SIAM/ASA J. Uncertainty Quantification
Vol. 13, No. 3, pp. 1028–1053

Optimal Control under Uncertainty with Joint Chance State Constraints: Almost-Everywhere Bounds, Variance Reduction, and Application to (Bi)linear Elliptic PDEs*

René Henrion†, Georg Stadler‡, and Florian Wechsung‡

Abstract. We study optimal control of partial differential equations (PDEs) under uncertainty with the state variable subject to joint chance constraints. The controls are deterministic, but the states are probabilistic due to random variables in the governing equation. Joint chance constraints ensure that the random state variable meets pointwise bounds with high probability. For linear governing PDEs and elliptically distributed random parameters, we prove existence and uniqueness results for almost-everywhere state bounds. Using the spherical-radial decomposition (SRD) of the uncertain variable, we prove that when the probability is very large or small, the resulting Monte Carlo estimator for the chance constraint probability exhibits substantially reduced variance compared to the standard Monte Carlo estimator. We further illustrate how the SRD can be leveraged to efficiently compute derivatives of the probability function, and we discuss different expansions of the uncertain variable in the governing equation. Numerical examples for linear and bilinear PDEs compare the performance of Monte Carlo and quasi-Monte Carlo sampling methods, examining probability estimation convergence as the number of samples increases. We also study how the accuracy of the probabilities depends on the truncation of the random variable expansion, and we numerically illustrate the variance reduction of the SRD.

Key words. optimal control under uncertainty, risk-averse, joint chance state constraints, spherical-radial decomposition, elliptic PDE, variance reduction

MSC codes. 90C15, 65K10, 35Q93, 60H35, 35R60

DOI. 10.1137/24M171557X

1. Introduction.

While deterministic optimization and control problems governed by partial differential equations (PDEs) make up a well-established research area [6, 23, 39], it has only been over the last one or two decades that researchers have started to systematically study these problems in the presence of uncertainty; see, e.g., [2, 10, 16, 25, 26, 29]. Optimization under uncertainty introduces new theoretical and algorithmic difficulties because uncertainty can substantially increase the dimensionality of the problem. This adds complexity to optimization problems governed by PDEs, as it necessitates random space approximation, and

*Received by the editors December 6, 2024; accepted for publication (in revised form) April 18, 2025; published electronically August 6, 2025.
https://doi.org/10.1137/24M171557X
Funding: The work of the first author was supported by the Deutsche Forschungsgemeinschaft within CRC 154 (project B04) and by the PGMO program funded by Électricité de France. The work of the second and third authors was partially supported by the Simons Foundation/SFARI (560651, AB). The second author was also supported by the U.S. Department of Energy, Office of Science Energy Earthshot Initiative under Award DE-SC0024721.

†Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany (henrion@wias-berlin.de).
‡Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 USA (stadler@cims.nyu.edu, fw18@nyu.edu).

one requires adapted objectives and constraints. Here we focus on probabilistic constraints on the state in optimal control problems governed by elliptic equations. In particular, we consider joint chance state constraints, which require that realizations of the distribution of states satisfy pointwise bound constraints with a certain probability. A precise formulation is given next.

1.1. Problem statement. For a bounded, sufficiently regular domain $\mathcal{D} \subset \mathbb{R}^d$, typically $d \in \{1, 2, 3\}$, we consider an optimal control problem under uncertainty, i.e.,

$$\min_{u \in U_{\text{ad}},\, y \in X} \mathcal{J}(u, y), \tag{1.1}$$

where U_{ad} is a convex closed set of admissible controls, and X is the function space of the random states. The uncertainty enters through the equation that relates the control u and the state y. We use a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, i.e., $\omega \in \Omega$ are events, $\mathcal{A}$ is a σ-algebra on Ω, and $\mathbb{P}$ is a probability measure. The governing equation is an elliptic PDE with random variable ξ defined in $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in a Hilbert space $\mathscr{H}$. The state $y(\omega)$ and the control u satisfy the equation

$$A(u)y(\omega) + B\xi(\omega) + Cu = f, \tag{1.2}$$

where for $u \in U_{\text{ad}}$, $A(u) : Y \mapsto H^{-1}(\mathcal{D})$ with Y a subspace of $H^1(\mathcal{D})$ is a linear invertible elliptic operator that is differentiable with respect to u, $B : \mathscr{H} \mapsto H^{-1}(\mathcal{D})$ and $C : U_{\text{ad}} \to L^2(\mathcal{D})$ are bounded linear operators, and $f \in L^2(\mathcal{D})$.

The equation form (1.2) includes linear and bilinear PDEs, and we will give concrete examples later in this introduction. In addition to (1.1) and (1.2), for $p \in [0, 1]$ we consider the following joint chance constraints for the state variable:

$$\mathbb{P}(\omega \in \Omega \,|\, \underline{y}(\boldsymbol{x}) \le y(\boldsymbol{x}, \omega) \le \bar{y}(\boldsymbol{x}) \text{ for a. a. } \boldsymbol{x} \in \mathcal{D}) \ge p, \tag{1.3}$$

where $\underline{y}, \bar{y} \in L^2(\mathcal{D})$ are given lower and upper bounds, respectively. The constraint (1.3) states that realizations of the state, which is a random variable, must be between $\underline{y}$ and $\bar{y}$ in a pointwise sense with probability of at least p. In this paper, we develop efficient methods to approximate and solve (1.1), (1.2), and (1.3) numerically.

1.2. Related literature. Chance constraints were introduced in the 1950s [9], and fundamental contributions to their theoretical and numerical foundations were later made by Prékopa [35]; see [37] for a more modern presentation. Recently, these probabilistic constraints have been considered in the context of infinite-dimensional optimization, in particular PDE-constrained optimization [10, 13, 14, 17, 20, 21, 22, 27, 34, 36, 38, 42]. These works focus on applications (for example, optimization of gas transport [22, 36], aeronautics [8], and population growth [34]), numerical approaches (see, for example, [10, 20, 27]), andtheoretical aspects (for example, existence and stability of solutions [14], convergence of algorithms [38], optimality conditions [17, 42], and explicit estimates for subdifferentials [21]).

Numerical approaches for chance constraints can roughly be classified in (1) methods that relax the constraints, resulting in formulations with nicer properties (e.g., convexity, differentiability); and (2) methods that approximate the original chance constraints. One

method in the first class is using the conditional value-at-risk, but other relaxations are possible (see, e.g., [27, 30]). Facilitating solution methods through relaxation of the chance constraint comes at the cost of potentially significant conservatism of the resulting solutions. The second class of approaches approximate the original chance constraints. This approximation could be discrete (as in sample average approximation [32]) or smooth (see, e.g., [8, 19, 33, 40]). Consequently, solution algorithms from discrete or nonlinear optimization are used. In this paper, we follow an approach based on the *spherical-radial decomposition* (SRD) of elliptically distributed (e.g., Gaussian) random vectors, which has been shown to be successful inboth theoretical analysis of probability functions [17, 21, 40, 41, 42] and numerical solutions of chance-constrained optimization problems [3, 13, 22, 34].

We emphasize that (1.3) is a *joint* chance constraint, that is, it considers the probability over a whole system of random inequalities. Alternatively, one could formulate *individual* chance constraints, where the probability is taken over all inequalities individually as follows:

$$\mathbb{P}\left(\omega \in \Omega \,|\, \underline{y}(\boldsymbol{x}) \le y(\boldsymbol{x},\omega) \le \bar{y}(\boldsymbol{x})\right) \ge p \qquad \text{for a. a. } \boldsymbol{x} \in \mathcal{D}.$$

The difference lies in enforcing uniform versus pointwise constraints, which must be satisfied with high probability. Individual constraints are less restrictive, allowing for lower optimal objective function values. Furthermore, in specific models (e.g., when control and random parameters are separated as in (1.2)), individual constraints can be reformulated into explicit deterministic equivalents. In contrast, joint constraints address the need for uniform constraint satisfaction, often reflecting practical requirements. In particular, random states that meet individual bounds with high probability at each point may still fail to meet these bounds uniformly over the entire domain with high probability (see, e.g., [3, sect. 4.2]).

Note that in the limit $p \to 1$, the chance constraint (1.3) turns into an "almost-sure" constraint. Nevertheless, in general, it is not recommended to formulate an almost-sure constraint as a chance constraint because the latter degenerates in the limit to a constraint violating standard constraint qualifications [17]. We refer the reader to [15, 17, 18] for theoretical and numerical approaches to treating almost-sure constraints.

1.3. Contributions, limitations, and overview. In this paper, we make the following contributions. (1) We extend existence, uniqueness, and convexity results to problems with joint chance *almost-everywhere* state constraints. (2) We prove theoretically and illustrate numerically that, compared to standard Monte Carlo (MC) sampling, using the spherical-radial decomposition (SRD) substantially reduces the variance for estimation of low, high, or radially close-to-symmetric probabilities. (3) Using smoothing properties characteristic of PDE operators, we illustrate the innate dimension reduction for the random state variable and discuss its use in approximating chance constraints. (4) We present a systematic numerical study of control problems governed by linear and bilinear elliptic problems, comparing standard, SRD-based MC, and quasi-MC samplings. We find that an SRD quasi-MC method requires 1–2 orders of magnitude fewer samples than a standard MC method for the same accuracy, which directly translates to a corresponding speedup.

Our approach also has some limitations. (1) For the problems we consider here, the map from the uncertainty to the PDE solution (and thus the constraint) is linear. The methods can be generalized to problems where the constraint function depends nonlinearly

on the random variables. However, some advantages of the proposed methods are lost with increasing nonlinearity. (2) We also assume that for fixed control u, the governing equation is linear. Note, however, that we do not require that the control enter linearly, as we illustrate using a bilinear control example. (3) For the problems considered here, the variance reduction achieved by using SRD MC sampling over standard MC sampling directly translates into computational speedup. Although variance reduction also applies to problems where the constraint depends nonlinearly on the random variable, the increased computational cost of SRD MC in nonlinear settings may negate the advantages gained from reduced variance.

Finally, we give an overview. Section 2 presents basic existence, uniqueness, and convexity results for the case where the governing equation (1.2) is linear in the control u. In particular, existing results are generalized to *almost-everywhere* state constraints. In section 3, the spherical-radial decomposition (SRD) is introduced for general elliptically distributed random variables, and reduced variance results are shown for the case of very small, large, or close-to-symmetric sets. Furthermore, we provide derivatives of the probability function. In section 4, we apply the general framework specifically to the control problem (1.1), discuss truncation of the random variable expansion, and compute explicit gradient formulas of the probability. Finally, in sections 5 and 6, we perform comprehensive numerical experiments for linear and bilinear control problems and discuss the results in the context of our theoretical findings.

1.4. Examples. Next, we present two examples that fit the setting described in subsection 1.1. These examples are studied numerically in sections 5 and 6.

Example 1.1 (tracking-type objective governed by linear PDE with uncertain Neumann data). We split the boundary of the physical domain $\mathcal{D} \subset \mathbb{R}^d$ into disjoint open sets $\partial\mathcal{D}_1, \partial\mathcal{D}_2$ and assume we are given a Gaussian law $d\mu$ with realizations $\xi(\omega) \in L^2(\partial\mathcal{D}_2)$. We consider the risk-neutral tracking-type optimal control problem

$$\underset{u\in L^2(\mathcal{D}),\, y\in X}{\text{minimize}} \quad \frac{1}{2}\int_\Omega \int_{\mathcal{D}} (y(\omega) - y_d)^2 \, d\boldsymbol{x}\, d\mu + \frac{\alpha}{2}\int_{\mathcal{D}} u^2 \, d\boldsymbol{x} \tag{1.4}$$

subject to a governing equation with uncertain Neumann data on $\partial\mathcal{D}_2$,

$$-\Delta y(\omega) = f + u \qquad \text{in } \mathcal{D}, \tag{1.5a}$$

$$y(\omega) = 0 \qquad \text{in } \partial\mathcal{D}_1, \tag{1.5b}$$

$$\nabla y(\omega) \cdot \boldsymbol{n} = \xi(\omega) \qquad \text{in } \partial\mathcal{D}_2, \tag{1.5c}$$

and joint state chance constraints (1.3). Here, $y_d, f \in L^2(\mathcal{D})$, $\alpha > 0$, and $\boldsymbol{n}$ is the unit boundary normal and X is the (Bochner) space of solutions of (1.5). In section 5, we show that the integration over Ω in (1.4) can be done analytically, and we study methods for approximating the joint chance constraint and the effect of the bound p on the controls u that minimize (1.4).

Example 1.2 (bilinear control with uncertain right-hand side). In bilinear control, the governing equation contains a term in which the control u multiplies the state y. This structure appears, for example, in the optimal control of quantum systems [4, 5, 24]. We assume an uncertain right-hand side with realizations $\xi(\omega) \in L^2(\mathcal{D})$ resulting in the governing equation

$$-\Delta y(\omega) + uy(\omega) = f + \xi(\omega) \quad \text{in } \mathcal{D}, \tag{1.6a}$$

$$y(\omega) = 0 \quad \text{in } \partial\mathcal{D}, \tag{1.6b}$$

where $f \in H^{-1}(\mathcal{D})$. We consider an objective that only involves the control cost; i.e., the optimization-under-uncertainty problem with joint chance state constraints is

$$\underset{u\in U_{\text{ad}},\, y\in X}{\text{minimize}} \ \frac{1}{2}\int_{\mathcal{D}} (u-u_0)^2\, d\boldsymbol{x}, \tag{1.7}$$

subject to y satisfying (1.6) and the joint chance constraint (1.3). Here, $u_0 \in L^2(\mathcal{D})$ is a given desired control, and if U_{ad} includes, for instance, the pointwise bound $u \geq 0$, invertibility of the PDE operator in (1.6) is guaranteed. Due to the bilinear equation, this is *not* a linear-quadratic problem. However, for fixed u, the map from the uncertain parameter ξ to the state variable y is linear.

