (26). Then, φ is locally Lipschitzian around $\bar{u}$ and, with the notation introduced in Lemma 2.12, the subdifferential of the probability function at $\bar{u}$ can be estimated from above by*

$$\partial^C \varphi(\bar{u}) \subseteq - \int\limits_{\{v \in \mathbb{S}^{m-1} | \rho(\bar{u}, v) < \infty\}} \chi(\rho(\bar{u}, v)) \qquad (40)$$

$$\cdot \operatorname{clco} \left\{ \frac{1}{[P(0, \Sigma^{1/2} v)](x)} \cdot u_x \,\Big|\, x \in M^*(\bar{u}, v) \right\} d\mu_\zeta(v).$$

Here, we use the notation $M^*(u, v)$ introduced in Lemma 2.12. The inclusion holds as an equality provided that $M^*(\bar{u}, v)$ is a singleton for μ_ζ a.e. $v \in \mathbb{S}^{m-1}$ with $\rho(\bar{u}, v) < \infty$ (implying condition (16) thanks to Lemmas 2.11 and 2.12 and, hence, strict differentiability of φ) or that the set (11) is bounded (implying the assumptions of Lemma 2.2).

Likewise, Corollary 2.10 yields the following necessary optimality conditions:

Theorem 2.14. *In the optimization problem (23), let there exist some control $\hat{u} \in L^2(D)$ such that the associated random state $\hat{y}$ solving (23b)–(23c) satisfies the probabilistic constraint (23d) strictly, i.e.,*

$$\varphi(\hat{u}) = \mathbb{P}(\hat{y}(x, \omega) \le \alpha \quad \forall x \in D) > p. \qquad (41)$$

Consider some $u^* \in L^2(D)$ that is feasible in (23). Assume that the associated expected state $\tilde{y} := S(u^*, 0)$ satisfies

$$\tilde{y}(x) < \alpha \quad \forall x \in \bar{D}. \qquad (42)$$

Then, for u^* to be a solution to (9), it is necessary that there exists some $\lambda \ge 0$ with

$$DF(u^*) \in -\lambda \int\limits_{\{v \in \mathbb{S}^{m-1} | \rho(u^*, v) < \infty\}} \chi(\rho(u^*, v)) \qquad (43)$$

$$\cdot \operatorname{clco} \left\{ \frac{1}{[P(0, \Sigma^{1/2} v)](x)} \cdot u_x \,\Big|\, x \in M^*(u^*, v) \right\} d\mu_\zeta(v)$$

and

$$\lambda(\mathbb{P}(y^*(x,\omega) \leq \alpha \quad \forall x \in D) - p) = 0, \tag{44}$$

where $y^*(\cdot, \omega) = S(u^*, \xi(\omega))$ is the random state associated with the control u^* and random event $\omega \in \Omega$. Moreover, this inclusion also becomes sufficient for u^* to be a solution to (9), if the conditions from Theorem 2.13 for (40) to hold as an equality are satisfied.

The following example shall illustrate Theorems 2.1 and 2.13, respectively. The data will be chosen in such as way so that (26) is satisfied and, hence, to apply both results. As a consequence, the probability function will be locally Lipschitzian. However, it will fail to be differentiable despite the fact that all problem data are smooth. This underlines why it is necessary to employ tools from generalized differential calculus in the analysis of the optimization problem (23). Moreover, we are going to calculate explicitly the upper estimate (40) for the subdifferential of the probability function provided in Theorem 2.13.

Example 2.15. *In our problem (23), let $m := 1, D := (0, 2\pi)^2, \alpha := 1, f_0 := 0, \xi \sim \mathcal{N}(0,1)$ and*

$$\phi_1(x_1, x_2) := 2\sin^2 x_1(\sin^2 x_2 - \cos^2 x_2) + 2\sin^2 x_2(\sin^2 x_1 - \cos^2 x_1).$$

Fix a control $\hat{u}(x_1, x_2) := \sin x_1 \sin x_2$. Then, for each $\tau \in \mathbb{R}$ and each $z_1 \in \mathbb{R}$,

$$y_{\tau, z_1}(x_1, x_2) := \tau \sin x_1 \sin x_2 + z_1 \sin^2 x_1 \sin^2 x_2$$

is the solution to the PDE

$$-\Delta y(x) = 2\tau \hat{u}(x) + f(x, z_1) \quad (x \in D), \quad y(x) = 0 \quad (x \in \partial D).$$

For the purpose of illustration, we consider the restriction of the probability function φ to the one-dimensional subspace $\mathbb{R}_+\{\hat{u}\}$, i.e., the probability function $\tilde{\varphi} \colon \mathbb{R} \to \mathbb{R}$ defined by

$$\tilde{\varphi}(\tau) := \varphi(2\tau\hat{u}) = \mathbb{P}(\omega \mid y_{\tau, \xi(\omega)}(x) \leq 1 \quad \forall x \in (0, 2\pi)^2).$$

One easily checks that (26) is satisfied at $2\tau\hat{u}$ for all $\tau \in (-1, 1)$. Accordingly, Corollary 2.9 (via Theorem 2.1) yields that, at all $2\tau\bar{u}$ with $\tau \in (-1, 1)$, φ is locally Lipschitzian (hence $\tilde{\varphi}$ is locally Lipschitzian at all $\tau \in (-1, 1)$) and the upper estimate of Theorem 2.13 holds true. We will derive an explicit formula for $\tilde{\varphi}$. Since $y_{\tau, z_1}(x_1, x_2) = 0$ whenever $\sin x_1 \sin x_2 = 0$, one gets that

$$\{z_1 \mid y_{\tau, z_1}(x) \leq \alpha\} = \left\{ z_1 \;\middle|\; z_1 \leq \frac{1 - \tau \sin x_1 \sin x_2}{\sin^2 x_1 \sin^2 x_2} \quad \forall x \in (0, 2\pi)^2 : \sin x_1 \sin x_2 \neq 0 \right\}.$$

If $0 \leq \tau < 1$, then the fraction on the right-hand side is minimized whenever $\sin x_1 \sin x_2 = 1$ (e.g., by choosing $x := (\pi/2, \pi/2)$). The achieved minimum equals $1 - \tau$. Likewise, if $-1 < \tau \leq 0$ then this fraction is minimized whenever $\sin x_1 \sin x_2 = -1$ (e.g., by choosing $x := (\pi/2, 3\pi/2)$). The achieved minimum equals $1 + \tau$. Accordingly,

$$\tilde{\varphi}(\tau) = \begin{cases} \mathbb{P}(\omega \mid \xi(\omega) \leq 1 + \tau) = \Phi(1 + \tau) & \text{if } -1 < \tau \leq 0 \\ \mathbb{P}(\omega \mid \xi(\omega) \leq 1 - \tau) = \Phi(1 - \tau) & \text{if } 0 \leq \tau < 1 \end{cases}$$

Here, Φ denotes the distribution function of the one-dimensional standard Gaussian distribution $\mathcal{N}(0, 1)$. It is locally Lipschitzian as expected, but it fails to be differentiable at zero because the derivatives at

zero of the two pieces become $F(1) > 0$ and $-F(1) < 0$, respectively, where $F = \Phi'$ refers to the standard Gaussian density.

Next, we are going to calculate explicitly the upper estimate (40) *from Theorem 2.13 at the reference point $\bar{u} = 0$. First, we note that $\Sigma^{1/2} = 1$ due to $\xi \sim \mathcal{N}(0,1)$ and $\mathbb{S}^{m-1} = \mathbb{S}^0 = \{-1, 1\}$. By definition of the operator P and of the state y_{τ, z_1}, we have that (for all $(x_1, x_2) \in (0, 2\pi)^2$):*

$$[P(0,1)](x) = y_{0,1}(x_1, x_2) = \sin^2 x_1 \sin^2 x_2; \quad [P(0,-1)](x) = y_{0,-1}(x_1, x_2) = -\sin^2 x_1 \sin^2 x_2.$$