1.5. Notation. We denote the spatial domain by $\mathcal{D}$ and use Ω for the random space. To distinguish vectors from scalars or scalar functions, we use a bold font. We use the common abbreviations SRD for spherical-radial decomposition and MC for Monte Carl. The number of unknowns to discretize $\mathcal{D}$ is denoted by n. When we use Karhunen–Loeve expansions to describe random variables, we use K for the number of expansion terms and the index k. To approximate the random space Ω, we denote by N the number of MC samples and use the index i for these samples. By M we denote the number of points in $\mathcal{D}$, where the state constraints are evaluated, and we will use the index j for this point set.

2. Solution existence, uniqueness, and convexity for linear governing equation.

In this section, we consider the case where the PDE operator A in (1.2) does not depend on the control; i.e., the governing equation is

$$Ay(\omega) + B\xi(\omega) + Cu = f. \tag{2.1}$$

Since A is assumed to be invertible, we can consider the (random) state variable y as a function of the (deterministic) control u and the (random) variable ξ as follows:

$$y^u(\omega) = A^{-1}(f - Cu - B\xi(\omega)). \tag{2.2}$$

We assume that the random variable ξ is defined as a linear combination of finitely many functions $\xi_0, e_k \in L^2(\mathcal{D})$ $(k = 0, \ldots, K)$,

$$\xi(\omega) := \xi_0 + \sum_{k=1}^{K} \zeta_k(\omega) e_k, \tag{2.3}$$

where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_K)$ is a K-dimensional random vector on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Assume further that $U_{\text{ad}} \subset L^2(\mathcal{D})$, and consider the probability function

$$\varphi(u) := \mathbb{P}\left(\omega \in \Omega : \underline{y}(\boldsymbol{x}) \leq y^u(\boldsymbol{x}, \omega) \leq \overline{y}(\boldsymbol{x}) \text{ for a. a. } \boldsymbol{x} \in \mathcal{D}\right) \tag{2.4}$$

and the associated set of controls that satisfy the chance constraint (1.3), i.e.,

$$U_{\text{pr}} := \left\{u \in L^2(\mathcal{D}) : \varphi(u) \geq p\right\}. \tag{2.5}$$

The next proposition summarizes the properties of $\varphi(\cdot)$ and U_{pr}.

Theorem 2.1. *The probability function* (2.4) *is well defined and weakly sequentially upper semicontinuous. As a consequence, the feasible set* U_{pr} *defined in* (2.5) *is weakly sequentially closed for arbitrary* $p \in [0,1]$. *If the random vector* $\boldsymbol{\zeta}$ *has a log-concave density (e.g., Gaussian and many others; see* [35, *sect.* 4.4]*), then* U_{pr} *is also convex for arbitrary* $p \in [0,1]$.

Proof. To show that φ is well defined, we verify that for any fixed control u and corresponding state y^u, the event set

$$\{\omega \in \Omega : \underline{y}(\boldsymbol{x}) \le y^u(\boldsymbol{x},\omega) \le \overline{y}(\boldsymbol{x}) \text{ for a. a. } \boldsymbol{x} \in \mathcal{D}\} \tag{2.6}$$

belongs to the σ-algebra $\mathcal{A}$ of the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. To this aim, we introduce the continuous linear solution operator $S : L^2(\mathcal{D}) \times \mathbb{R}^K \to Y$ as

$$S(u,\boldsymbol{z}) := A^{-1}\left(Cu + \sum_{k=1}^{K} z_k B e_k\right) \qquad \left(u \in L^2(\mathcal{D}),\, \boldsymbol{z} := (z_1,\dots,z_K) \in \mathbb{R}^K\right)$$

and define the convex and closed set

$$\hat{K} := \{\tilde{y} \in H^1(\mathcal{D}) : A^{-1}(f - B\xi_0) - \overline{y}(\boldsymbol{x}) \le \tilde{y}(\boldsymbol{x}) \le A^{-1}(f - B\xi_0) - \underline{y}(\boldsymbol{x}) \text{ for a. a. } \boldsymbol{x} \in \mathcal{D}\}.$$

The closedness of $\hat{K}$ follows from the facts that converging sequences in $\hat{K}$ also converge in $L^2(\mathcal{D})$ and that the set of L^2-functions with almost-everywhere bounds is closed [39, sect. 2.5]. Then, the event set in (2.6) coincides with the set

$$\{\omega \in \Omega : S(u,\boldsymbol{\zeta}(\omega)) \in \hat{K}\} = \boldsymbol{\zeta}^{-1}([S(u,\cdot)]^{-1}(\hat{K})), \tag{2.7}$$

which belongs to the σ-algebra $\mathcal{A}$ because the set $[S(u,\cdot)]^{-1}(\hat{K})$ is closed as the preimage of the closed set $\hat{K}$ under the continuous mapping $S(u,\cdot)$. Therefore, φ is well defined.

To show the remaining statements, we apply Lemmas A.1 and A.2 from the appendix. In order to do so, we specify the abstract spaces and functions from these lemmas as follows:

$$U := L^2(\mathcal{D}),\ \tilde{Y} := Y,\ K := \hat{K},\ g := S.$$

We verify first that the assumptions of Lemma A.2 are satisfied, namely that K is weakly closed and g is weakly sequentially continuous. This is indeed the case because, first, $\hat{K}$ is weakly closed as a closed, convex subset of the Banach space Y and, second, because S is weakly sequentially continuous as a continuous linear operator. Then, by Lemma A.2, the probability function $\tilde{\varphi}$ defined there is weakly sequentially upper semicontinuous. On the other hand, by the coincidence of (2.6) and (2.7), the probability function $\tilde{\varphi}$ of Lemma A.2 is identical to the probability function φ in (2.4) in the concrete setting of the current theorem. Thus, φ is weakly sequentially upper semicontinuous.

Likewise, the assumptions of Lemma A.1 are satisfied when we additionally assume in our theorem that $\boldsymbol{\zeta}$ has a log-concave density. Hence, with the definitions made before, the set

$$M := \{u \in U : \mathbb{P}(g(u,\boldsymbol{\zeta}) \in K) \ge p\} = \{u \in U : \mathbb{P}(S(u,\boldsymbol{\zeta}) \in \hat{K}) \ge p\}$$

defined in that lemma is convex for arbitrary $p \in [0,1]$. Taking into account the equivalence of (2.7) and (2.6) mentioned above as well as the definition of (2.5) via (2.4), we end up at the identity $M = U_{\rm pr}$, so that $U_{\rm pr}$ is convex for any $p \in [0,1]$, as was to be shown. ∎

Note that in the original formulation (1.1), we consider the control u and the state y as independent variables that are connected by the equality constraint (1.2). The previous result allows us to argue existence and uniqueness for the reduced form of (1.1), (1.2), (1.3), in which we consider the state $y = y^u$ as a function of the control u using (2.1) as follows:

$$\min_{u \in U_{\rm ad} \cap U_{\rm pr}} \hat{\mathcal{J}}(u), \quad \text{where } \hat{\mathcal{J}}(u) := \mathcal{J}(u, y^u). \tag{2.8}$$

Theorem 2.2. *Assume that $\hat{\mathcal{J}}$ is coercive, convex, and lower semicontinuous and that $U_{\rm ad} \cap U_{\rm pr} \neq \emptyset$. Then, (2.8) admits a solution. The same holds if we replace the coercivity of $\hat{\mathcal{J}}$ by the additional boundedness of $U_{\rm ad}$. If, in either of the two settings, $\hat{\mathcal{J}}$ is strictly convex and $\boldsymbol{\zeta}$ has a log-concave density, then (2.8) has a unique solution.*

Proof. By convexity and lower semicontinuity, $\hat{\mathcal{J}}$ is weakly sequentially lower semicontinuous. Since $U_{\rm ad}$ is closed and convex, it is also weakly sequentially closed. With $U_{\rm pr}$ being weakly sequentially closed by Theorem 2.1, $U_{\rm ad} \cap U_{\rm pr}$ is weakly sequentially closed. Since a coercive, weakly sequentially lower semicontinuous function attains its infimum over a nonempty and weakly sequentially closed set, there exists a solution to (2.8).

Assume now that $\hat{\mathcal{J}}$ is not necessarily coercive and that $U_{\rm ad}$ is additionally bounded. Then, as a closed, convex, and bounded subset of a reflexive Banach space, $U_{\rm ad}$ is weakly sequentially compact. As a consequence, the nonempty intersection with the weakly sequentially closed set $U_{\rm pr}$ is weakly sequentially compact, too. Now the existence of a solution to problem (2.8) follows in a classical way from the weak sequential lower semicontinuity of $\hat{\mathcal{J}}$. If $\boldsymbol{\zeta}$ has a log-concave density, then $U_{\rm pr}$ is convex by Theorem 2.1. Therefore, $U_{\rm ad} \cap U_{\rm pr}$ is convex, and the result follows from the strict convexity of $\hat{\mathcal{J}}$. ∎

3. Spherical-radial decomposition.

In this section, we revisit the spherical-radial decomposition (SRD) as a methodology for addressing chance constraints in a slightly more general context than required for our target optimal control problem (1.1)–(1.3). In particular, in subsections 3.1 and 3.2, we assume an elliptically distributed random variable (generalizing the Gaussian distribution in our target problem) and that the constraint function is convex (generalizing the linear constraint function needed in (1.3)).

The SRD has a somewhat greater level of complexity compared to smoothing methods [8, 33, 36] and sample average approximation [32], which are applicable to any distribution accessible through sampling. Although the SRD is confined to a specific class of multivariate distributions, including Gaussian distributions, it capitalizes on their intrinsic properties to provide three significant benefits. First, it facilitates the analysis of the original problem (e.g., Lipschitz continuity or differentiability of the chance constraint) without necessitating approximations at the outset [21, 40, 41]. This approach has been utilized in [17, 42] to derive optimality conditions for a continuous PDE optimal control problem. Second, the SRD has reduced variance in probability estimation compared to both direct MC and quasi-MC samplings, which is advantageous for numerically solving chance-constrained optimization problems. Third, the SRD provides explicit gradient expressions for the probability function

in both the original and numerically approximated formulations. This is crucial for addressing chance constraints through nonlinear optimization methods. While gradients for numerical approximations are also obtainable with smoothing methods, they cannot be derived from single samples, unlike when using the SRD. The first aspect is less significant in this context as our focus is on numerical solutions rather than theoretical analysis. The second aspect, concerning variance reduction, has not been systematically explored so far, and we shall examine it more thoroughly in subsection 3.2. As for the third aspect, we will derive and employ explicit gradient formulas for the probability functions to solve the optimal control problems outlined in Examples 1.1 and 1.2.

3.1. General setting. It is well known [12, eq. (2.12)] that *elliptically* distributed n-dimensional random vectors $\boldsymbol{\zeta}$ admit a decomposition

$$\boldsymbol{\zeta} = \boldsymbol{m} + \tau L\boldsymbol{\theta}, \tag{3.1}$$

where $\boldsymbol{m} \in \mathbb{R}^n$, $L \in \mathbb{R}^{n\times k}$, $k \leq n$, τ is a one-dimensional nonnegative random variable, and $\boldsymbol{\theta}$ is a random vector uniformly distributed on the unit sphere $\mathbb{S}^{k-1}$ of $\mathbb{R}^k$. Moreover, τ and $\boldsymbol{\theta}$ are independently distributed. With τ and $\boldsymbol{\theta}$ playing the roles of a radial variable and a spherical variable (in the sense of polar coordinates), respectively, (3.1) is referred to as the spherical-radial decomposition of $\boldsymbol{\zeta}$. A prominent example is a (nondegenerate) Gaussian random vector $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{m}, \Sigma)$, which yields the decomposition (3.1) with $k = n$, τ having a chi distribution with n degrees of freedom, and L satisfying $LL^T = \Sigma$. Accordingly, the Gaussian probability of some measurable subset $M \subseteq \mathbb{R}^n$ can be represented as

$$\mathbb{P}(\boldsymbol{\zeta} \in M) = \int_{\boldsymbol{v}\in\mathbb{S}^{n-1}} \mu_\chi(\{r \geq 0 : \boldsymbol{m} + rL\boldsymbol{v} \in M\}) d\mu_U(\boldsymbol{v}), \tag{3.2}$$

where μ_χ is the one-dimensional chi distribution with n degrees of freedom, and μ_U is the uniform distribution over $\mathbb{S}^{n-1}$.

In this paper, we use the SRD in combination with the inequality $g(u, \boldsymbol{\zeta}) \leq 0$, where $g : U \times \mathbb{R}^n \to \mathbb{R}$, U is a Banach space of controls, and $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{m}, \Sigma)$ is an n-dimensional Gaussian random vector. We assume that $g(u, \cdot)$ is convex for all $u \in U$. Defining the probability function $\varphi : U \to [0, 1]$ as

$$\varphi(u) := \mathbb{P}(g(u, \boldsymbol{\zeta}) \leq 0) \qquad (u \in U), \tag{3.3}$$

(3.2) yields that

$$\varphi(u) = \int_{\boldsymbol{v}\in\mathbb{S}^{n-1}} \mu_\chi(\{r \geq 0 : g(u, \boldsymbol{m} + rL\boldsymbol{v} \leq 0\}) d\mu_U(\boldsymbol{v}) \qquad (u \in U).$$

A considerably more convenient representation can be provided if the mean $\boldsymbol{m}$ is strictly feasible for $u \in U$ as follows:

$$g(u, \boldsymbol{m}) < 0. \tag{3.4}$$

In [40, Prop. 3.11] it was shown that (3.4) holds if $\varphi(u) \geq 0.5$ and there exists *any* Slater point $\boldsymbol{z} \in \mathbb{R}^n$ (not necessarily $\boldsymbol{m}$), i.e., $g(u, \boldsymbol{z}) < 0$. Both conditions are mild; the latter holds

because the target probability level p in a chance constraint is usually close to one. Under (3.4), the one-dimensional set $\{r \geq 0 : g(u, \boldsymbol{m} + rL\boldsymbol{v}) \leq 0\}$ reduces to the interval $[0, \rho(u, \boldsymbol{v})]$, where

$$\rho(u, \boldsymbol{v}) := \sup\{r \geq 0 : g(u, \boldsymbol{m} + rL\boldsymbol{v}) \leq 0\}. \tag{3.5}$$

Note that $\rho(u, \boldsymbol{v}) = \infty$ is possible. Hence,

$$\mu_\chi(\{r \geq 0 : g(u, \boldsymbol{m} + rL\boldsymbol{v}) \leq 0\}) = \mu_\chi([0, \rho(u, \boldsymbol{v})]) = F_\chi(\rho(u, \boldsymbol{v})) - F_\chi(0) = F_\chi(\rho(u, \boldsymbol{v})),$$

where F_χ denotes the distribution function of the one-dimensional chi distribution with n degrees of freedom. Here, to include the case $\rho(u, \boldsymbol{v}) = \infty$, we adopt the convention that $F_\chi(\infty) = 1$. Consequently, the probability function simplifies to

$$\varphi(u) = \int_{\boldsymbol{v} \in \mathbb{S}^{n-1}} F_\chi(\rho(u, \boldsymbol{v})) d\mu_U(\boldsymbol{v}) \qquad (u \in U). \tag{3.6}$$

For numerical purposes, the spherical integral (3.6) is usually approximated by a finite sum

$$\varphi(u) \approx \tilde{\varphi}(u) := N^{-1} \sum_{i=1}^{N} F_\chi(\rho(u, \boldsymbol{v}_i)) \qquad (u \in U), \tag{3.7}$$

where $\{\boldsymbol{v}_i\}_{i=1}^N \subseteq \mathbb{S}^{n-1}$ are samples uniformly distributed on the sphere. Such samples can be obtained, for example, by normalizing to unit length a set of MC or quasi-MC samples of the standard Gaussian distribution $\mathcal{N}(0, I)$ in $\mathbb{R}^n$. For more efficient methods of creating such samples, see, e.g., [1].

Although the following analysis focuses on Gaussian distributions, the methods also apply to other elliptical distributions, including multivariate t-distributions, symmetric Laplace, or hyperbolic distributions. The main difference lies in the precise form of the one-dimensional radial distribution (i.e., the chi distribution for the Gaussian case), whereas the spherical component remains consistent. Additionally, techniques from the Gaussian framework can be adapted to Gaussian-like distributions, including log-normal, truncated Gaussian, and Gaussian mixture distributions.

3.2. Variance reduction by spherical-radial decomposition. Here, we assume (3.4) to ensure the validity of the form (3.6). We will show that for large sets (and thus high probability), small sets (and thus low probability), and radially near-symmetric sets about the mean, the variance ratio of the probability estimators obtained with the SRD and standard MC approaches zero (as both tend to zero individually). We fix a control u (which does not play any role in this analysis) and recall the standard MC estimator

$$\varphi(u) \approx N^{-1} \#\{i \in \{1, \ldots, N\} : g(u, \boldsymbol{z}_i) \leq 0\} = N^{-1} \sum_{i=1}^{N} \mathcal{I}_{\{\boldsymbol{z} \in \mathbb{R}^n : g(u, \boldsymbol{z}) \leq 0\}}(\boldsymbol{z}_i), \tag{3.8}$$

where $\{\boldsymbol{z}_i\}_{i=1}^N \subseteq \mathbb{R}^n$ are samples of the Gaussian random vector $\boldsymbol{\zeta}$, and $\mathcal{I}_C$ refers to the characteristic function of the set C. In contrast, the estimator based on the SRD with the

MC samples $\{\boldsymbol{v}_i\}_{i=1}^N \subseteq \mathbb{S}^{n-1}$ on the sphere is (3.7). Since the effect of the sample size N on the variance is the same for both estimators, we may restrict ourselves to comparing the variances of the elementary estimators

$$\mathrm{V_{MC}} := \operatorname{Var} \mathcal{I}_{\{\boldsymbol{z}\in\mathbb{R}^n : g(u,\boldsymbol{z})\le 0\}}(\boldsymbol{\zeta}), \qquad \mathrm{V_{SRD}} := \operatorname{Var} F_\chi(\rho(u,\boldsymbol{\theta})), \tag{3.9}$$

where $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{m}, \Sigma)$, and $\boldsymbol{\theta} \sim \mathcal{U}(\mathbb{S}^{n-1})$ is uniformly distributed on the sphere. It is well known that $\mathrm{V_{MC}} = p(1-p)$, where

$$p = \varphi(u) = \mathbb{P}(\{g(u,\boldsymbol{\zeta}) \le 0\}) = \mathbb{E}\mathcal{I}_{\{\boldsymbol{z}\in\mathbb{R}^n : g(u,\boldsymbol{z})\le 0\}}(\boldsymbol{\zeta}) = \mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta})). \tag{3.10}$$

As shown in [40, eq. (1.5)], the estimate $\mathrm{V_{SRD}} \le \mathrm{V_{MC}}$ holds. The following example shows that the variances can be equal.

Example 3.1. Consider the function $g(u,\boldsymbol{z}) := z_1$, so that the inequality $z_1 \le 0$ defines a closed half-space with the origin on its boundary. Suppose further that $\boldsymbol{\zeta} \sim \mathcal{N}(0, I)$. Then clearly, $p = 1/2$, from which follows $\mathrm{V_{MC}} = 1/4$. On the other hand, definition (3.5) of ρ implies that

$$\rho(u,\boldsymbol{v}) := \sup\{r \ge 0 : rv_1 \le 0\} = \begin{cases} 0 & \text{if } v_1 \ge 0, \\ \infty & \text{if } v_1 < 0. \end{cases}$$

Since $\boldsymbol{\theta}$ is uniformly distributed on the sphere, one concludes that, by symmetry, the events $\theta_1 \ge 0$ and $\theta_1 < 0$ occur with the same probability $1/2$. Thus, $\mathbb{P}(\rho(u,\boldsymbol{\theta}) = 0) = \mathbb{P}(\rho(u,\boldsymbol{\theta}) = \infty) = 1/2$, and therefore both events $F_\chi(\rho(u,\boldsymbol{\theta})) = 0$ and $F_\chi(\rho(u,\boldsymbol{\theta})) = 1$ happen with probability $1/2$. This means that $\mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta})) = \mathbb{E}[F_\chi(\rho(u,\boldsymbol{\theta}))]^2 = 1/2$. Therefore,

$$\mathrm{V_{SRD}} = \mathbb{E}[F_\chi(\rho(u,\boldsymbol{\theta}))]^2 - [\mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta}))]^2 = 1/4 = p(1-p) = \mathrm{V_{MC}}\,.$$

This example shows that $\mathrm{V_{SRD}} \le \mathrm{V_{MC}}$ cannot be improved in general. However, in some situations, $\mathrm{V_{SRD}}$ is substantially smaller than $\mathrm{V_{MC}}$, as shown in the following two lemmas.

Lemma 3.2. *With p from (3.10) and $\rho^{\inf/\sup} := \inf/\sup\{\rho(u,\boldsymbol{v}) \mid \boldsymbol{v} \in \mathbb{S}^{K-1}\}$, we have*

$$\mathrm{V_{SRD}} \le \min\{(1-p)(p - F_\chi(\rho^{\inf})), p(F_\chi(\rho^{\sup}) - p)\}.$$

Proof. From (3.10), it follows that

$$\mathrm{V_{SRD}} = \mathbb{E}[F_\chi(\rho(u,\boldsymbol{\theta}))]^2 - [\mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta}))]^2 \le \mathbb{E}[F_\chi(\rho^{\sup})F_\chi(\rho(u,\boldsymbol{\theta}))] - p^2 = p(F_\chi(\rho^{\sup}) - p).$$

For the second estimate, we use the identity $\operatorname{Var} X = \operatorname{Var}(1-X)$, which holds for any random variable X, and obtain that

$$\begin{aligned} \mathrm{V_{SRD}} &= \mathbb{E}[1 - F_\chi(\rho(u,\boldsymbol{\theta}))]^2 - [\mathbb{E}(1 - F_\chi(\rho(u,\boldsymbol{\theta})))]^2 \\ &\le \mathbb{E}[(1 - F_\chi(\rho^{\inf}))(1 - F_\chi(\rho(u,\boldsymbol{\theta})))] - (1-p)^2 \\ &= (1 - F_\chi(\rho^{\inf}))(1-p) - (1-p)^2 = (1-p)(p - F_\chi(\rho^{\inf})). \end{aligned}$$

■

This lemma implies that for $\rho^{\inf} \to \infty$ (growing sets with growing probability), one has that

$$\frac{V_{SRD}}{V_{MC}} \leq \frac{(1-p)(p - F_\chi(\rho^{\inf}))}{p(1-p)} = 1 - \frac{F_\chi(\rho^{\inf})}{p} \to 0. \tag{3.11}$$

Here, we used that $\rho^{\inf} \to \infty$ implies $F_\chi(\rho^{\inf}) \to 1$ and that $F_\chi(\rho^{\inf}) \leq \mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta})) = p$ necessarily implies $p \to 1$. Thus, the ratio between the variances of the SRD and the standard MC estimators becomes arbitrarily small for high probabilities p.

Similarly, for $\rho^{\sup} \to 0$ (shrinking sets with decreasing probability, yet always containing the mean $\boldsymbol{m}$ of the given Gaussian distribution), one has that

$$\frac{V_{SRD}}{V_{MC}} \leq \frac{p(F_\chi(\rho^{\sup}) - p)}{p(1-p)} = \frac{F_\chi(\rho^{\sup}) - p}{1-p} \to 0.$$

Similarly, we used that $F_\chi(\rho^{\sup}) \to 0$ for $\rho^{\sup} \to 0$ and that $F_\chi(\rho^{\sup}) \geq p$ also implies $p \to 0$. This means that the ratio between the variances of the SRD and the standard MC estimators goes to zero as $p \to 0$ (under the additional assumption that the mean of the Gaussian distribution is contained in the set, which—following the remarks below (3.4)—was automatically satisfied in the first case because $p \geq 0.5$ may be assumed due to $p \to 1$).

We may interpret $\rho_\delta := \rho^{\sup} - \rho^{\inf} \geq 0$ as a measure of radial asymmetry of the set $\{\boldsymbol{z} \in \mathbb{R}^n : g(u,\boldsymbol{z}) \leq 0\}$. Clearly, for $\rho_\delta = 0$, one has that $\rho(u,\boldsymbol{v})$ is constant for all $\boldsymbol{v} \in \mathbb{S}^{n-1}$ so that the set becomes perfectly symmetric (i.e., a ball) around zero.