Since $y_0 = 0$ by our data assumption $f_0 = 0$, we infer from (29) *that*

$$\left\{ r \geq 0 \,\bigg|\, \max_{x \in \bar{D}} \, [P(\bar{u}, r\Sigma^{1/2} \cdot (-1))](x) + y_0(x) \leq \alpha \right\} = \left\{ r \geq 0 \,\bigg|\, \max_{x \in [0,2\pi]^2} \, [P(0, -r)](x) \leq 1 \right\}$$

$$= \left\{ r \geq 0 \,\bigg|\, \max_{x \in [0,2\pi]^2} \, -r \sin^2 x_1 \sin^2 x_2 \leq 1 \right\} = [0, \infty),$$

hence $\rho(\bar{u}, -1) = \infty$ by definition of ρ. Therefore, $v = -1$ does not contribute to the integral in (40). *Likewise,*

$$\left\{ r \geq 0 \,\bigg|\, \max_{x \in \bar{D}} \, [P(\bar{u}, r\Sigma^{1/2} \cdot 1)](x) + y_0(x) \leq \alpha \right\} = \left\{ r \geq 0 \,\bigg|\, \max_{x \in [0,2\pi]^2} \, r \sin^2 x_1 \sin^2 x_2 \leq 1 \right\}$$

$$= [0, 1],$$

hence $\rho(\bar{u}, 1) = 1 < \infty$ and so $v = 1$ does contribute to the integral in (40). *Since the uniform distribution on $\mathbb{S}^0$ can be represented as $\mu_\zeta = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, where δ_t is the Dirac measure at $t \in \mathbb{R}$, the negative value of this integral appearing in* (40) *reduces to*

$$-\frac{1}{2}\chi(\rho(0,1)) \cdot \text{clco} \left\{ \frac{1}{[P(0,1)](x)} \cdot u_x \mid x \in M^*(0,1) \right\}$$

$$= -\frac{1}{2}\chi(1) \cdot \text{clco} \left\{ \frac{1}{\sin^2 x_1 \sin^2 x_2} \cdot u_x \mid x \in M^*(0,1) \right\}.$$

It remains to determine the index set $M^*(0,1)$. First note that $\bar{y} = 0$ holds true for the expected state (associated with control $\bar{u} = 0$ and $z_1 = 0$) by our data assumption $f_0 = 0$. By definition of M^* in Lemma 2.12 we arrive at

$$M^*(0,1) = \{x \in [0, 2\pi]^2 \mid \sin^2 x_1 \sin^2 x_2 = 1\}$$
$$= \{(\pi/2, \pi/2), (3\pi/2, \pi/2), (\pi/2, 3\pi/2), (3\pi/2, 3\pi/2)\}.$$

This yields the final explicit upper estimate for the subdifferential of the probability function

$$\partial^C \varphi(0) \subseteq -\frac{1}{2}\chi(1) \cdot \text{co} \left\{ u_{(\pi/2,\pi/2)}, u_{(3\pi/2,\pi/2)}, u_{(\pi/2,3\pi/2)}, u_{(3\pi/2,3\pi/2)} \right\},$$

where the $u_x \in L^2((0, 2\pi)^2)$ have been introduced in the statement of Lemma 2.12 and where the closure operation in the convex hull can be omitted due to the set being finite.

We want to illustrate Theorem 2.14 in an example.

Example 2.16. *In our problem* (23)*, let (with Φ being the distribution function of $\mathcal{N}(0,1)$)*

$$D := (0,1)^2, m := 1, \alpha := 1/16, f_0 := 0, \xi \sim \mathcal{N}(0,1),$$
$$\phi_1(x_1, x_2) := 2(x_1(1-x_1) + x_2(1-x_2)), p := \Phi(1) \approx 0.841.$$

Fix a control $\tilde{u} \in L^2(D)$ defined by the relation

$$(\tilde{u}, h)_{L^2(D)} = [P(h, 0)](1/2, 1/2) = [A^{-1}h](1/2, 1/2) \quad \forall h \in L^2(D). \tag{45}$$

We define the objective in (23a) *by $F(u) := \|u - \tilde{u}\|^2_{L^2(D)}$ and claim that $u^* := 0$ is a solution of the optimization problem. Evidently, for each $z_1 \in \mathbb{R}$, the function $y(x_1, x_2) := z_1 x_1(1 - x_1)x_2(1 - x_2)$ is a solution to the PDE*

$$-\Delta y(x) = z_1 \phi_1(x) \quad (x \in D), \quad y(x) = 0 \quad (x \in \partial D), \tag{46}$$

which is (8) *with $u = 0$. It follows from* (5) *that*

$$[S(0, z_1)](x) = [P(0, z_1)](x) = [P(z_1\phi_1, 0)](x) = z_1 x_1(1 - x_1)x_2(1 - x_2) \quad \forall z_1 \in \mathbb{R}. \tag{47}$$

In particular, $S[(u^*, 0)](x) = 0$ for all $x \in \bar{D}$ so that (42) *is satisfied. Moreover,*

$$\varphi(u^*) = \varphi(0) = \mathbb{P}(\omega \mid \xi(\omega)x_1(1 - x_1)x_2(1 - x_2) \leq 1/16 \; \forall x \in (0,1)^2)$$
$$= \mathbb{P}(\omega \mid \xi(\omega) \in (-\infty, 1]) = \Phi(1) = p.$$

This means that $u^* = 0$ is a feasible control in problem (23) *at which the probabilistic constraint is active. Next, we calculate the Clarke subdifferential $\partial^C \varphi(u^*)$. Recalling that, due to the assumptions on our data, $\Sigma = 1$ and $\mathbb{S}^{m-1} = \mathbb{S}^0 = \{-1, 1\}$, we identify the two radii $\rho(0, 1), \rho(0, -1)$ via* (29) *from the relations*

$$[0, \rho(0, 1)] = \left\{ r \geq 0 \,\middle|\, \max_{x \in [0,1]^2} [P(0, r)](x) \leq 1/16 \right\} = [0, 1],$$

$$[0, \rho(0, -1)] = \left\{ r \geq 0 \,\middle|\, \max_{x \in [0,1]^2} [P(0, -r)](x) \leq 1/16 \right\} = [0, \infty).$$

Accordingly, $\rho(0, 1) = 1$ and $\rho(0, -1) = \infty$. By definition and by virtue of (47),

$$M^*(0, 1) = \{x \in [0, 1]^2 \mid [P(0, 1)](x) > 0, \; [P(0, 1)](x) = 1/16\} = \{(1/2, 1/2)\}.$$

Consequently, $\#M^*(0, 1) = 1$ and, hence, Theorem 2.13 *yields that*

$$\partial^C \varphi(u^*) = -\int_{v=1} \frac{\chi(\rho(0, 1))}{[P(0, 1)](1/2, 1/2)} \cdot u_{(1/2,1/2)} d\mu_\zeta(v)$$
$$= \left\{ -\frac{1}{2} \frac{\chi(1)}{[P(0, 1)](1/2, 1/2)} \cdot u_{(1/2,1/2)} \right\} = \{-8\chi(1)\tilde{u}\}, \tag{48}$$

where $\tilde{u}$ has been introduced above. In particular, φ is strictly differentiable with derivative $\nabla \varphi(u^*) = -8\chi(1)\tilde{u}$.

Finally, we want to apply the optimality conditions of Theorem 2.14. *In order to do so, it remains first to check condition* (41). *Thanks to* (47) *we have that $[P(\phi_1, 0)](1/2, 1/2) = 1/16 > 0$. This implies that $\tilde{u} \neq 0$ because otherwise a contradiction with* (45) *would occur. Now, assuming that condition* (41) *is violated, it would follow that $\varphi(u) \leq p$ for all $u \in L^2((0, 1)^2)$. Owing to $\varphi(u^*) = p$ (see above), it would follow that u^* is a (global) maximum of φ. By Clarke's necessary optimality condition [7, Proposition 2.3.2] and by* (48) *we would end up at the contradiction $0 \in \partial^C(\varphi)(u^*) = \{-8\chi(1)\tilde{u}\}$ with $\tilde{u} \neq 0$ (see above). Summarizing, conditions* (41), (42), *and $\#M^*(0, 1) = 1$ (while $\rho(0, -1) = \infty$) are satisfied, which allows us to apply Theorem* 2.14. *In order to prove our claim that $u^* := 0$ is a*

solution of (23), *it remains to verify* (43) *and* (44). *We define a multiplier* $\lambda := \frac{-1}{4\chi(1)} < 0$. *As for the complementarity relation, it is satisfied thanks to the already stated equality* $\varphi(0) = p$. *On the other hand,* (48) *yields that*

$$
\begin{aligned}
\nabla F(0) &= -2\tilde{u} = \lambda 8\chi(1)\tilde{u} = \lambda \partial^C \varphi(u^*) = \lambda \int\limits_{v=1} \frac{\chi(\rho(0,1))}{[P(0,1)](1/2,1/2)} \cdot u_{(1/2,1/2)} d\mu_\zeta(v) \\
&= \lambda \int\limits_{\{v \in \mathbb{S}^0 \mid \rho(u^*,v) < \infty\}} \chi(\rho(u^*,v)) \cdot \mathrm{clco} \left\{ \frac{1}{[P(0,\Sigma^{1/2}v)](x)} \cdot u_x \,\Big|\, x \in M^*(u^*,v) \right\} d\mu_\zeta(v)
\end{aligned}
$$

where we recall that $\rho(0,-1) = \infty$ *and* $M^*(0,1) = \{(1/2,1/2)\}$. *Now,* (43) *as a sufficient condition proves that* $u^* = 0$ *is a solution of the formulated optimization problem.*

3 Almost sure and robust constraints

In Section 2, we considered optimization problems with chance constraints of the form $\mathbb{P}(g(u,\xi(\omega)) \leq 0) \geq p$. In this section, we will consider two closely related models:

$$
\begin{aligned}
g(u,\xi(\omega)) \leq 0 \quad &\mathbb{P}\text{-almost surely} \qquad &\text{(almost sure)}, \\
g(u,z) \leq 0 \quad &\forall z \in \mathsf{supp}\,\xi \qquad &\text{(robust)}.
\end{aligned}
$$

Optimality conditions for PDE-constrained optimization problems with almost sure state constraints were recently been derived in [13, 14] and robust formulations have also been seen in the literature; see, e.g., [22]. We will show optimality conditions for the almost sure and robust cases (Section 3.1 and Section 3.2, respectively) using our model PDE. First, we will discuss some nuances in the various models.

The probabilistic model from Section 2 is equivalent to the the almost sure model if $p = 1$. However, as mentioned in Remark 2.6, the optimality conditions obtained there exclude the case $p = 1$. One might hope that optimality conditions for the probability level $p = 1$ could be obtained in the limit $p \to 1$, but this is not the case, as demonstrated in the following example.

Example 3.1. *Let* ξ *be a one-dimensional random variable with density* $f(x) := 2(1-x)\chi_{[0,1]}$ *and consider the problem*

$$
\min_{u \in \mathbb{R}} \{F(u) := u\} \quad \text{s.t.} \quad \varphi(u) := \mathbb{P}(g(u,\xi(\omega)) \leq 0) \geq p, \tag{49}
$$

where $g(u,z) := z - u$ *for* $z, u \in \mathbb{R}$. *Then, the probability function is given by*

$$
\varphi(u) = \begin{cases} 0 & u < 0, \\ 2u - u^2 & u \in [0,1], \\ 1 & u > 1. \end{cases}
$$

At $u > 0$, φ *is continuously differentiable. For* $p > 0$, *the constraint set* $\varphi(u) \geq p$ *is identified as the interval* $[u_p^*, \infty)$ *with* $u_p^* := 1 - \sqrt{1-p} \in (0,1]$ *necessarily being the solution of* (49). *Note that* $\varphi(u_p^*) = p$, *so the inequality constraint is binding at* u_p^*. *If* $p = 1$, *then* $\varphi'(u_p^*) = \varphi'(1) = 0$. *Hence, the basic constraint qualification is violated at the solution* $u_1^* = 1$ *of problem* (49) *and no optimality condition can be established. If* $p < 1$, *then*

$$
\varphi'(u_p^*) = 2 - 2(1 - \sqrt{1-p}) > 0
$$

and the basic constraint qualification for necessary optimality conditions is satisfied. It follows that the Lagrange multiplier of the KKT conditions is uniquely determined by the relation

$$1 = F'(u_p^*) = \lambda_p \varphi'(u_p^*)$$

yielding $\lambda_p = 1/(2\sqrt{1-p})$. Now, for $p \to 1$, we get $F'(u_p^*) \to 1$ on the one hand, but $\varphi'(u_p^*) \to 0$ and $\lambda_p \to \infty$ on the other hand, so that the limit of KKT conditions does not provide any information.

The negative observations made in the previous example result from the fact that setting $p = 1$ in the probabilistic version provides a degenerate problem description. In contrast, the almost sure and robust perspective (which are equivalent in this example) both work well:

Example 3.2. *We reconsider problem* (49) *but in its equivalent robust description*

$$\min_{u \in \mathbb{R}} \{F(u) := u\} \quad \text{s.t.} \quad \tilde{\varphi}(u) := \sup_{z \in [0,1]} g(u, z) \le 0 \tag{50}$$

(note that $[0,1]$ is the support of ξ in Example 3.1). The fact that this problem is equivalent to (49) will follow from Lemma 3.4 below, because g is continuous in z, and hence the robust description is equivalent to the almost-sure description, which in turn is always equivalent with the probabilistic one. Clearly, the sup-function is given by

$$\tilde{\varphi}(u) = \sup_{z \in [0,1]} z - u = 1 - u.$$

Then, $\tilde{\varphi}(2) < 0$, so that $\hat{u} = 2$ is a Slater point of the problem and equivalent optimality conditions can be formulated. The feasible set equals the interval $[1, \infty)$ and at the solution $u^* = 1$ of the problem, one has the KKT relation

$$1 = F'(1) = (-1)(\tilde{\varphi})'(1) = (-1)(-1).$$

The following example shows that generally, an almost sure model is not equivalent to a robust one.

Example 3.3. *Consider the one-dimensional control problem*

$$\min_{u \in L^2(0,1)} \|u\|_{L^2(0,1)}^2 \quad \text{s.t.} \quad y(x, \omega) \le 1 \quad \forall x \in [0, 1] \quad \mathbb{P}\text{-a.s.},$$

where $y(x, \omega)$ is the solution to

$$-\Delta y(x, \omega) = u(x) + 2f(\xi(\omega)) \quad x \in (0, 1) \; \mathbb{P}\text{-a.s.}; \quad y(x, \omega) = 0 \quad x \in \{0, 1\} \; \mathbb{P}\text{-a.s.},$$

ξ is a one-dimensional random variable uniformly distributed on $[0, 1]$, and

$$f(z) := \begin{cases} 4 & z \ne 0.5, \\ 5 & z = 0.5. \end{cases}$$

Then, $u := 0$ is a feasible control of the problem. Indeed, the random states associated with this control are given by

$$y(x, \omega) = f(\xi(\omega))x(1 - x) \quad \forall x \in [0, 1], \forall \omega \in \Omega.$$

It follows that

$$y(x, \omega) \le \frac{1}{4} f(\xi(\omega)) \le 1 \quad \forall \omega : \xi(\omega) \ne 0.5.$$

Since $\xi(\omega) \neq 0.5$ $\mathbb{P}$-almost surely, feasibility of $u = 0$ follows. In particular, this control is optimal in the given problem. Now, adopting the robust perspective, we observe that $[0, 1]$ is the support of the random variable and formulate the problem

$$\min \|u\|_{L^2(0,1)}^2 \quad \text{s.t.} \quad y(x, z) \leq 1 \quad \forall x \in [0, 1] \quad \forall z \in [0, 1],$$

where $y(x, z)$ is the solution to

$$-\Delta y(x, z) = u(x) + 2f(z) \quad x \in (0, 1), z \in [0, 1]; \quad y(x, z) = 0 \quad x \in \{0, 1\}, z \in [0, 1].$$

Then, the parameterized state $y(x, z) = f(z)x(1-x)$ solving the PDE under the control $u = 0$ fails to be feasible because of $y(0.5, 0.5) = 5/4 > 1$. Consequently, $u = 0$ is not feasible, so cannot be optimal for the robust problem.

The reason for the difference between the robust and almost sure perspectives in the example above is that the state fails to be lower semicontinuous with respect to the parameter z, which is generated by the same deficiency of the function f. We have the following relation between the two models.

Lemma 3.4. *Let ξ be an m-dimensional random vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $\Xi \subseteq \mathbb{R}^m$ the support of its law $\mathbb{P} \circ \xi^{-1}$. Furthermore, let $h \colon \mathbb{R}^m \to \mathbb{R}$ be a lower semicontinuous function. Then,*

$$h(\xi(\omega)) \leq 0 \quad \mathbb{P}\text{-a.s.} \quad \Longleftrightarrow \quad h(z) \leq 0 \quad \forall z \in \Xi.$$

Proof. ($\Longrightarrow$): Assume there exists $z \in \Xi$ with $h(z) > 0$. Since h is lower semicontinuous, there exists some $r > 0$ such that $h(z') > 0$ for all z' in the ball $\mathbb{B}(z, r) \in \mathbb{R}^m$ of radius r centered around the point z. From here and from $z \in \Xi$, we get the contradiction

$$\mathbb{P}\{\omega \mid h(\xi(\omega)) > 0\} \geq \mathbb{P}\{\omega \mid \xi(\omega) \in \mathbb{B}(z, r)\} > 0.$$

($\Longleftarrow$): The reverse direction follows readily from

$$1 = \mathbb{P}\{\omega \mid \xi(\omega) \in \Xi\} \leq \mathbb{P}\{\omega \mid h(\xi(\omega)) \leq 0\}.$$

$\square$

3.1 Almost sure state constraints

In this section, we will derive optimality conditions for a model problem with almost sure state constraints. As mentioned in the introduction, the functional-analytic techniques applied here largely come from traditional approaches developed by Rockafellar and Wets for two-stage problems in stochastic programming [33, 34, 35, 36]. Much of the analysis there hinged on an assumption of *relatively complete recourse*, a requirement that every control u must result in a feasible state y. This assumption is too strong for our model problem (51) below, since this would make the state constraint redundant for any choice of u. In [33] and in [14], this assumption was lifted, resulting in singular Lagrange multipliers. Unlike in the previous section, there is little hope for a more explicit representation and these elements do not provide a structure that is amenable to numerical approximation. In practice, a regularization as in [14] can be used so that the singular terms only appear in the limit. Here, however, we study the unregularized problem, which also shares structural similarities to that of its robust counterpart, as we will see in Section 3.2. As a main result, we will see that we can more efficiently arrive at

the conditions provided by [13, 14] by first applying standard optimality theory and then refining the Lagrange multipliers using a Yosida–Hewitt-type decomposition.