Lemma 3.3. *Under the assumptions of Lemma 3.2 and also by assuming that $p \in (0,1)$, one has that $V_{SRD} \leq L^2\rho_\delta^2$, where L is the maximum of the density $\mu_\chi = F_\chi'$ of the given χ-distribution (which is bounded for any dimension).*

Proof. Thanks to $p \in (0,1)$, there exists a unique p-quantile ρ^* of the one-dimensional chi distribution such that $F_\chi(\rho^*) = p$. One easily derives $\rho^* \in [\rho^{\inf}, \rho^{\sup}]$, because otherwise a contradiction with $\mathbb{E}F_\chi(\rho(u,\boldsymbol{\theta})) = p$ would arise from the fact that the distribution function F_χ is strictly increasing. Now, exploiting this last property once more, we arrive at

$$\begin{aligned} V_{SRD} &= \mathrm{Var}(F_\chi(\rho(u,\boldsymbol{\theta})) = \int_{\boldsymbol{v}\in\mathbb{S}^{K-1}} [F_\chi(\rho(u,\boldsymbol{v})) - \mathbb{E}F_\chi(\rho(u,\boldsymbol{v}))]^2 d\mu_U(\boldsymbol{v}) \\ &= \int_{\boldsymbol{v}\in\mathbb{S}^{K-1}} [F_\chi(\rho(u,\boldsymbol{v})) - F_\chi(\rho^*)]^2 d\mu_U(\boldsymbol{v}) \leq (F_\chi(\rho^{\sup}) - F_\chi(\rho^{\inf}))^2 \\ &= (F_\chi'(\tilde\rho)(\rho^{\sup} - \rho^{\inf}))^2 \leq L^2\rho_\delta^2 \qquad (\tilde\rho \in [\rho^{\inf}, \rho^{\sup}]). \end{aligned}$$

∎

Note that Lemma 3.3 implies that the SRD estimator has zero variance for balls centered at zero. Therefore, a single ray is sufficient to compute the exact probability of a ball. This is of course well known. Now, consider a sequence of sets with the same probability p converging to a centered ball. In this sequence, $V_{MC} = p(1-p)$ is the fixed variance of the MC estimator, while V_{SRD} converges to zero. Thus, the variance ratio between the estimators reduces in favor of the SRD as shown in Lemma 3.2.

Finally, we revisit Example 3.1. For the half-space considered there, $\rho^{\inf} = 0$ and $\rho^{\sup} = \infty$. Hence, the estimate from Lemma 3.2 only yields the already known relation $V_{SRD} \leq p(1-p) =$

V_{MC}. Moreover, $\rho_\delta = \infty$, for which Lemma 3.3 does not provide additional information either. The reason is that this half-space is neither a large nor a small space (its probability is equal to 1/2, which is far from 1 and 0) nor a symmetric set. This explains why in this situation the variance of the SRD MC estimator is not reduced compared to the classical MC estimator.

3.3. Differentiability and computation of gradients of the probability function. When numerically addressing a chance constraint such as (1.3), compactly written as $\varphi(u) \geq p$ with φ defined as in (2.4), it is essential to compute not only the value of $\varphi(u)$ but also its derivatives. Analytical properties of φ, such as Lipschitz continuity or differentiability, have been studied before [21, 40, 41, 42]. Given our focus on numerical methods, we concentrate on the discretized probability function $\tilde{\varphi}$ defined in (3.7), which uses the samples $\{\boldsymbol{v}_i\}_{i=1}^N \subseteq \mathbb{S}^{n-1}$. Specifically, we aim at identifying conditions guaranteeing its differentiability and derive an explicit form of the derivative suitable for implementation. Since the chi distribution function F_χ is continuously differentiable, with its derivative being the density μ_χ, the problem reduces to verifying the partial differentiability of the function $\rho(u,\cdot)$ with respect to u. We now fix some $\bar{u} \in X$ and assume that the random constraint function g from subsection 3.1 has the typical form relevant for joint chance constraints, namely

$$g := \max_{j=1,\dots,p} g_j, \tag{3.12}$$

where the $g_j : U \times \mathbb{R}^n \to \mathbb{R}$ are affine. In particular, $g(u,\cdot)$ is convex for all $u \in U$ as required in subsection 3.1. Analogously to (3.5), we define

$$\rho_j(u,\boldsymbol{v}) := \sup\{r \geq 0 : g_j(u, \boldsymbol{m} + rL\boldsymbol{v}) \leq 0\} \qquad (j = 1,\dots,p). \tag{3.13}$$

Under the additional assumption

$$g_j(\bar{u}, \boldsymbol{m}) < 0 \qquad (j = 1,\dots,p), \tag{3.14}$$

it follows that (3.4) holds. Moreover,

$$\rho(\bar{u},\boldsymbol{v}) = \min_{j=1,\dots,p} \rho_j(\bar{u},\boldsymbol{v}). \tag{3.15}$$

In the following we denote by f_χ the density of the one-dimensional chi distribution μ_χ with n degrees of freedom and adopt the convention that $f_\chi(\infty) = \lim_{t\to\infty} f_\chi(t) = 0$.

Proposition 3.4. *Let U be a Banach space. In the setting above and under assumption (3.14), if for all $i \in \{1,\dots,N\}$ satisfying $\rho(\bar{u},\boldsymbol{v}_i) < \infty$ it holds that*

$$\#\{j \in \{1,\dots,p\} : \rho(\bar{u},\boldsymbol{v}_i) = \rho_j(\bar{u},\boldsymbol{v}_i)\} = 1, \tag{3.16}$$

then $\tilde{\varphi}$ is continuously differentiable at $\bar{u}$ with derivative

$$\nabla\tilde{\varphi}(\bar{u}) = -N^{-1}\sum_{i=1}^N \frac{f_\chi(\rho_{j_i}(\bar{u},\boldsymbol{v}_i))}{\langle \nabla_z g_{j_i}(\bar{u}, \boldsymbol{m} + \rho_{j_i}(\bar{u},\boldsymbol{v}_i)L\boldsymbol{v}_i), L\boldsymbol{v}_i\rangle} \nabla_u g_{j_i}(\bar{u}, \boldsymbol{m} + \rho_{j_i}(\bar{u},\boldsymbol{v}_i)L\boldsymbol{v}_i). \tag{3.17}$$

Here, for $i \in \{1,\dots,N\}$, the index j_i satisfies (3.16), which is unique if $\rho(\bar{u},\boldsymbol{v}_i) < \infty$.

Proof. Our setting allows us to invoke [42, Cor. 3.2] for each $j \in \{1,\dots,p\}$ to argue that the functions $F_\chi(\rho_j(u,\boldsymbol{v}))$ are continuously differentiable at all $(u,\boldsymbol{v}) \in \mathcal{V}_j(\bar{u}) \times \mathbb{S}^{n-1}$, where $\mathcal{V}_j(\bar{u})$ is some neighborhood of $\bar{u}$. Moreover, for all $\boldsymbol{v} \in \mathbb{S}^{n-1}$, one has that

$$\nabla[F_\chi(\rho_j(\cdot,\boldsymbol{v}))](\bar{u}) = \frac{-f_\chi(\rho_j(\bar{u},\boldsymbol{v}))}{\langle \nabla_z g_j(\bar{u},\boldsymbol{m}+\rho_j(\bar{u},\boldsymbol{v})L\boldsymbol{v}), L\boldsymbol{v}\rangle} \nabla_u g_j(\bar{u},\boldsymbol{m}+\rho_j(\bar{u},\boldsymbol{v})L\boldsymbol{v}). \tag{3.18}$$

Actually, the application of [42, Cor. 3.2] requires the verification of a certain growth condition. However, since g_j is affine, this condition is easily verified along the lines of [17, Lem. 6]. To prove (3.16), consider first an index $i \in \{1,\dots,N\}$ with $\rho(\bar{u},\boldsymbol{v}_i) < \infty$. It follows from [21, Lem. 1] and (3.16) that $\rho(\cdot,\boldsymbol{v}_i) = \rho_{j_i}(\cdot,\boldsymbol{v}_i)$ locally around $\bar{u}$. Therefore by (3.18),

$$\nabla[F_\chi(\rho(\cdot,\boldsymbol{v}_i))](\bar{u}) = \frac{-f_\chi(\rho_{j_i}(\bar{u},\boldsymbol{v}_i))}{\langle \nabla_z g_{j_i}(\bar{u},\boldsymbol{m}+\rho_{j_i}(\bar{u},\boldsymbol{v}_i)L\boldsymbol{v}_i), L\boldsymbol{v}_i\rangle} \nabla_u g_{j_i}(\bar{u},\boldsymbol{m}+\rho_{j_i}(\bar{u},\boldsymbol{v}_i)L\boldsymbol{v}_i), \tag{3.19}$$

which yields (3.17). Now, consider an index i where $\rho(\bar{u},\boldsymbol{v}_i) = \infty$, whence $\rho_j(\bar{u},\boldsymbol{v}_i) = \infty$ for all $j = 1,\dots,p$. Consequently, $F_\chi(\rho(\bar{u},\boldsymbol{v}_i)) = F_\chi(\rho_j(\bar{u},\boldsymbol{v}_i))$ for all $j \in \{1,\dots,p\}$. As argued above, all $F_\chi(\rho_j(\cdot,\boldsymbol{v}_i))$ are continuously differentiable at $\bar{u}$ with vanishing derivative [42, Lem. 3.3]. This shows that $F_\chi(\rho(\cdot,\boldsymbol{v}_i))$ itself is continuously differentiable with zero derivative at $\bar{u}$, and hence (3.19) also holds in this second case thanks to $f_\chi(\infty) = 0$. Finally, (3.17) follows from (3.19) and the definition of $\tilde{\varphi}$ in (3.7). ∎

The next proposition shows that the difficult-to-verify condition (3.16), required for differentiability of $\tilde{\varphi}$, is generically satisfied under a common constraint qualification that is weaker than the prominent *linear independence constraint qualification* (LICQ).

Proposition 3.5. *If the so-called rank-2-constraint qualification*

$$\begin{aligned} &\forall \boldsymbol{z} \in \mathbb{R}^n, i,j \in \{1,\dots,p\}, i \neq j: \\ &0 = g(\bar{u},\boldsymbol{z}) = g_i(\bar{u},\boldsymbol{z}) = g_j(\bar{u},\boldsymbol{z}) \Longrightarrow \ \mathrm{rank}\,\{\nabla_{\boldsymbol{z}} g_i(\bar{u},\boldsymbol{z}), \nabla_{\boldsymbol{z}} g_j(\bar{u},\boldsymbol{z})\} = 2 \end{aligned} \tag{3.20}$$

holds at $\bar{u} \in U$, then a random sample $\{\boldsymbol{v}_i\}_{i=1}^N \subseteq \mathbb{S}^{n-1}$ satisfies (3.16) with probability one.

Proof. Following [41, Lem. 4.3], (3.20) implies that

$$\mu_U(\{\boldsymbol{v}_i \in \mathbb{S}^{n-1} : \rho(\bar{u},\boldsymbol{v}) < \infty, \#\{j \in \{1,\dots,p\} : \rho(\bar{u},\boldsymbol{v}_i) = \rho_j(\bar{u},\boldsymbol{v}_i)\} = 1\}) = 1.$$

This shows that (3.16) is satisfied with probability one for each $i = 1,\dots,N$ separately, and hence it is satisfied with probability one simultaneously for all $i = 1,\dots,N$. ∎

4. Application of spherical-radial decomposition to the control problem.

4.1. Discretization of joint state constraints. To replace the continuous state constraint with a finite number of inequalities, we use a continuous linear map $E : Y \to \mathbb{R}^M$. If Y continuously embeds into the space of continuous function, then the map E can be chosen as the evaluation of the state $y(\omega)$ at the points $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_M\} \subseteq \mathcal{D}$. If the dimension of $\mathcal{D}$ is $d = 1$, $Y \subset H^1(\mathcal{D}) \hookrightarrow \mathcal{C}^0(\mathcal{D})$ and thus point evaluation is continuous. If $d = 2,3$ and the data is smooth, then additional regularity arguments for elliptic operators can be used to show that $Y \subset H^2(\mathcal{D})$, which again continuously embeds into the space of continuous functions,

and thus choosing E as point evaluation is feasible. If $d > 3$ or no additional regularity can be used, E can be chosen as local averages of the state variable $y(\omega)$ centered at the points $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$. For ease of notation, in the following we write E as point evaluation, i.e., $E(y) := (y(\boldsymbol{x}_1), \dots, y(\boldsymbol{x}_M))$. The following derivations also apply to E being a local averaging operator. Note that here we consider the points $\boldsymbol{x}_j$ as given, e.g., as a uniform grid of points covering the domain $\mathcal{D}$. It is also possible to choose the set of points $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$ adaptively using probabilistic information, thus keeping the number of points M moderate, [3]. Now, regardless of the choice of a discrete subset $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M\}$ of the domain, we end up at the accordingly modified chance constraint (1.3):

$$\mathbb{P}(\omega \in \Omega \,|\, \underline{y}(\boldsymbol{x}_j) \le y(\boldsymbol{x}_j, \omega) \le \bar{y}(\boldsymbol{x}_j) \quad \forall j = 1, \dots, M) \ge p. \tag{4.1}$$

To write the probability in this chance constraint for the concrete governing equation (1.2) in the form $\mathbb{P}(\max_{1 \le j \le M} g_j(u, \boldsymbol{z}))$, we define the functions g_j as

$$\begin{aligned} g_j(u, \boldsymbol{z}) &:= \left[A(u)^{-1} \left(f - Cu - B\xi_0 - \sum_{k=1}^{K} z_k B e_k \right) - \bar{y} \right] (\boldsymbol{x}_j), \\ g_{M+j}(u, \boldsymbol{z}) &:= \left[\underline{y} - A(u)^{-1} \left(f - Cu - B\xi_0 - \sum_{k=1}^{K} z_k B e_k \right) \right] (\boldsymbol{x}_j), \end{aligned} \tag{4.2}$$

where we assume the form (2.3) for the random process ξ, with the concrete distribution $\boldsymbol{\zeta} \sim \mathcal{N}(0, \Sigma)$. In particular, $\boldsymbol{z} \in \mathbb{R}^K$ represents the realizations $\boldsymbol{\zeta}(\omega)$ of $\boldsymbol{\zeta}$. This fits our abstract setting from section 3 with the choices $p := 2M, n := K$. Clearly, the g_j are affine in $\boldsymbol{z}$ for $j = 1, \dots, 2M$ as required in subsection 3.3.