Consider the problem

$$\min_{u \in L^2(D)} F(u) \tag{51a}$$

$$\text{s.t.} \quad -\Delta y(x,\omega) = u(x) + \tilde{f}(x,\omega), \quad x \in D \quad \mathbb{P}\text{-a.s.}, \tag{51b}$$

$$y(x,\omega) = 0, \qquad\qquad x \in \partial D \quad \mathbb{P}\text{-a.s.}, \tag{51c}$$

$$y(x,\omega) \leq \alpha \qquad\qquad x \in D \quad \mathbb{P}\text{-a.s.} \tag{51d}$$

We have replaced the structured source term f with a random field $\tilde{f}$; while more general differential operators would be possible using the results from [13, 14], we choose the simple model here for the sake of comparison with the model introduced in Section 2.2. Now, we will assume the following.

Assumption 3.5. *The open and bounded set $D \subseteq \mathbb{R}^d$ ($d = 1, 2, 3$) is of class S. Additionally, $u \in L^2(D)$ and $\tilde{f} \in L_\mathbb{P}^\infty(\Omega, L^2(D))$.*

We recall the continuous linear operator $A^{-1} \colon L^2(D) \to \mathcal{C}(\bar{D})$ defined after Lemma 1.2.

Lemma 3.6. *Suppose Assumption 3.5 holds. Then for any $u, \tilde{f}(\cdot,\omega) \in L^2(D)$, and $\mathbb{P}$-almost every $\omega \in \Omega$, there a unique solution $y(\cdot,\omega) \in H_0^1(D) \cap \mathcal{C}(\bar{D})$ to (51b)–(51c). Moreover, $y \in L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D}))$ and there exists a constant $C > 0$ such that*

$$\|y\|_{L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D}))} \leq C(\|u\|_{L^2(D)} + \|\tilde{f}\|_{L_\mathbb{P}^\infty(\Omega, L^2(D))}). \tag{52}$$

Proof. Following the proof of Lemma 1.2, we obtain the existence of a unique solution $y(\cdot,\omega) \in H_0^1(D) \cap \mathcal{C}(\bar{D})$ and a $C > 0$ such that

$$\|y(\cdot,\omega)\|_{\mathcal{C}(\bar{D})} \leq C\|u + \tilde{f}(\cdot,\omega)\|_{L^2(D)} \leq C(\|u\|_{L^2(D)} + \|\tilde{f}\|_{L_\mathbb{P}^\infty(\Omega, L^2(D))}),$$

which shows (52) due to our assumption on $\tilde{f}$. Moreover, $\omega \mapsto y(\cdot,\omega)$ is measurable since $\tilde{f}$ is measurable and the operator $A^{-1} \colon L^2(D) \to \mathcal{C}(\bar{D})$ is continuous. $\qquad\square$

Analogously to (5), we will define a control-to-state operator $\mathcal{S} \colon L^2(D) \to L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D}))$ by

$$\mathcal{S}(u) = \mathcal{A}^{-1}(\mathcal{B}u + \tilde{f}).$$

Here, $\mathcal{A}^{-1} \colon L_\mathbb{P}^\infty(\Omega, L^2(D)) \to L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D}))$ is defined by $[\mathcal{A}^{-1}y](\omega) = A^{-1}y(\cdot,\omega)$ a.s. The operator $\mathcal{B}$ is the continuous injection that maps each element of $L^2(D)$ to itself in $L_\mathbb{P}^\infty(\Omega, L^2(D))$. For the state constraint, let ψ denote the constant counterpart to α in $\mathcal{C}(\bar{D})$ and define

$$\mathcal{G} \colon L^2(D) \to L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D})), u \mapsto \mathcal{S}(u) - \psi.$$

Additionally, we define the (convex and closed) cone $\mathcal{K} := \{y \in L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D})) : y(x,\omega) \leq 0$ a.e. in D $\mathbb{P}$-a.s.$\}$ and its polar cone $\mathcal{K}^- := \{y^* \in (L_\mathbb{P}^\infty(\Omega, \mathcal{C}(\bar{D})))^* : \langle y^*, y \rangle \leq 0 \quad \forall y \in \mathcal{K}\}$. It is straightforward to verify the $(-\mathcal{K})$-convexity of $\mathcal{G}$, i.e.,

$$\lambda\mathcal{G}(u) + (1-\lambda)\mathcal{G}(v) - \mathcal{G}(\lambda u + (1-\lambda)v) \in -\mathcal{K} \quad \forall u, v \in L^2(D), \forall \lambda \in [0,1],$$

using the affine linearity of $\mathcal{G}$. First, we present optimality conditions obtained using standard convex analysis.

Lemma 3.7. *Suppose Assumption 3.5 holds and F is convex and continuously Fréchet differentiable on $L^2(D)$. Assume the constraint qualification $0 \in \operatorname{int}\{\mathcal{G}(L^2(D)) - \mathcal{K}\}$ holds. Then, for u^* to be a solution to (51), it is necessary and sufficient that there exists some multiplier $\lambda^* \in (L^\infty_\mathbb{P}(\Omega, \mathcal{C}(\bar{D})))^*$ with*

$$\langle DF(u^*), h \rangle + \langle \lambda^*, \mathcal{A}^{-1}\mathcal{B}h \rangle = 0 \quad \forall h \in L^2(D), \tag{53a}$$

$$\mathcal{G}(u^*) \in \mathcal{K}, \quad \lambda^* \in \mathcal{K}^-, \quad \langle \lambda^*, \mathcal{G}(u^*) \rangle = 0. \tag{53b}$$

Proof. Let $L(u, \lambda) := F(u) + \langle \lambda, \mathcal{G}(u) \rangle$. Since F is convex and $\mathcal{G}$ is $(-\mathcal{K})$-convex, the problem (51) is convex. The constraint qualification implies that the set of Lagrange multipliers is nonempty (see [4, Theorem 3.6]). In particular, we have that there exists $\lambda^* \in (L^\infty_\mathbb{P}(\Omega, \mathcal{C}(\bar{D})))^*$ such that

$$0 \in \partial_u L(u^*, \lambda^*) \quad \text{and} \quad \lambda^* \in N_\mathcal{K}(\mathcal{G}(u^*)), \tag{54}$$

where ∂_u denotes the partial subdifferential with respect to u in the sense of convex analysis. The operator $\mathcal{G} \colon L^2(D) \to L^\infty_\mathbb{P}(\Omega, \mathcal{C}(\bar{D}))$ is affine linear and is bounded thanks to (52). Therefore, it is Fréchet differentiable with $D\mathcal{G}(u)[h] = \mathcal{A}^{-1}\mathcal{B}h$ for all $h \in L^2(D)$. This means the first condition in (54) simplifies to (53a). The conditions (53b) follow from the fact that $\mathcal{K}$ is a convex cone; see [4, p. 150]. Conditions (53) are necessary and sufficient by [4, Proposition 3.3]. □

Remark 3.8. *The constraint qualification $0 \in \operatorname{int}\{\mathcal{G}(L^2(D)) - \mathcal{K}\}$ means that for every w in a neighborhood of zero in $L^\infty_\mathbb{P}(\Omega, \mathcal{C}(\bar{D}))$ there exists a $u \in L^2(D)$ satisfying $\mathcal{G}(u) + w \in \mathcal{K}$. In particular, there exists a u such that $\mathcal{G}(u)(x, \omega) < 0$ a.e. in D $\mathbb{P}$-a.s. If $y = \mathcal{S}(u)$, we obtain the existence of a state that is strictly feasible, i.e., $y(x, \omega) < \alpha$ a.e. in D $\mathbb{P}$-a.s. Consider now the value function $v(w) := \inf\{F(u) \mid u \in L^2(D), \mathcal{G}(u) + w \in \mathcal{K}\}$. Then $0 \in \operatorname{int}\{\mathcal{G}(L^2(D)) - \mathcal{K}\}$ is equivalent to continuity of v at $0 \in L^\infty_\mathbb{P}(\Omega, \mathcal{C}(\bar{D}))$, i.e., $0 \in \operatorname{int} \operatorname{dom} v$. This provides the connection to [13, Assumption 3.5 (ii)]. Note that [13, Assumption 3.5 (i)] ensures that a minimizer exists, something we do not require here.*