4.2. Implicit dimension reduction in probability space. We now turn to the approximation in random space. A standard method to efficiently describe a Gaussian random variable or field is its Karhunen–Loeve (KL) expansion. For instance, for a random field $\xi \sim \mathcal{N}(\xi_0, \mathcal{C}_0)$ over the physical domain $\mathcal{D}$, such a (possibly infinite) expansion is

$$\xi(\omega) = \xi_0 + \sum_k \zeta_k^{\xi}(\omega) e_k^{\xi}, \tag{4.3}$$

where $e_k^{\xi} \in L^2(\mathcal{D})$ are eigenvectors (or functions) of $\mathcal{C}_0$, and $\zeta_k^{\xi}(\omega)$ are independent Gaussian random variables with variances λ_k^{ξ}, where $\lambda_1^{\xi} \ge \lambda_2^{\xi} \ge \cdots \ge 0$ are the eigenvalues corresponding to e_k^{ξ}. The rate at which the sequence $\{\lambda_k^{\xi}\}_k$ decays controls the approximation error when the KL expansion (4.3) is truncated.

Although a truncated version of (4.3) is an efficient (in terms of the number of terms and thus the dimension) representation for the random variable $\xi(\omega)$, it may not be efficient for the state random variable $y(\omega)$, which is subject to the joint chance constraints. To find a representation tailored to the state, note that (2.2) implies

$$y(\omega) = y_0^u - A(u)^{-1} B \xi(\omega) =: y_0^u - \sum_k \zeta_k^y(\omega) e_k^y, \tag{4.4}$$

where $y_0^u := A(u)^{-1}(f - B\xi_0 - Cu)$ is the mean of the state variable. For fixed u, e_k^y are the eigenfunctions of $A(u)^{-1}B\mathcal{C}_0B^\star A(u)^{-\star}$, with the "$-\star$" denoting the inverse adjoint operator. Moreover, in this KL-type expansion, $\zeta_k^y(\omega)$ are independent Gaussian random variables with variance λ_k^y, where $\lambda_1^y \geq \lambda_2^y \geq \cdots \geq 0$ are the eigenvalues corresponding to e_k^y.

Due to smoothing properties of the inverse elliptic operator $A(u)^{-1}$, the variances λ_k^y of the random variables in (4.4) typically decay faster than the variances λ_k^ξ for (4.3). Thus, using (4.4) instead of (4.3) allows a more efficient finite-dimensional approximation of the random variable. To illustrate this, we show in the numerical example in section 5 that $K = 20$ coefficients allow a highly accurate representation of $y(\omega)$, while an accurate representation of $\xi(\omega)$ would require a substantially larger number of expansion coefficients; see Figure 1.

Note that in (4.3), λ_k^ξ and e_k^ξ depend only on the distribution of the random variable and not on the control u. When the PDE operator A is independent of u, only the mean y_0^u in (4.4) depends on the control, whereas λ_k^y and e_k^y are independent of u. For both scenarios, eigenfunctions and eigenvalues can be computed upfront and used throughout a numerical optimization algorithm to find the optimal control. However, if $A(u)$ is a function of u, then λ_k^y and e_k^y are also functions of u and need to be recomputed whenever the control changes (such as in a numerical optimization algorithm). Thus, while (4.4) provides an efficient way to represent the random state variable, in situations where the PDE operator depends on u, utilizing (4.4) may not be computationally efficient for evaluating the probability $\varphi(u)$ (or its approximation (3.7)) and its derivative with respect to u.

4.3. Probability function and its gradient for the concrete control problem. We start by observing that condition (3.14) at some $u \in U$ is translated in our concrete control problem via (4.2) and by virtue of $\boldsymbol{m} = 0$ as

$$\underline{y}(\boldsymbol{x}_j) < y_0^u(\boldsymbol{x}_j) < \bar{y}(\boldsymbol{x}_j) \quad (j = 1, \ldots, M), \tag{4.5}$$

where $y_0^u := A(u)^{-1}(f - Cu - B\xi_0)$ refers to the state associated with the control u and the mean field ξ_0 of the random source term ξ according to (2.3). This being satisfied, we may represent the sample-based approximation of the probability function associated with our concrete control problem as (3.7) with $\rho(\cdot,\cdot)$ defined as in (3.5). In order to make formula (3.7) explicit, we have to represent ρ in terms of the original data of our control problem. Using (4.2), the functions ρ_j from (3.13) are calculated, for $j = 1, \ldots, M$ and $\boldsymbol{v} \in \mathbb{S}^{K-1}$, as

$$\rho_j(u, \boldsymbol{v}) = \begin{cases} \dfrac{y_0^u(\boldsymbol{x}_j) - \bar{y}(\boldsymbol{x}_j)}{\delta(u, \boldsymbol{v})(\boldsymbol{x}_j)} & \text{if } \delta(u, \boldsymbol{v})(\boldsymbol{x}_j) < 0, \\ \infty & \text{if } \delta(u, \boldsymbol{v})(\boldsymbol{x}_j) \geq 0, \end{cases} \tag{4.6}$$

$$\rho_{M+j}(u, \boldsymbol{v}) = \begin{cases} \dfrac{y_0^u(\boldsymbol{x}_j) - \underline{y}(\boldsymbol{x}_j)}{\delta(u, \boldsymbol{v})(\boldsymbol{x}_j)} & \text{if } \delta(u, \boldsymbol{v})(\boldsymbol{x}_j) > 0, \\ \infty & \text{if } \delta(u, \boldsymbol{v})(\boldsymbol{x}_j) \leq 0. \end{cases} \tag{4.7}$$

Here, $\delta(u, \boldsymbol{v}) := A(u)^{-1}B\sum_{k=1}^K (L\boldsymbol{v})_k e_k$. Now, the approximating probability function $\tilde{\varphi}(u)$ in (3.7) is immediately calculated from the above formulas using the representation (3.15). Regarding the derivative of $\tilde{\varphi}$, we can calculate the partial derivatives of g_j with respect to u and $\boldsymbol{z}$ as in (4.2). Doing so, we obtain, for all $h \in U$, $j = 1, \ldots, p$, and $k = 1, \ldots, K$ as follows:

$$\begin{aligned}\langle \nabla_u g_j(u,\boldsymbol{z}),h\rangle &= -\langle \nabla_u g_{M+j}(u,\boldsymbol{z}),h\rangle \\ &= \left[\left(\left(A(u)^{-1}\right)' h\right)\left(f - Cu - B\xi_0 - \sum_{k=1}^{K} z_k Be_k\right) - A(u)^{-1}Ch\right](\boldsymbol{x}_j), \end{aligned} \tag{4.8}$$

$$\frac{\partial g_j}{\partial z_k}(u,\boldsymbol{z}) = -\frac{\partial g_{M+j}}{\partial z_k}(u,\boldsymbol{z}) = -\left[A(u)^{-1}Be_k\right](\boldsymbol{x}_j). \tag{4.9}$$

Now, under assumption (3.16), we obtain that $\tilde{\varphi}$ is continuously differentiable at $\bar{u}$ with the derivative in (3.17), which can be made explicit using (4.8), (4.9).

Finally, we make assumption (3.16) more explicit in light of the rank-2 constraint qualification in Proposition 3.5. For that purpose, for a fixed control $\bar{u}$, we denote by $y^{(k)}$ the solutions of

$$A(\bar{u})y + Be_k = 0 \quad (k = 1,\dots,K).$$

Proposition 4.1. *If for any $i,j \in \{1,\dots,M\}$ with $i \neq j$, the two vectors*

$$\begin{pmatrix} y^{(1)}(\boldsymbol{x}_i) \\ \vdots \\ y^{(K)}(\boldsymbol{x}_i) \end{pmatrix}, \begin{pmatrix} y^{(1)}(\boldsymbol{x}_j) \\ \vdots \\ y^{(K)}(\boldsymbol{x}_j) \end{pmatrix}$$

are linearly independent, then the rank-2-constraint qualification from Proposition 3.5 is satisfied at $\bar{u}$, and, consequently, for a random sample $\{\boldsymbol{v}_i\}_{i=1}^N \subseteq \mathbb{S}^{n-1}$, the probability function $\tilde{\varphi}$ from (3.7) is continuously differentiable at $\bar{u}$ with probability one.

Proof. First, note that due to (4.5), the constraint functions (4.2) cannot satisfy the condition $0 = g_i(\bar{u},\boldsymbol{z}) = g_{M+i}(\bar{u},\boldsymbol{z})$ for some $\boldsymbol{z} \in \mathbb{R}^n$ and some $i \in \{1,\dots,M\}$. Therefore, the set $\{i,j\} \subseteq \{1,\dots,p = 2M\}$ with $i \neq j$ in (3.20) can always be written in one of the following forms:

$$\{i,j\} = \{i',j'\} \quad \text{or} \quad \{i,j\} = \{i', M+j'\} \quad \text{or} \quad \{i,j\} = \{M+i', M+j'\}$$

for some $i',j' \in \{1,\dots,M\}$ with $i' \neq j'$. In any case, due to (4.9) and the definition of $y^{(k)}$, the gradient $\nabla_{\boldsymbol{z}} g_i(\bar{u},\boldsymbol{z})$ coincides, up to the sign, with the first vector above. Likewise, $\nabla_{\boldsymbol{z}} g_j(\bar{u},\boldsymbol{z})$ coincides (up to the sign) with the second vector above. Therefore, our assumption on the linear independence of these two vectors implies the conclusion of (3.20). ∎

5. Numerical results for linear governing equation. We now use the linear-quadratic problem from Example 1.1 to study the approximation of the joint chance constraint and the behavior of optimal controls. First, we show that the integration over Ω in the objective (1.4) can be done analytically for a Gaussian random field. For that purpose, assume that the law of the Gaussian measure is $\mathcal{N}(\xi_0,\mathcal{C}_0)$. Using the linearity of the governing equation implies that, for fixed control u, the state y is also Gaussian. Thus, integration over Ω of the quadratic in (1.4) can be performed analytically [11, Remark 1.2.9] and results in the reduced problem

$$\begin{aligned} &\underset{u\in L^2(\mathcal{D})}{\text{minimize}} && \tilde{\mathcal{J}}(u) := \frac{1}{2}\int_{\mathcal{D}} (y_0^u - y_d)^2\,d\boldsymbol{x} + \frac{\alpha}{2}\int_{\mathcal{D}} u^2\,d\boldsymbol{x} + \operatorname{Tr}(A^{-1}B\mathcal{C}_0B^\star A^{-\star}) \\ &\text{subject to} && \text{the joint chance constraint (1.3)}, \end{aligned} \tag{5.1}$$

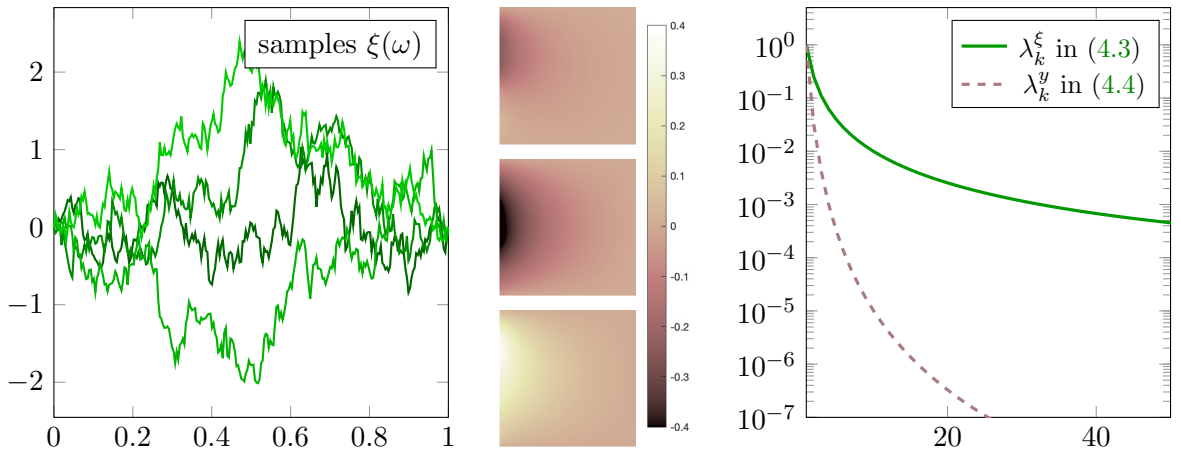

Figure 1. *The left figure shows samples from the distribution of the random boundary data $\xi(\omega)$. The middle figure shows samples from the distribution of $y(\omega) - y_0^u$, the state with zero mean. The right figure shows the normalized (i.e., λ_1 is scaled to 1) eigenvalue factors in the KL expansion (4.3) of ξ (green solid line), and in the KL expansion (4.3) of y (pink dashed line).*

where $A = -\Delta$ and $B : L^2(\partial\mathcal{D}_2) \mapsto H^{-1}(\mathcal{D})$, and $y_0^u = A^{-1}(f + u + B\xi_0)$ is the mean of the distribution of states. Moreover, "Tr" denotes the trace of an operator. Note that the last term in (5.1) is independent of u and thus can be neglected in optimization. Next, we describe the concrete domain and parameters for the problem (1.3)–(1.5).