It is possible to refine these optimality conditions by taking advantage of the following Yosida–Hewitt-type decomposition result from [19, Appendix 1, Theorem 3] and [25]. Let X be a separable Banach space. A continuous linear functional $v \in (L^\infty_\mathbb{P}(\Omega, X))^*$ of the form

$$\langle v, x \rangle = \int_\Omega \langle x^*(\omega), x(\omega) \rangle \, \mathrm{d}\mathbb{P}(\omega) \tag{55}$$

for some $x^* \in L^1_\mathbb{P}(\Omega, X^*)$ is called *absolutely continuous*. We call $v^\circ \in (L^\infty_\mathbb{P}(\Omega, X))^*$ *singular (relative to $\mathbb{P}$)* if there exists a sequence $\{F_n\} \subset \mathcal{F}$ with $F_{n+1} \subset F_n$ for all n and $\mathbb{P}(F_n) \to 0$ as $n \to \infty$ such that v° is concentrated on $\{F_n\}$, i.e., $\langle v^\circ, x \rangle = 0$ for all $x \in L^\infty_\mathbb{P}(\Omega, X)$, that vanish on some F_n. The set of all absolutely continuous functionals form a closed subspace of $(L^\infty_\mathbb{P}(\Omega, X))^*$ that is isometric to $L^1_\mathbb{P}(\Omega, X^*)$, meaning that for any element $x \in L^\infty_\mathbb{P}(\Omega, X)$, we have

$$\langle r^*, x \rangle = \langle r^\circ, x \rangle + \langle r^a, x \rangle = \langle r^\circ, x \rangle + \mathbb{E}[\langle r(\cdot), x(\cdot) \rangle],$$

where r is the element in $L^1_\mathbb{P}(\Omega, X^*)$ corresponding to r^a.

Theorem 3.9 (Ioffe and Levin)**.** *Each functional $v^* \in (L^\infty_\mathbb{P}(\Omega, X))^*$ has a unique decomposition*

$$v^* = v + v^\circ,$$

where v is absolutely continuous, v° is singular relative to $\mathbb{P}$, and

$$\|v^*\|_{(L^\infty_\mathbb{P}(\Omega, X))^*} = \|v\|_{(L^\infty_\mathbb{P}(\Omega, X))^*} + \|v^\circ\|_{(L^\infty_\mathbb{P}(\Omega, X))^*}.$$

We now have the following result.

Theorem 3.10. *With the assumptions from Lemma 3.7, for u^* to be a solution to (51), it is necessary and sufficient that there exist multipliers $\lambda \in L^1_{\mathbb{P}}(\Omega, \mathcal{C}(\bar{D})^*)$ and $\lambda^\circ \in (L^\infty_{\mathbb{P}}(\Omega, \mathcal{C}(\bar{D})))^*$ with*

$$DF(u^*) + \mathbb{E}[A^{-*}\lambda] + \mathcal{B}^*\mathcal{A}^{-*}\lambda^\circ = 0, \tag{56a}$$

$$Ay^*(\cdot, \omega) - u^* - \tilde{f}(\cdot, \omega) = 0, \tag{56b}$$

$$y^*(\cdot, \omega) \le \alpha \text{ a.e. in } D, \quad \lambda(\cdot, \omega) \ge 0 \text{ a.e. in } D, \quad \langle \lambda(\cdot, \omega), y^*(\cdot, \omega) - \psi \rangle = 0, \tag{56c}$$

$$\lambda^\circ \in \mathcal{K}^-, \quad \langle \lambda^\circ, y^* - \psi \rangle = 0, \tag{56d}$$

with pointwise conditions (in ω) holding $\mathbb{P}$-a.s.

Proof. Fix $h \in L^2(D)$ and let the unique decomposition of λ^* be given by $\lambda^* = \lambda^a + \lambda^\circ$. Note first that $\langle \lambda^*, \mathcal{A}^{-1}\mathcal{B}h \rangle = \langle \lambda^\circ, \mathcal{A}^{-1}\mathcal{B}h \rangle + \langle \lambda^a, \mathcal{A}^{-1}\mathcal{B}h \rangle$. Letting $\lambda \in L^1_{\mathbb{P}}(\Omega, \mathcal{C}(\bar{D})^*)$ be the integrable element identified with λ^a, and recalling that $\mathcal{B}$ is the identity in $L^2(D)$ for every ω, we have

$$\langle \lambda^*, \mathcal{A}^{-1}\mathcal{B}h \rangle = \mathbb{E}[\langle \lambda, A^{-1}h \rangle] + \langle \lambda^\circ, \mathcal{A}^{-1}\mathcal{B}h \rangle = \mathbb{E}[\langle A^{-*}\lambda, h \rangle] + \langle \mathcal{B}^*\mathcal{A}^{-*}\lambda^\circ, h \rangle.$$

Therefore, (53a) is equivalent to (56a). Condition (56b) is equivalent to setting $y^* = \mathcal{S}(u^*)$ and the first inequality in (56c) is evidently equivalent to $\mathcal{G}(u^*) \in \mathcal{K}$.

We now observe the decomposition of the condition $\lambda^* \in \mathcal{K}^-$ and claim that since $\langle \lambda^*, y \rangle = \langle \lambda^\circ + \lambda^a, y \rangle \le 0$ for all $y \in \mathcal{K}$, we must have both

$$\langle \lambda^\circ, y \rangle \le 0 \quad \text{and} \quad \langle \lambda^a, y \rangle = \mathbb{E}[\langle \lambda, y \rangle] \le 0 \quad \forall y \in \mathcal{K}. \tag{57}$$

To show this, we take the sequence of sets $\{F_n\}$ on which λ° is concentrated and construct the functions

$$y_n(\cdot, \omega) := \begin{cases} y(\cdot, \omega), & \omega \in F_n, \\ 0, & \omega \in \Omega \backslash F_n \end{cases} \quad (y \in \mathcal{K}). \tag{58}$$

Since $y \le 0$ a.e. $\mathbb{P}$-a.s., y_n also has this property, so it belongs to $\mathcal{K}$, too. We have $\langle \lambda^\circ, y_n - y \rangle = 0$ since $y_n = y$ in F_n for every n, and moreover (since the following integrand is bounded and $\mathbb{P}(F_n) \to 0$),

$$\langle \lambda^a, y_n \rangle = \int_{F_n} \langle \lambda(\cdot, \omega), y(\cdot, \omega) \rangle \, \mathrm{d}\mathbb{P}(\omega) \to 0 \quad \text{as} \quad n \to \infty.$$

We conclude that for any $y \in \mathcal{K}$, we have

$$\langle \lambda^\circ, y \rangle = \lim_{n \to \infty} \langle \lambda^\circ + \lambda^a, y_n \rangle \le 0.$$

The inequality $\mathbb{E}[\langle \lambda, y \rangle] \le 0$ can evidently be shown by reversing the roles of 0 and y in (58). We have shown that (57) is true.

To conclude, the first inequality is equivalent to $\lambda^\circ \in \mathcal{K}^-$; the inequality $\mathbb{E}[\langle \lambda, y \rangle] \le 0$ for any $y \in \mathcal{K}$ implies $\lambda(\cdot, \omega) \ge 0$ a.e. in D $\mathbb{P}$-a.s. Due to (57), the condition

$$\langle \lambda^*, \mathcal{G}(u^*) \rangle = \langle \lambda^\circ + \lambda^a, y^* - \psi \rangle = 0$$

implies that both

$$\langle \lambda^\circ, y^* - \psi \rangle = 0 \quad \text{and} \quad \langle \lambda^a, y^* - \psi \rangle = \mathbb{E}[\langle \lambda, y^* - \psi \rangle] = 0 \tag{59}$$

must be satisfied. Finally, since $\langle \lambda(\cdot, \omega), y^*(\cdot, \omega) - \psi \rangle \le 0$ $\mathbb{P}$-a.s., the second condition in (59) can only be satisfied if $\langle \lambda(\cdot, \omega), y^*(\cdot, \omega) - \psi \rangle = 0$ $\mathbb{P}$-a.s. $\quad\square$

Remark 3.11. *The arguments used in the proof of Theorem 3.10 greatly simplify the derivation of optimality conditions when compared to [13, 33]. The crucial observation used here is that one can start from the standard result (Lemma 3.7) and directly use the structure of the decomposition to derive the more explicit conditions. Additionally, unlike in [13, 14], we did not need to require the reflexivity of X in the state space $L_{\mathbb{P}}^{\infty}(\Omega, X)$. This suggests that it may be generally possible to relax the assumption of reflexivity in other applications in PDE-constrained optimization under uncertainty.*

The equations (56) are not yet in the form frequently used in PDE-constrained optimization. We can introduce the adjoint variable in (56a) by defining $p(\cdot, \omega) := A^{-*}\lambda(\cdot, \omega)$ and $p^{\circ} := \mathcal{A}^{-*}\lambda^{\circ}$, from which we obtain the adjoint equations

$$\mathcal{A}^*p^{\circ} = \lambda^{\circ} \quad \text{in } (L_{\mathbb{P}}^{\infty}(\Omega, \mathcal{C}(\bar{D})))^* \quad \text{and} \quad A^*p = \lambda \quad \text{in } L_{\mathbb{P}}^1(\Omega, \mathcal{C}(\bar{D})^*). \tag{60}$$

Therefore, (53a) can be expressed in the equivalent form

$$DF(u^*) + \mathbb{E}[p] + \mathcal{B}^*p^{\circ} = 0 \quad \text{in } (L^2(D))^* \tag{61}$$

where p° and p satisfy the respective conditions in (60). This allows us to recover conditions analogous to [13, Lemma 3.9]:

Proposition 3.12. *Under the same assumptions as Lemma 3.7, u^* is a solution to (51) if and only if there exist multipliers $\lambda \in L_{\mathbb{P}}^1(\Omega, \mathcal{C}(\bar{D})^*)$ and $\lambda^{\circ} \in (L_{\mathbb{P}}^{\infty}(\Omega, \mathcal{C}(\bar{D})))^*$ as well as adjoint variables $p \in L_{\mathbb{P}}^1(\Omega, L^2(D)^*)$ and $p^{\circ} \in (L_{\mathbb{P}}^{\infty}(\Omega, L^2(D)))^*$ such that*

$$DF(u^*) + \mathbb{E}[p] + \mathcal{B}^*p^{\circ} = 0, \tag{62a}$$

$$\mathcal{A}^*p^{\circ} - \lambda^{\circ} = 0, \tag{62b}$$

$$A^*p(\cdot, \omega) - \lambda(\cdot, \omega) = 0, \tag{62c}$$

$$Ay^*(\cdot, \omega) - u^* - \tilde{f}(\cdot, \omega) = 0, \tag{62d}$$

$$\lambda(\cdot, \omega) \geq 0 \text{ a.e. in } D, \quad \langle \lambda(\cdot, \omega), y^*(\cdot, \omega) - \psi \rangle = 0, \tag{62e}$$

$$y^*(\cdot, \omega) \leq \alpha \text{ a.e. in } D, \quad \lambda^{\circ} \in \mathcal{K}^-, \quad \langle \lambda^{\circ}, y^* - \psi \rangle = 0, \tag{62f}$$

with pointwise conditions (in ω) holding $\mathbb{P}$-a.s.

As mentioned at the beginning of this section, the appearance of singular terms in the system (62) is inconvenient for numerical approximations. These are typically handled by regularization, even in the deterministic setting (since the corresponding multipliers there are generally only measures). Should additional structure (see Lemma 3.4) be available so that the almost sure and robust formulations are equivalent, the KKT conditions derived for the latter case may be helpful in some settings for the design of numerical methods. The KKT conditions for the robust problem will be the subject of the following section.

3.2 Robust state constraints

In this section, we analyze the following robust counterpart of problem (51):

$$\min_{u \in L^2(D)} F(u) \tag{63a}$$

$$\text{s.t.} \quad -\Delta \hat{y}(x, z) = u(x) + f(x, z), \quad (x, z) \in D \times \Xi, \tag{63b}$$

$$\hat{y}(x, z) = 0, \quad (x, z) \in \partial D \times \Xi, \tag{63c}$$

$$\hat{y}(x, z) \leq \alpha \quad (x, z) \in D \times \Xi. \tag{63d}$$

We start with a robust version of Lemma 3.6.

Lemma 3.13. *Suppose Assumption 1.1 holds and suppose $\Xi \subset \mathbb{R}^m$ is a compact set. Then for any $u \in L^2(D)$, there exists a unique solution $\hat{y} \in \mathcal{C}(\bar{D} \times \Xi)$ of (63b)–(63c). Moreover, there exists a constant $C > 0$ such that*

$$\|\hat{y}\|_{\mathcal{C}(\bar{D}\times\Xi)} \leq C(\max_{z\in\Xi}\|z\|_2 + \|u + f_0\|_{L^2(D)}). \tag{64}$$

Proof. From Lemma 1.2, we have

$$\|\hat{y}\|_{\mathcal{C}(\Xi,\mathcal{C}(\bar{D}))} = \sup_{z\in\Xi}\|\hat{y}(\cdot,z)\|_{\mathcal{C}(\bar{D})} \leq C(\sup_{z\in\Xi}\|z\|_2 + \|u + f_0\|_{L^2(D)}) < \infty$$

and the continuity of the mapping $z \mapsto \hat{y}(\cdot, z)$, meaning $\hat{y} \in \mathcal{C}(\Xi, \mathcal{C}(\bar{D}))$. After identifying $\mathcal{C}(\Xi, \mathcal{C}(\bar{D}))$ with $\mathcal{C}(\bar{D} \times \Xi)$, which is justified since Ξ and $\bar{D}$ are compact sets (cf. [37, p. 50]), we have shown (64). □

Lemma 3.13 justifies the definition of a new control-to-state operator $\hat{S}\colon L^2(D) \to \mathcal{C}(\bar{D} \times \Xi)$ by

$$\hat{S}(u) = \hat{\mathcal{A}}^{-1}(\hat{\mathcal{B}}u + f),$$

where $f \in \mathcal{C}(\Xi, L^2(D))$, $\hat{\mathcal{B}}$ is the continuous injection that maps each element of $L^2(D)$ to itself in $\mathcal{C}(\Xi, L^2(D))$, and $\hat{\mathcal{A}}^{-1}\colon \mathcal{C}(\Xi, L^2(D)) \to \mathcal{C}(\bar{D} \times \Xi)$ is defined by $(\hat{\mathcal{A}}^{-1}y)(\cdot, z) = A^{-1}y(\cdot, z)$. Analogously to the previous section, we introduce $\hat{\mathcal{G}}\colon L^2(D) \to \mathcal{C}(\bar{D} \times \Xi)$ by $[\hat{\mathcal{G}}(u)](x, z) = [\hat{S}(u)](x, z) - \alpha$, the (convex and closed) cone $\hat{\mathcal{K}} := \{y \in \mathcal{C}(\bar{D} \times \Xi) : y(x, z) \leq 0 \text{ in } D \times \Xi\}$, and the polar cone $\hat{\mathcal{K}}^- := \{y^* \in \mathcal{C}(\bar{D} \times \Xi)^* : \langle y^*, y\rangle \leq 0 \,\forall y \in \hat{\mathcal{K}}\}$.

Lemma 3.14. *With the same assumptions as in Lemma 3.13, let F be convex and continuously Fréchet differentiable on $L^2(D)$. Assume the constraint qualification $0 \in \text{int}\{\hat{\mathcal{G}}(L^2(D)) - \hat{\mathcal{K}}\}$ holds. Then, for u^* to be a solution to (63), it is necessary and sufficient that there exists some multiplier $\lambda^* \in \mathcal{C}(\bar{D} \times \Xi)^*$ with*

$$\langle DF(u^*), h\rangle + \langle\lambda^*, \hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}h\rangle = 0 \quad \forall h \in L^2(D), \tag{65a}$$

$$\hat{\mathcal{G}}(u^*) \in \hat{\mathcal{K}}, \quad \lambda^* \in \hat{\mathcal{K}}^-, \quad \langle\lambda^*, \hat{\mathcal{G}}(u^*)\rangle = 0. \tag{65b}$$

Proof. The proof follows identical arguments to those used in Lemma 3.7. □

The dual space $\mathcal{C}(\bar{D} \times \Xi)^*$ can be identified with the set $\mathcal{M}(\bar{D} \times \Xi)$ of all regular Borel measures[1] μ supported on $\bar{D} \times \Xi$ by means of

$$\langle r^*, r\rangle = \int_{\bar{D}\times\Xi} r(x, z)\,\mathrm{d}\mu(x, z).$$

This means that there exists a measure μ^* such that

$$\langle\lambda^*, \hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}h\rangle = \int_{\bar{D}\times\Xi}[\hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}h](x, z)\,\mathrm{d}\mu^*(x, z) = \int_{\bar{D}\times\Xi}[A^{-1}h](x)\,\mathrm{d}\mu^*(x, z).$$

---

[1]A regular Borel measure is in this context the same as a Radon measure due to the compactness of $\bar{D} \times \Xi$; see [2, p. 152].