We choose the domain $\mathcal{D} = (0,1)^2 \subset \mathbb{R}^2$ and divide the boundary into $\partial D_2 = \{0\} \times [0,1]$ and $\partial D_1 = \partial D \setminus \partial D_2$. The Laplace operator uses zero Dirichlet boundary conditions on ∂D_1 and Neumann boundary conditions on ∂D_2. In addition, $y_d = \frac{1}{10}\cos(2\pi x_1)\sin(2\pi x_2)$, $f \equiv 0$, $\alpha = 10^{-5}$. The uncertain parameter field enters as Neumann data on the one-dimensional domain ∂D_2. This data follows an infinite-dimensional Gaussian distribution with mean $\xi_0 \equiv 0$, and a covariance operator given by the inverse elliptic PDE operator $\mathcal{C}_0 = \gamma(-\partial_{x_2x_2})^{-1}$, with homogeneous Dirichlet conditions at the boundary of ∂D_2, i.e., at the two points $(0,0)$ and $(0,1)$, and with $\gamma = 4$. This covariance operator has the eigenfunctions $\sin(k\boldsymbol{x}_2)$, $k = 1,2,\ldots$, with corresponding eigenvalues $4(k\pi)^{-2}$. Samples from this distribution, and from the distribution of the two-dimensional state variable $y(\omega)$, are shown in Figure 1. This figure also shows a comparison of the KL expansion eigenvalues of ξ and y; see (4.3) and (4.4), respectively. As can be seen, the KL factors λ_i^y of $y(\omega)$ decay much faster despite corresponding to a Gaussian random field defined over a two-dimensional domain (note that ξ is defined over a one-dimensional domain). This faster decay shows that the random variable $y(\omega)$ can be well approximated in a lower dimension K.

5.1. Computational aspects. A finite difference approximation (i.e., the five-point stencil) is used on a mesh of $n \times n$ points to discretize the Laplacian Δ in the governing equation on the spatial domain $\mathcal{D}$. Finite differences are also used to discretize the one-dimensional second derivative operator in the definition of the covariance $\mathcal{C}_0$. Solving (5.1) requires repeated solution of the governing equation and thus inversion of the discretization of A. Thus, we compute the Cholesky factorization of the system matrix upfront and reuse these factors

throughout the solution process. Unless otherwise specified, we use $n = 128$; i.e., the discretized state has a dimension of $n^2 = 16,384$. We discretize the state constraint on the same grid as the governing equation, i.e., $M = n^2$. Due to the rapid decay of the Karhunen–Loeve factors for the expansion of $y(\omega)$ shown in Figure 1, we use $K = 20$ in the expansion unless otherwise specified. To obtain samples of the uniform distribution on the sphere, we first generate MC or QMC (quasi-Monte Carlo) samples from the standard Gaussian distribution in $\mathbb{R}^K$, where we use the Halton sequence for QMC. These samples are then normalized to unit length, a well-known technique for obtaining samples from the uniform distribution on the sphere. In the context of QMC, this approach can be further improved by using more efficient low-discrepancy sequences specifically designed for the sphere; see e.g., [1].

5.2. Approximation of chance constraint probability function. First, we compare different approximations of the probability (1.3), where all tests are performed at the nominal control $u = \frac{1}{5}\sin(2\pi x_1)\cos(\pi x_2)$. We compare different sampling methods for an increasing number of samples, namely standard MC sampling in $\mathbb{R}^K$, as well as spherical-radial MC and QMCsampling on $\mathbb{S}^{K-1}$.

The results for the root mean square error for two different sets of upper and lower bounds for the state are shown in Figure 2. Here, we used 10^8 samples from the standard MC method to compute a highly accurate estimate of the exact probability. We observe the expected $1/\sqrt{N}$ convergence of the standard and the SRD-based MC methods. Moreover, we find that SRD-MC requires as little as one-quarter of the samples that the standard MC method needs for the same accuracy. This difference is more pronounced in the right figure, which is for probability (1.3) close to 1. This is a first numerical confirmation of Lemma 3.2, that is, the

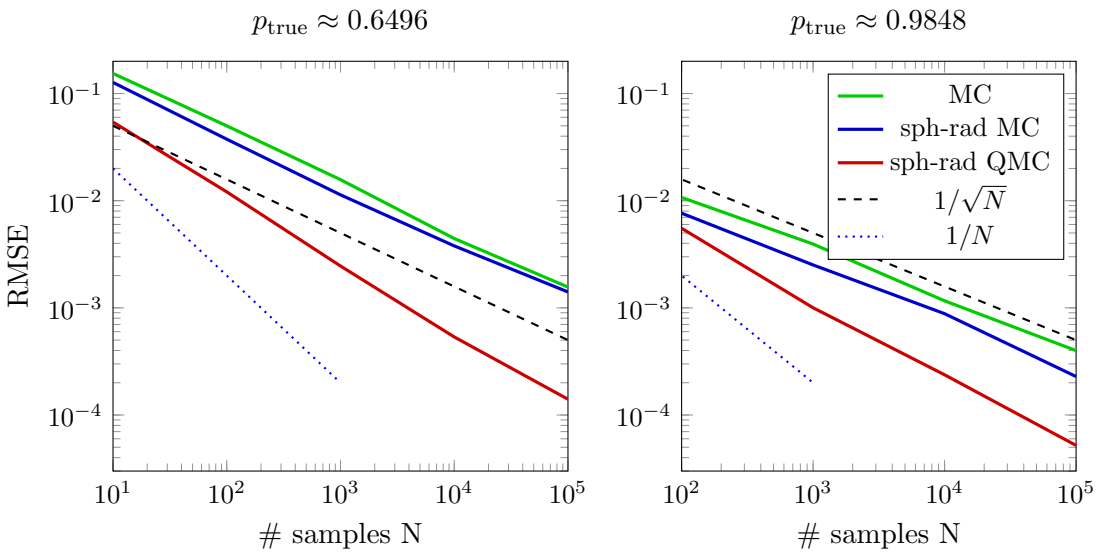

Figure 2. *Root mean squared error (RMSE) of probability estimation is shown on the y-axis for different sampling methods and different numbers N of MC samples (x-axis). The left plot is for* $\underline{y} \equiv -0.3$, $\bar{y} \equiv 0.3$, *corresponding to the probability* $p \approx 0.6496$. *The right plot uses the wider bounds* $\underline{y} \equiv -0.7$, $\bar{y} \equiv 0.7$, *resulting in* $p \approx 0.9848$.

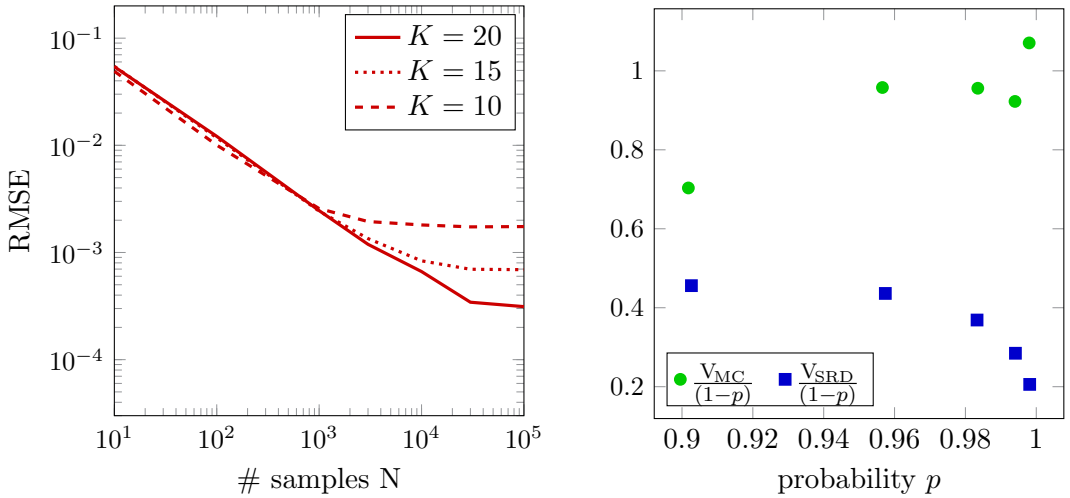

Figure 3. *Left: QMC sampling with SRD for different numbers K of KL modes. Shown is the RMSE of the probability estimate for the same setup used on the left of Figure 2, i.e., with $\underline{y} \equiv -0.3$, $\bar{y} \equiv 0.3$. Right: Variance normalized by $(1-p)$ using standard and SRD MC samples for the problem described in section 5. The difference in p is due to using different bounds $\bar{y} = -\underline{y} \equiv 0.5, 0.6, 0.7, 0.8, 0.9$ (from left to right, going from more to less restrictive). The variance is estimated using 100 MC simulations, each using $N = 500$ samples.*

result that SRD-MC outperforms standard MC as $p \to 1$. In both figures, spherical-radial QMC significantly outperforms MC methods, yielding a convergence rate in between $1/\sqrt{N}$ and $1/N$. For example, the QMC-based estimate with 2000 samples achieves an error similar to the other methods with 10^5 samples (left figure). The right figure (i.e., case p close to 1) shows that the spherical-radial QMC sampling obtains an error similar to that of SRD MC with 40,000 samples and of standard MC with 100,000 samples.

Furthermore, on the left in Figure 3, we study the influence of the number of KL modes (and therefore the dimension) used for the state variable y defined in (4.4). To compute a reference solution, we use $K = 30$ KL modes and 10^7 QMC samples. As can be seen, down to an RMSE of about $2 \cdot 10^{-3}$ (obtained with 1000 samples), there is little difference between $K = 10, 15, 20$. The KL truncation error starts to dominate the overall error at an RMSE of $2 \cdot 10^{-3}$ for 10 KL modes and an RMSE of $5 \cdot 10^{-4}$ for 20 KL modes.

Finally, to verify Lemma 3.2 numerically, we compute the variance of the standard MC estimator (3.8) and the SRD MC estimator (3.7). The results are shown on the right in Figure 3, where we divide the numerical approximation for the variances $\mathrm{V}_{\mathrm{SRD}}$ and V_{MC} defined in (3.9) by $(1-p)$ for visualization purposes. The figure shows the reduced variance of the SRD estimator; i.e., the MC estimator based on the SRD substantially improves over the standard estimator for large p. In particular, the ratio (3.11) decreases as $p \to 1$.

5.3. Optimal controls under chance constraints. To solve the optimal control problem, we employ a sequential quadratic programming (SQP) method, in which we approximate the second derivative of the chance constraint using the BFGS method [31]. For ease of visualization, we use the previously discussed problem but only with an upper-state constraint of $\bar{y} \equiv 0.3$. In Figure 4 we show the optimal control for $p = 0.9$ and samples from the state

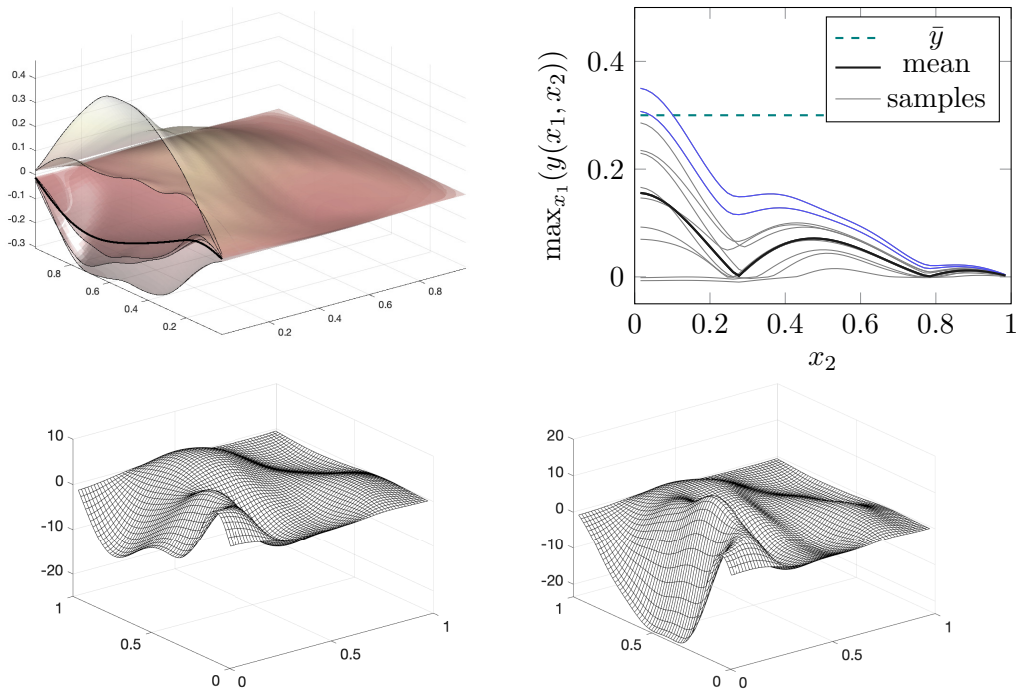

Figure 4. *Optimal control solution for $p = 0.9$ and unilateral chance constraints with $\bar{y} = 0.3$. Shown on the top left are samples from the state variable y. The graph on the top right visualizes $\max_{x_1}(y(x_{1,2}))$ as a function of x_2 for different samples of the state. The mean of the optimal state is shown in black; samples are in gray and blue. The upper bound $\bar{y}$ is shown in green (dotted line). The two blue samples exceed the bound and thus do not satisfy the bound constraint. Shown at the bottom left is the corresponding optimal control u. For comparison, the optimal control u for $p = 0.98$ is shown on the bottom right.*

variable. Furthermore, we show the optimal control for $p = 0.98$. We observe that the optimal control is quite sensitive to p; for p closer to 1, it takes much smaller values to ensure that the states are sufficiently likely to satisfy the bound constraint at all points.

6. Numerical results for the bilinear control problem.

We now extend the results from the previous section to Example 1.2. Due to the bilinear structure of the equation, for a fixed control u, the map from the uncertain parameter ξ to the state variable y is linear. Although many methods from the previous section generalize to this case, the dependence of the PDE operator on the control makes the computation of derivatives of the probability function more involved and does not allow for the straightforward dimension reduction for the uncertain variable from subsection 4.2.

6.1. Setup and discretization.

In Example 1.2, we let $\xi \sim \mathcal{N}(\xi_0, \mathcal{C}_0)$, where the covariance operator $\mathcal{C}_0 : L^2(D) \to L^2(D)$ is given by the inverse elliptic operator $(-\alpha\Delta + I)^{-2}$; see [7]. We only consider an upper bound for the state variable, i.e., formally set $\underline{y} = -\infty$ in (1.3). We use finite elements to approximate the governing equation as well as the prior covariance operator. In particular, we use an n-dimensional finite element subspace $V_h \subset H_0^1(\mathcal{D})$ and

write the governing equation (1.6) in the form $[A + M(u)]y(\omega) = M(f + \xi(\omega))$, where, for $v, g \in V_h$, we define the operators $A, M(u)$ and M through the weak forms

$$\langle Ay, v\rangle := \int_{\mathcal{D}} \nabla y \cdot \nabla v \, dx, \quad \langle M(u)y, v\rangle := \int_{\mathcal{D}} uyv \, dx, \quad \langle Mg, v\rangle := \int_{\mathcal{D}} gv \, dx.$$

In the following, we associate each operator with its corresponding matrix, i.e., $A_{ij} = \langle A\phi_i, \phi_j\rangle$, where $\{\phi_1, \ldots, \phi_n\}$ is a basis for the space V_h. As is usual in the finite element method, we denote by $\boldsymbol{u} \in \mathbb{R}^n$ the coefficient vector corresponding to the finite element function u_h (which we simply denote by u again) and use similar notation for all the other variables. Expressed in terms of the above operators, the discretized prior covariance is $\mathcal{C}_0 = ([\alpha A + M]^{-1}M)^2$, and samples of $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\xi}_0, \mathcal{C}_0)$ can then be generated (see [7, sect. 3.6]) via $\boldsymbol{\xi} = \boldsymbol{\xi}_0 + [\alpha A + M]^{-1}L\boldsymbol{z}$, where $\boldsymbol{z} \in \mathbb{R}^n$ is a draw from an independent and indentically distributed (i.i.d.) standard normal distribution, and L satisfies $LL^T = M$. For brevity and to emphasize that the following analysis does not depend on the choice of the covariance operator, we denote $\tilde{L} := [\alpha A + M]^{-1}L$. Next, we sketch the computation of the probability function and its gradient for this bilinear control problem.

6.2. Chance constraint probability function and its gradient. As before, for fixed control u we use the SRD to generate samples of the (finite element coefficient) of the state variable as follows:

$$\boldsymbol{y}_i = [A + M(u)]^{-1}M\boldsymbol{\xi}_0 + r_i[A + M(u)]^{-1}M(\boldsymbol{f} + \tilde{L}\boldsymbol{v}_i) \qquad (i = 1, \ldots, N), \tag{6.1}$$

where r_i are random draws from the chi distribution, $\boldsymbol{v}_i$ draws from the uniform distribution of the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, and $\boldsymbol{f}$ are the coefficients of the discretization of f. We denote $\boldsymbol{y}_0^u := [A + M(u)]^{-1}M(\boldsymbol{f} + \boldsymbol{\xi}_0)$ and obtain the approximating probability function

$$\tilde{\varphi}(u) := N^{-1}\sum_{i=1}^{N} F_\chi(\rho(u, \boldsymbol{v}_i)) \quad \text{with } \rho(u, \boldsymbol{v}_i) \text{ defined as in (4.6).} \tag{6.2}$$

Note that the partial derivatives (4.8) and (4.9), which are needed in the expressions for $\nabla\tilde{\varphi}(u)$ given by (3.17), involve the term $(A(u)^{-1})'h$. To obtain an explicit form of this derivative, we use the adjoint method to compute the gradient of $\tilde{\varphi}(u)$, i.e., on the right-hand side in the definition of $\tilde{\varphi}(u)$, we consider the variables $u, \boldsymbol{y}_i$, and $\boldsymbol{y}_0^u$ to be independent of one another, but constrain we the objective by the equation for the mean state and for the state samples (6.1). Following the (formal) Lagrange method, we compute $\nabla\tilde{\varphi}(u)$ by taking partial variations of the corresponding Lagrangian function [39]. We find that in addition to solving the equation for the state mean and the N equations in (6.1), the gradient computation requires another $(N + 1)$ solutions of adjoint equations (which coincide with the state equation as the PDE operator is self-adjoint). Using this gradient, one can now solve optimal control problems with chance constraints, as illustrated next for concrete data.

6.3. Optimal controls under chance constraints. We consider a problem of the form of Example 1.2, without the control bound $u \geq 0$. Using the domain $\mathcal{D} = [0, 1]^2$, we define the auxiliary variable $y_d := \sin(2\pi x_1)\sin(2\pi x_2)$ and choose $f := -\Delta y_d + y_d$, $u_0 := 1$, $\xi_0 := 0$.

This data is constructed in such a way that the optimal control is known in the absence of uncertainty. Namely, when the random variable $\xi(\omega)$ reduces to its mean $\xi_0 = 0$, the optimal control is $u \equiv 1$, and the optimal state is $y = y_d$. For the (inverse) PDE operator that defines the covariance $\mathcal{C}_0$, we use $\alpha = 0.1$ and homogeneous boundary conditions.

We choose the upper bound $\bar{y} = 1.10$ for the states and first solve the optimization problem without the joint chance constraint. For the resulting optimal control, the probability of exceeding $\bar{y}$ is 19.6%. We then add the chance constraint and solve the optimization problem for $p = 0.82, 0.84, 0.86$, i.e., enforce an exceedance probability of less than $(1 - p)$. The resulting optimal controls u are shown in Figure 5, and the objective values corresponding to these controls are displayed on the left in Figure 6. As can be seen, the objective increases rapidly as p increases. Also shown on the right in Figure 6 is the mean of the optimal state corresponding to $p = 0.84$. Note that the optimal controls u are increased, compared to the solution $u \equiv 1$ without chance constraints, at the two points $(1/4, 1/4)$ and $(3/4, 3/4)$. This is due to the damping effect that the control has on the states, which is necessary when the mean of the state variable is close to the upper bound $\bar{y}$. The increasingly singular behavior of controls for increasing p is a consequence of the unboundedness of Gaussian distributions, and the analysis of its behavior and its accurate approximation are interesting future research questions.

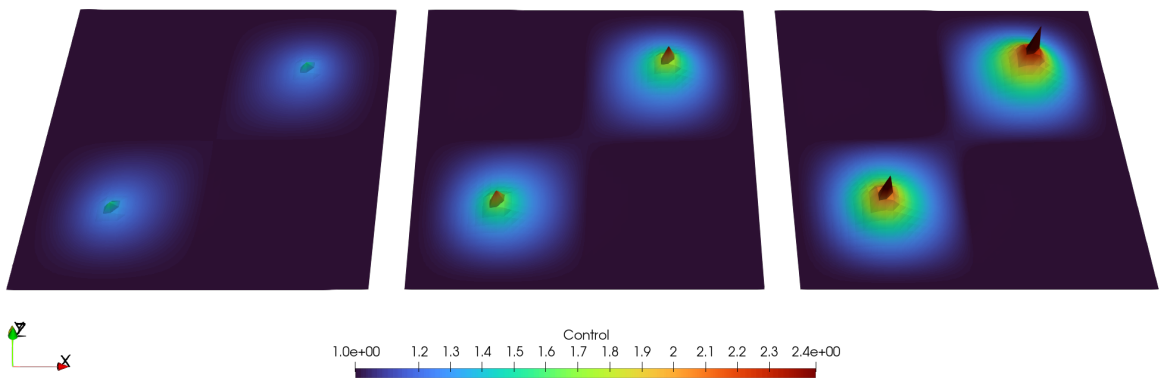

Figure 5. *Optimal controls u for $p = 0.82, 0.84, 0.86$ (from left to right) for the bilinear example. Note that the optimal controls develop spikes near the points where the state distribution is close to the upper bound.*

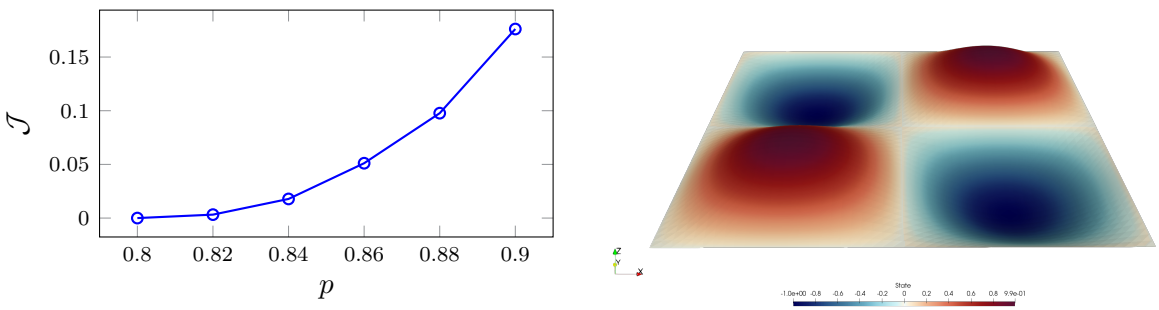

Figure 6. *Left: Value of objective $\mathcal{J}$ at the optimal control u as p in the chance constraint is increased. Right: Mean of state corresponding to optimal control for $p = 0.84$ shown in the middle image of Figure 5.*

Appendix. The subsequent two lemmas are slight extensions of the results from [35, Thm. 10.2.1] and [14, Lem. 2], adapted for infinite-dimensional image spaces. Since a direct derivation from these results is not possible, we provide independent proofs for the reader's convenience.

Lemma A.1. *Let $U, \tilde{Y}$ be normed spaces, $K \subseteq \tilde{Y}$ be a convex subset, and $g : U \times \mathbb{R}^n \to \tilde{Y}$ be a linear mapping. Suppose further that $\boldsymbol{\zeta}$ is an n-dimensional random vector on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ whose distribution has a log-concave density. Then, the set*

$$M := \{u \in U : \mathbb{P}(g(u, \boldsymbol{\zeta}) \in K) \geq p\}$$

is convex for arbitrary $p \in [0, 1]$.