Moreover, $\lambda^* \in \hat{\mathcal{K}}^-$ implies for the associated measure μ^* that $\int_{\bar{D} \times \Xi} r(x,z) \, \mathsf{d}\mu^*(x,z) \leq 0$ for all $r \in \mathcal{C}(\bar{D} \times \Xi)$ such that $r \leq 0$ a.e. in $\bar{D} \times \Xi$. This implies the nonnegativity of μ^*. Using $y^* := \hat{\mathcal{S}}(u^*)$ and the definition of $\hat{\mathcal{G}}$, (65) can be refined to the conditions

$$\langle DF(u^*), h \rangle + \int_{\bar{D} \times \Xi} [A^{-1}h](x) \, \mathsf{d}\mu^*(x,z) = 0 \quad \forall h \in L^2(D), \tag{66a}$$

$$y^*(x,z) \leq \alpha \quad \forall (x,z) \in \bar{D} \times \Xi, \quad \int_{\bar{D} \times \Xi} (y^*(x,z) - \alpha) \, \mathsf{d}\mu^*(x,z) = 0, \tag{66b}$$

$$\mu^* \in \mathcal{M}(\bar{D} \times \Xi) \text{ nonnegative.} \tag{66c}$$

Connection to semi-infinite programming. Up until now in Section 3, we have used a functional-analytic approach to derive optimality conditions. It is worth mentioning, however, that one can use arguments from semi-infinite programming to arrive at the equivalent conditions for the problem

$$\min_{u \in L^2(D)} F(u) \quad \text{s.t.} \quad g(u) := \max_{(x,z) \in D \times \Xi} [\hat{\mathcal{S}}(u)](x,z) - \alpha \leq 0.$$

We define the Lagrange function $\hat{L}(u, \lambda) := F(u) + \lambda g(u)$ with $\lambda \in \mathbb{R}$. If there exists a u such that $g(u) < 0$, then the set of Lagrange multipliers is nonempty. Then a feasible point u^* is a solution if and only if there exists $\lambda^* \geq 0$ such that

$$0 \in \partial_u L(u^*, \lambda^*) = DF(u^*) + \lambda^* \partial g(u^*) \quad \text{and} \quad \lambda^* g(u^*) = 0.$$

We will give a couple of equivalent characterizations of $\partial g(u^*)$. Let $g_{x,z}(u) := [\hat{\mathcal{S}}(u)](x,z) - \alpha$. First, we consider the viewpoint of Clarke [7, Section 2.8], where this subdifferential can be represented by the set

$$\partial g(u^*) = \left\{ \int_{\bar{D} \times \Xi} \partial g_{x,z}(u^*) \, \mathsf{d}\mu(x,z) \,\middle|\, \mu \text{ is a Radon probability measure supported on } M(u^*) \right\},$$

and

$$M(u) := \{(x,z) \in \bar{D} \times \Xi : [\hat{\mathcal{S}}(u)](x,z) = \max_{(x',z') \in \bar{D} \times \Xi} [\hat{\mathcal{S}}(u)](x',z')\}.$$

Since $\partial g_{x,z}(u) = Dg_{x,z}(u) = [\hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}(\cdot)](x,z) = [A^{-1}(\cdot)](x)$, we have that there exists a Radon probability measure $\tilde{\mu}$ supported on $M(u^*)$ such that

$$\langle DF(u^*), h \rangle + \lambda^* \int_{\bar{D} \times \Xi} [A^{-1}h](x) \, \mathsf{d}\tilde{\mu}(x,z) = 0 \quad \forall h \in L^2(D), \tag{67a}$$

$$\lambda^* \geq 0, \quad \max_{(x,z) \in \bar{D} \times \Xi} y^*(x,z) \leq \alpha, \quad \lambda^* \left(\max_{(x,z) \in \bar{D} \times \Xi} y^*(x,z) - \alpha \right) = 0, \tag{67b}$$

Lemma 3.15. *The conditions* (66) *are equivalent to those in* (67).

Proof. Let us assume we have (66) and we set $\mu^* =: \lambda^* \tilde{\mu}$ such that $\tilde{\mu} = (\mu^*(\bar{D} \times \Xi))^{-1}\mu^*$ and $\lambda^* = \mu^*(\bar{D} \times \Xi)$. Evidently, we have $\tilde{\mu}(\bar{D} \times \Xi) = 1$, making $\tilde{\mu}$ is a Radon probability measure and $\lambda^* \geq 0$ since μ^* is nonnegative. This construction immediately yields (67a). Now, the integral condition in (66b) becomes $\lambda^* \int_{\bar{D} \times \Xi} (y^*(x,z) - \alpha) \, \mathsf{d}\tilde{\mu}(x,z) = 0$. On account of the nonnegativity of $\tilde{\mu}$, the integrand can only be equal to zero if $y^*(x,z) = \alpha$ on a set of positive measure, and this

in turn only occurs on those points satisfying $y^*(x,z) = \max_{(x',z') \in \bar{D} \times \Xi} y^*(x',z')$. In particular, this means that $\tilde{\mu}$ is supported on $M(u^*)$ and we obtain the final condition in (67b) due to

$$\lambda^* \int_{\bar{D} \times \Xi} (y^*(x,z) - \alpha) \, \mathsf{d}\tilde{\mu}(x,z) = \lambda^* \int_{\bar{D} \times \Xi} \left(\max_{(x',z') \in \bar{D} \times \Xi} y^*(x',z') - \alpha \right) \, \mathsf{d}\tilde{\mu}(x,z)$$

$$= \lambda^* \left(\max_{(x',z') \in \bar{D} \times \Xi} y^*(x',z') - \alpha \right) = 0.$$

On the other hand, suppose we have (67). It is clear that $\mu^* := \lambda^* \tilde{\mu}$ is a nonnegative Radon measure (and hence a nonnegative regular Borel measure). One can use identical arguments to those above to reconstruct the conditions (66). □

We move on to another representation of the optimality conditions. The Ioffe-Tikhomirov Theorem [20, Theorem 4.2.3] yields the following representation of the convex subdifferential of g:

$$\partial g(u^*) = \mathrm{clco}\{\partial g_{x,z}(u^*) \mid (x,z) \in M(u^*)\},$$

where "clco" denotes the weak closure (weak-star closure in general Banach spaces) in U. Recall that $\partial g_{x,z}(u^*) = [A^{-1}(\cdot)](x)$. Then we have the optimality conditions (67b) with

$$\langle DF(u^*), h \rangle + \lambda^* \mathsf{co}\{[A^{-1}h](x) \mid (x,z) \in M(u^*)\} \ni 0 \quad \forall h \in L^2(D). \tag{68}$$

Here, "clco" could be replaced by "co" since the (finite-dimensional) set $\{[A^{-1}h](x) \mid (x,z) \in M(u^*)\}$ is compact by the compactness of $M(u^*)$ and $\hat{\mathcal{A}}^{-1}\hat{\mathcal{B}}h \in \mathcal{C}(\bar{D} \times \Xi)$. We have the following:

Lemma 3.16. *The condition* (68) *is equivalent to the existence of a Radon probability measure* $\tilde{\mu}$ *such that* (67a) *holds.*

Proof. Let $S := \{[A^{-1}h](x) \mid (x,z) \in M(u^*)\}$. From [31, Proposition 1.2] we have that a point s is an element of $\mathrm{clco}(S)$ if and only if there is a Radon probability measure on S (i.e., μ is supported on $M(u^*)$) representing s, meaning in this case that we have

$$f(s) = \int_{\bar{D} \times \Xi} f(x,z) \, \mathsf{d}\mu(x,z) \quad \forall f \in \mathcal{C}(\mathbb{R}^{d+m}).$$

Choosing $f = A^{-1}h$, we obtain the equivalence of the formulations (68) and (67a). □

References

[1] A. Alphonse, C. Geiersbach, M. Hintermüller, and T. M. Surowiec. Risk-averse optimal control of random elliptic variational inequalities. *arXiv preprint arXiv:2210.03425*, 2022.

[2] H. W. Alt. *Lineare Funktionalanalysis*. Springer, Berlin, Heidelberg, 2012. URL https://doi.org/10.1007/978-3-662-08386-4.

[3] H. Berthold, H. Heitsch, R. Henrion, and J. Schwientek. On the algorithmic solution of optimization problems subject to probabilistic/robust (probust) constraints. *Math. Oper. Res.*, 96(1):1–37, 2022. URL https://doi.org/10.1080/02331934.2019.1576670.

[4] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Science & Business Media, 2013. URL https://doi.org/10.1007/978-1-4612-1394-9.

[5] J.-B. Caillau, M. Cerf, A. Sassi, E. Trélat, and H. Zidani. Solving chance constrained optimal control problems in aerospace via kernel density estimation. *Optim. Control Appl. Methods*, 39 (5):1833–1858, 2018. URL https://doi.org/10.1002/oca.2445.

[6] A. Charnes, W. W. Cooper, and G. H. Symonds. Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Manage. Sci*, 4(3):235–263, 1958. URL https://doi.org/10.1287/mnsc.4.3.235.

[7] F. Clarke. *Optimization and Nonsmooth Analysis*. Wiley New York, 1983. URL https://doi.org/10.1137/1.9781611971309.

[8] S. Conti, H. Held, M. Pach, M. Rumpf, and R. Schultz. Shape optimization under uncertainty—a stochastic programming perspective. *SIAM J. Optim.*, 19(4):1610–1632, 2009. URL 10.1137/070702059.

[9] S. Conti, M. Rumpf, R. Schultz, and S. Tölkes. Stochastic dominance constraints in elastic shape optimization. *SIAM J. Optim.*, 56(4):3021–3034, 2018. URL https://doi.org/10.1137/16M108313X.

[10] M. H. Farshbaf-Shaker, R. Henrion, and D. Hömberg. Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. *Set-Valued Var. Anal*, 26:821–841, 2018. URL https://doi.org/10.1007/s11228-017-0452-5.

[11] M. H. Farshbaf-Shaker, M. Gugat, H. Heitsch, and R. Henrion. Optimal Neumann boundary control of a vibrating string with uncertain initial data and probabilistic terminal constraints. *SIAM J. Optim.*, 58(4):2288–2311, 2020.

[12] D. B. Gahururu, M. Hintermüller, and T. M. Surowiec. Risk-neutral PDE-constrained generalized Nash equilibrium problems. *Mathematical Programming*, pages 1–51, 2022. URL https://doi.org/10.1007/s10107-022-01800-z.

[13] C. Geiersbach and M. Hintermüller. Optimality conditions and Moreau–Yosida regularization for almost sure state constraints. *ESAIM: COCV*, 28:Paper No. 80, 36, 2022. URL https://doi.org/10.1051/cocv/2022070.

[14] C. Geiersbach and W. Wollner. Optimality conditions for convex stochastic optimization problems in Banach spaces with almost sure state constraints. *SIAM J. Optim.*, 4(31):2455–2480, 2021. URL https://doi.org/10.1137/20M1363558.

[15] Geletu, Abebe, Hoffmann, Armin, Schmidt, Patrick, and Li, Pu. Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. *ESAIM: COCV*, 26:70, 2020. URL https://doi.org/10.1051/cocv/2019077.

[16] S. Göttlich, O. Kolb, and K. Lux. Chance-constrained optimal inflow control in hyperbolic supply systems with uncertain demand. *Optim. Control Appl. Methods*, 42:566–589, 2021. URL https://doi.org/10.1002/oca.2689.

[17] R. Haller, H. Meinlschmidt, and J. Rehberg. Hölder regularity for domains of fractional powers of elliptic operators with mixed boundary conditions. 2022. URL https://www.wias-berlin.de/publications/wias-publ/run.jsp?template=abstract&type=Preprint&year=2022&number=2959.

[18] A. Hantoute, R. Henrion, and P. Pérez-Aros. Subdifferential characterization of probability functions under Gaussian distribution. *Math. Program.*, 174(1-2, Ser. B):167–194, 2019. URL https://doi.org/10.1007/s10107-018-1237-9.

[19] A. D. Ioffe and V. L. Levin. Subdifferentials of convex functions. *Trudy Moskovskogo Matematicheskogo Obshchestva*, 26:3–73, 1972.

[20] A. D. Ioffe and V. M. Tikhomirov. *Theory of extremal problems*. Elsevier North-Holland Amsterdam, 1979.

[21] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*, volume 88 of *Pure and Applied Mathematics*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. URL https://doi.org/10.1137/1.9780898719451.

[22] P. Kolvenbach, O. Lass, and S. Ulbrich. An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty. *Optim. Eng.*, 19(3):697–731, Sep 2018. URL https://doi.org/10.1007/s11081-018-9388-3.

[23] D. P. Kouri and T. M. Surowiec. Existence and optimality conditions for risk-averse PDE-constrained optimization. *SIAM-ASA J. Uncertain. Quantif.*, 6(2):787–815, 2018. URL https://doi.org/10.1137/16m1086613.

[24] D. P. Kouri, M. Staudigl, and T. M. Surowiec. A relaxation-based probabilistic approach for PDE-constrained optimization under uncertainty with pointwise state constraints. *Comput. Optim. Appl., URL = https://doi.org/10.1007/s10589-023-00461-8,*, Feb. 2023.

[25] V. L. Levin. The Lebesgue decomposition for functionals on the vector-function space L_X^∞. *Funct. Anal. its Appl.*, 8(4):314–317, 1974. URL https://doi.org/10.1007/bf01075488.

[26] J. Martínez-Frutos and F. Esparza. *Optimal Control of PDEs Under Uncertainty: An Introduction with Application to Optimal Shape Design of Structures*. BCAM SpringerBriefs. Springer, 2018. URL https://doi.org/10.1007/978-3-319-98210-6.

[27] J. Milz. Consistency of Monte Carlo estimators for risk-neutral PDE-constrained optimization. *Appl. Math. Optim.*, 87(3):57, 2023. URL https://doi.org/10.1007/s00245-023-09967-3.

[28] B. S. Mordukhovich. *Variational Analysis and Generalized Differentiation Vol. I: Basic Theory*. Springer Berlin Heidelberg New York, 2006.

[29] P. Pérez-Aros, C. Quiñinao, and M. Tejo. Control in probability for sde models of growth population. *Appl. Math. Optim.*, 86(3):44, 2022. URL https://doi.org/10.1007/s00245-022-09915-7.

[30] J. Peypouquet. *Convex Optimization in Normed Spaces: Theory, Methods and Examples*. SpringerBriefs in Optimization. Springer International Publishing, 2015. URL https://books.google.de/books?id=puwEogEACAAJ.

[31] R. R. Phelps. *Lectures on Choquet's theorem*. Springer, 2001. URL https://doi.org/10.1007/b76887.

[32] A. Prékopa. *Stochastic programming*. Dordrecht: Kluwer Academic Publishers, 1995. URL https://doi.org/10.1007/978-94-017-3087-7.

[33] R. Rockafellar and R. Wets. Stochastic convex programming: singular multipliers and extended duality singular multipliers and duality. *Pac. J. Appl. Math.*, 62(2):507–522, 1976. URL https://doi.org/10.2140/pjm.1976.62.507.

[34] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: Kuhn–Tucker conditions. *Journal of Mathematical Economics*, 2(3):349–370, 1975. URL https://doi.org/10.1016/0304-4068(75)90003-8.

[35] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: basic duality. *Pac. J. Appl. Math.*, 62(1):173–195, 1976. URL https://doi.org/10.2140/pjm.1976.62.173.

[36] R. T. Rockafellar and R. J.-B. Wets. Stochastic convex programming: relatively complete recourse and induced feasibility. *SIAM J. Optim.*, 14(3):574–589, 1976. URL https://doi.org/10.1137/0314038.

[37] R. A. Ryan. *Introduction to tensor products of Banach spaces*. Springer Monographs in Mathematics. Springer-Verlag London, Ltd., London, 2002. URL https://doi.org/10.1007/978-1-4471-3903-4.

[38] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming—modeling and theory*, volume 28 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2021. URL https://doi.org/10.1137/1.9781611976595. Third edition.

[39] K. Teka, A. Geletu, and P. Li. Structural properties and convergence approach for chance-constrained optimization of boundary-value elliptic partial differential equation systems. In B. Yang and D. Stipanovic, editors, *Nonlinear Systems*, chapter 7. IntechOpen, Rijeka, 2023. URL https://doi.org/10.5772/intechopen.104620.

[40] W. van Ackooij. A discussion of probability functions and constraints from a variational perspective. *Set-Valued Var. Anal*, 28(4):585–609, 2020. URL https://doi.org/10.1007/s11228-020-00552-2.

[41] W. van Ackooij and R. Henrion. Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. *SIAM J. Optim.*, 24(4):1864–1889, 2014. URL https://doi.org/10.1137/130922689.

[42] W. van Ackooij, R. Henrion, and P. Pérez-Aros. Generalized gradients for probabilistic/robust (probust) constraints. *Optimization*, 69:1451 – 1479, 2020. URL https://doi.org/10.1080/02331934.2019.1576670.