Proof. Let $p \in [0, 1]$ be arbitrary. Define the set-valued mapping $H : U \rightrightarrows \mathbb{R}^n$ by

$$H(u) := \{\boldsymbol{z} \in \mathbb{R}^n : g(u, \boldsymbol{z}) \in K\} \qquad (u \in U).$$

Note that the images $H(u)$ are convex subsets of $\mathbb{R}^n$ for all $u \in U$ thanks to K being convex and g being linear. Although convex sets in $\mathbb{R}^n$ need not be Borel measurable, they are Lebesgue measurable (see [28]), and hence, with the random vector $\boldsymbol{\zeta}$ having a density by assumption, the probability $\mathbb{P}(\boldsymbol{\zeta} \in H(u))$ is well defined for all $u \in U$. Now, let $u^1, u^2 \in M$ and $\lambda \in [0, 1]$ be arbitrarily given. This means that $\mathbb{P}(\boldsymbol{\zeta} \in H(u^1)) \geq p$ and $\mathbb{P}(\boldsymbol{\zeta} \in H(u^2)) \geq p$. We have to show that $\lambda u^1 + (1 - \lambda)u^2 \in M$. We claim that

$$H(\lambda u^1 + (1 - \lambda)u^2) \supseteq \lambda H(u^1) + (1 - \lambda)H(u^2), \tag{A.1}$$

where the right-hand side is the set of elementwise sums. In fact, assuming $\boldsymbol{z} \in \lambda H(u^1) + (1 - \lambda)H(u^2)$, there must exist $\boldsymbol{z}^1 \in H(u^1)$ and $\boldsymbol{z}^2 \in H(u^2)$ such that $\boldsymbol{z} = \lambda \boldsymbol{z}^1 + (1 - \lambda)\boldsymbol{z}^2$. Accordingly, $g(u^1, \boldsymbol{z}^1), g(u^2, \boldsymbol{z}^2) \in K$ and, hence, by linearity of g and convexity of K,

$$g(\lambda u^1 + (1 - \lambda)(u^2, \boldsymbol{z})) = g(\lambda(u^1, \boldsymbol{z}^1) + (1 - \lambda)(u^2, \boldsymbol{z}^2)) = \lambda g(u^1, \boldsymbol{z}^1) + (1 - \lambda)g(u^2, \boldsymbol{z}^2) \in K.$$

This proves that $\boldsymbol{z} \in H(\lambda u^1 + (1 - \lambda)u^2)$. By (A.1), we continue as

$$\begin{aligned} \mathbb{P}(\boldsymbol{\zeta} \in H(\lambda u^1 + (1 - \lambda)u^2)) &\geq \mathbb{P}(\boldsymbol{\zeta} \in \lambda H(u^1) + (1 - \lambda)H(u^2)) \\ &\geq [\mathbb{P}(\boldsymbol{\zeta} \in H(u^1))]^{\lambda} \cdot [\mathbb{P}(\boldsymbol{\zeta} \in H(u^2))]^{1-\lambda} \\ &\geq p^{\lambda} p^{1-\lambda} = p. \end{aligned}$$

Here, the second inequality follows from Prékopas' theorem stating that a log-concave density (assumed here) induces a log-concave probability measure [35, Thm. 4.2.1]. Thus, we have shown the desired relation $\lambda u^1 + (1 - \lambda)u^2 \in M$. ∎

Lemma A.2. *Let $U, \tilde{Y}$ be Banach spaces, $K \subseteq \tilde{Y}$ a weakly closed subset, and $g : U \times \mathbb{R}^n \to \tilde{Y}$ a weakly sequentially continuous mapping. Suppose further that $\boldsymbol{\zeta}$ is an n-dimensional random vector on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then, the probability function $\tilde{\varphi} : U \to [0, 1]$ defined by*

$$\tilde{\varphi}(u) := \mathbb{P}(g(u, \boldsymbol{\zeta}) \in K) \qquad (u \in U)$$

is weakly sequentially upper semicontinuous.

Proof. Observe first that $\{\omega \in \Omega \mid g(u, \boldsymbol{\zeta}(\omega)) \in K\} \in \mathcal{A}$ for arbitrary $u \in U$ because the sets $\{\boldsymbol{z} \in \mathbb{R}^n \mid g(u, \boldsymbol{z}) \in K\}$ are closed by weak sequential continuity of g and weak closedness of K. Fix an arbitrary $\bar{u}$, and let $u_n \rightharpoonup \bar{u}$ be a weakly convergent sequence. Denote by u_{n_l} a subsequence such that

$$\limsup_{n\to\infty} \tilde{\varphi}(u_n) = \lim_{l\to\infty} \tilde{\varphi}(u_{n_l}). \tag{A.2}$$

Define the sets $A, A_n \in \mathcal{A}$ by

$$A := \{\omega \in \Omega : g(\bar{u}, \boldsymbol{\zeta}(\omega)) \in K\}; \quad A_n := \{\omega \in \Omega : g(u_n, \boldsymbol{\zeta}(\omega)) \in K\} \quad (n \in \mathbb{N}).$$

Now, consider an arbitrary $\omega \in \Omega \setminus A$. Then, $g(\bar{u}, \boldsymbol{\zeta}(\omega)) \notin K$. Since g is weakly sequentially continuous, we have $g(u_n, \boldsymbol{\zeta}(\omega)) \rightharpoonup g(\bar{u}, \boldsymbol{\zeta}(\omega))$. Since $Y \setminus K$ is weakly open, it follows that

$$\forall \omega \in \Omega \setminus A \quad \exists n_0(\omega) : g(u_n, \boldsymbol{\zeta}(\omega)) \notin K \quad \forall n \geq n_0(\omega).$$

Denoting by χ_C the characteristic function of a set C, it follows that $\chi_{A_n}(\omega) \to 0$ as $n \to \infty$ for all $\omega \in \Omega \setminus A$. By the dominated convergence theorem,

$$\int_{\Omega\setminus A} \chi_{A_n}(\omega)\, \mathbb{P}(d\omega) \to 0 \quad \forall \omega \in \Omega \setminus A \text{ as } n \to \infty.$$

On the other hand, $\chi_{A_n}(\omega) \leq \chi_A(\omega) = 1$ for $\omega \in A$, whence

$$\begin{aligned}
\lim_{l\to\infty} \tilde{\varphi}(u_{n_l}) &= \lim_{l\to\infty} \mathbb{P}(g(u_{n_l}, \boldsymbol{\zeta}) \in K) = \lim_{l\to\infty} \int_\Omega \chi_{A_{n_l}}(\omega)\, \mathbb{P}(d\omega) \\
&\leq \limsup_{l\to\infty} \int_{\Omega\setminus A} \chi_{A_{n_l}}(\omega)\, \mathbb{P}(d\omega) + \limsup_{l\to\infty} \int_A \chi_{A_{n_l}}(\omega)\, \mathbb{P}(d\omega) \\
&= \limsup_{l\to\infty} \int_A \chi_{A_{n_l}}(\omega)\, \mathbb{P}(d\omega) \\
&\leq \limsup_{l\to\infty} \int_A \mathbb{P}(d\omega) = \mathbb{P}(A) = \mathbb{P}(g(\bar{u}, \boldsymbol{\zeta}) \in K) = \tilde{\varphi}(\bar{u}).
\end{aligned}$$

Combining this with (6.2) results in φ being weakly sequentially upper semicontinuous in $\bar{u}$. ∎

Acknowledgments. The authors thank two anonymous referees for their valuable comments and suggestions that improved this article.

REFERENCES

[1] C. Aistleitner, J. S. Brauchart, and J. Dick, *Point sets on the sphere $\mathbb{S}^2$ with small spherical cap discrepancy*, Discrete Comput. Geom., 48 (2012), pp. 990–1024, https://doi.org/10.1007/s00454-012-9451-3.

[2] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas, *Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 1166–1192, https://doi.org/10.1137/16M106306X.

[3] H. Berthold, H. Heitsch, R. Henrion, and J. Schwientek, *On the algorithmic solution of optimization problems subject to probabilistic/robust (probust) constraints*, Math. Methods Oper. Res., 96 (2022), pp. 1–37, https://doi.org/10.1007/s00186-021-00764-8.

[4] A. Borzì, G. Ciaramella, and M. Sprengel, *Formulation and Numerical Solution of Quantum Control Problems*, Comput. Sci. Engrg. 16, SIAM, Philadelphia, 2017, https://doi.org/10.1137/1.9781611974843.

[5] A. Borzì, J. Salomon, and S. Volkwein, *Formulation and numerical solution of finite-level quantum optimal control problems*, J. Comput. Appl. Math., 216 (2008), pp. 170–197, https://doi.org/10.1016/j.cam.2007.04.029.

[6] A. Borzì and V. Schulz, *Computational Optimization of Systems Governed by Partial Differential Equations*, Comput. Sci. Engrg. 8, SIAM, Philadelphia, 2012, https://doi.org/10.1137/1.9781611972054.

[7] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems Part* I*: The Linearized Case, with Application to Global Seismic Inversion*, SIAM J. Sci. Comput., 35 (2013), pp. A2494–A2523, https://doi.org/10.1137/12089586X.

[8] 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 (2018), pp. 1833–1858, https://doi.org/10.1002/oca.2445.

[9] 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 (1958), pp. 235–263, https://doi.org/10.1287/mnsc.4.3.235.

[10] P. Chen and O. Ghattas, *Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters*, SIAM/ASA J. Uncertain. Quantif., 9 (2021), pp. 1381–1410, https://doi.org/10.1137/20M1381381.

[11] G. Da Prato and J. Zabczyk, *Second-Order Partial Differential Equations in Hilbert Spaces*, Cambridge University Press, Cambridge, UK, 2002.

[12] K. Fang, S. Kotz, and K. W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London, 1990.

[13] 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. Control Optim., 58 (2020), pp. 2288–2311, https://doi.org/10.1137/19M1269944.

[14] 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 (2018), pp. 821–841, https://doi.org/10.1007/s11228-017-0452-5.

[15] D. B. Gahururu, M. Hintermüller, and T. M. Surowiec, *Risk-neutral PDE-constrained generalized Nash equilibrium problems*, Math. Program., 198 (2023), pp. 1287–1337, https://doi.org/10.1007/s10107-022-01800-z.

[16] S. Garreis, T. M. Surowiec, and M. Ulbrich, *An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures*, SIAM J. Optim., 31 (2021), pp. 1–29, https://doi.org/10.1137/19M125039X.

[17] C. Geiersbach and R. Henrion, *Optimality conditions in control problems with random state constraints in probabilistic or almost sure form*, Math. Oper. Res., published online July 15, 2024, https://doi.org/10.1287/moor.2023.0177.

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

[19] A. Geletu, A. Hoffmann, M. Klöppel, and P. Li, *An inner-outer approximation approach to chance constrained optimization*, SIAM J. Optim., 27 (2017), pp. 1834–1857, https://doi.org/10.1137/15M1049750.

[20] A. Geletu, A. Hoffmann, P. Schmidt, and P. Li, *Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation*, ESAIM COCV, 26 (2020), 70, https://doi.org/10.1051/cocv/2019077.

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

[22] H. Heitsch, R. Henrion, and C. Tischendorf, *Probabilistic maximization of time-dependent capacities in a gas network*, Optim. Eng., 26 (2025), pp. 365–400, https://doi.org/10.1007/s11081-024-09908-1.

[23] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization with PDE Constraints*, Math. Model. Theory Appl. 23, Springer, New York, 2009.
[24] K. Ito and K. Kunisch, *Optimal bilinear control of an abstract Schrödinger equation*, SIAM J. Control Optim., 46 (2007), pp. 274–287, https://doi.org/10.1137/05064254X.
[25] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, *A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty*, SIAM J. Sci. Comput., 35 (2013), pp. A1847–A1879, https://doi.org/10.1137/120892362.
[26] D. P. Kouri and A. Shapiro, *Optimization of PDEs with uncertain inputs*, in Frontiers in PDE-Constrained Optimization, IMA Vol. Math. Appl. 163, Springer, New York, 2018, pp. 41–81.
[27] 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., 85 (2023), pp. 441–478, https://doi.org/10.1007/s10589-023-00461-8.
[28] R. Lang, *A note on the measurability of convex sets*, Arch. Math., 47 (1986), pp. 90–92, https://doi.org/10.1007/BF01202504.
[29] J. Milz and M. Ulbrich, *Sample size estimates for risk-neutral semilinear PDE-constrained optimization*, SIAM J. Optim., 34 (2024), pp. 844–869, https://doi.org/10.1137/22M1512636.
[30] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM J. Optim., 17 (2006), pp. 969–996, https://doi.org/10.1137/050622328.
[31] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed., Springer-Verlag, Berlin, 2006.
[32] B. K. Pagnoncelli, S. Ahmed, and A. Shapiro, *Sample average approximation method for chance constrained programming: Theory and applications*, J. Optim. Theory Appl., 142 (2009), pp. 399–416, https://doi.org/10.1007/s10957-009-9523-6.
[33] A. Peña Ordieres, J. R. Luedtke, and A. Wächter, *Solving chance-constrained problems via a smooth sample-based nonlinear approximation*, SIAM J. Optim., 30 (2020), pp. 2221–2250, https://doi.org/10.1137/19M1261985.
[34] P. Pérez-Aros, C. Quiñinao, and M. Tejo, *Control in probability for SDE models of growth population*, Appl. Math. Optim., 86 (2022), 44, https://doi.org/10.1007/s00245-022-09915-7.
[35] A. Prékopa, *Stochastic Programming*, Math. Appl. 324, Springer, Dordrecht, 1995, https://doi.org/10.1007/978-94-017-3087-7.
[36] M. Schuster, E. Strauch, M. Gugat, and J. Lang, *Probabilistic constrained optimization on flow networks*, Optim. Eng., 23 (2022), pp. 1–50, https://doi.org/10.1007/s11081-021-09619-x.
[37] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming:Modeling and Theory*, 3rd ed., MOS-SIAM Ser. Optim. 28, SIAM, Philadelphia, 2021, https://doi.org/10.1137/1.9781611976595.
[38] 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 Nonlinear Systems—Recent Developments and Advances, B. Yang and D. Stipanović, eds., IntechOpen, 2023, ch. 7, https://doi.org/10.5772/intechopen.104620.
[39] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, translated from the 2005 German original by J. Sprekels, Grad. Stud. Math. 112, American Mathematical Society, Providence, RI, 2010.
[40] W. van Ackooij and R. Henrion, *Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions*, SIAM J. Optim., 24 (2014), pp. 1864–1889, https://doi.org/10.1137/130922689.
[41] W. van Ackooij and R. Henrion, *(Sub-)gradient formulae for probability functions of random inequality systems under Gaussian distribution*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 63–87, https://doi.org/10.1137/16M1061308.
[42] W. van Ackooij, R. Henrion, and H. Zidani, *Pontryagin's principle for some probabilistic control problems*, Appl. Math. Optim., 90 (2024), 5, https://doi.org/10.1007/s00245-024-10151-4.