Sample average approximations of strongly convex stochastic programs in Hilbert spaces

Johannes Milz[1]

Abstract

We analyze the tail behavior of solutions to sample average approximations (SAAs) of stochastic programs posed in Hilbert spaces. We require that the integrand be strongly convex with the same convexity parameter for each realization. Combined with a standard condition from the literature on stochastic programming, we establish non-asymptotic exponential tail bounds for the distance between the SAA solutions and the stochastic program's solution, without assuming compactness of the feasible set. Our assumptions are verified on a class of infinite-dimensional optimization problems governed by affine-linear partial differential equations with random inputs. We present numerical results illustrating our theoretical findings.

1 Introduction

We apply the sample average approximation (SAA) to a class of strongly convex stochastic programs posed in Hilbert spaces, and study the tail behavior of the distance between SAA solutions and their true counterparts. Our work sheds light on the number of samples needed to reliably estimate solutions to infinite-dimensional, linear-quadratic optimal control problems governed by affine-linear partial differential equations (PDEs) with random inputs, a class of optimization problems that has received much attention recently [30, 40, 42]. Our analysis requires that the

---

✉ Johannes Milz
  milz@ma.tum.de

1   Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany

integrand be strongly convex with the same convexity parameter for each random element's sample. This assumption is fulfilled for convex optimal controls problems with a strongly convex control regularizer, such as those considered in [30, 40, 42]. Throughout the paper, a function $f : H \to \mathbb{R} \cup \{\infty\}$ is α-strongly convex with parameter $\alpha > 0$ if $f(\cdot) - (\alpha/2)\| \cdot \|_H^2$ is convex, where H is a real Hilbert space with norm $\| \cdot \|_H$. Moreover, a function on a real Hilbert space is strongly convex if it is α-strongly convex with some parameter $\alpha > 0$.

We consider the potentially infinite-dimensional stochastic program

$$\min_{u \in U} \{ f(u) = \mathbb{E}[J(u, \xi)] + \Psi(u) \}, \tag{1}$$

where U is a real, separable Hilbert space, $\Psi : U \to \mathbb{R} \cup \{\infty\}$ is proper, lower-semicontinuous and convex, and $J : U \times \Xi \to \mathbb{R}$ is the integrand. Moreover, ξ is a random element mapping from a probability space to a complete, separable metric space Ξ equipped with its Borel σ-field. We also use $\xi \in \Xi$ to represent a deterministic element.

Let $\xi^1, \xi^2, \ldots$ be independent identically distributed Ξ-valued random elements defined on a complete probability space $(\Omega, \mathcal{F}, P)$ such that each ξ^i has the same distribution as that of ξ. The SAA problem corresponding to (1) is

$$\min_{u \in U} \left\{ f_N(u) = \frac{1}{N} \sum_{i=1}^{N} J(u, \xi^i) + \Psi(u) \right\}, \tag{2}$$

We define $F : U \to \mathbb{R} \cup \{\infty\}$ and the sample average function $F_N : U \to \mathbb{R}$ by

$$F(u) = \mathbb{E}[J(u, \xi)] \quad \text{and} \quad F_N(u) = \frac{1}{N} \sum_{i=1}^{N} J(u, \xi^i). \tag{3}$$

Since we assume that the random elements $\xi^1, \xi^2, \ldots$ are defined on the common probability space $(\Omega, \mathcal{F}, P)$, we can view the functions f_N and F_N as defined on $U \times \Omega$ and the solution u_N^* to (2) as a mapping from Ω to U. The second argument of f_N and of F_N is often dropped.

Let u^* be a solution to (1) and u_N^* be a solution to (2). We assume that $J(\cdot, \xi)$ is α-strongly convex with parameter $\alpha > 0$ for each $\xi \in \Xi$. Furthermore, we assume that $F(\cdot)$ and $J(\cdot, \xi)$ for all $\xi \in \Xi$ are Gâteaux differentiable. Under these assumptions, we establish the error estimate

$$\alpha \|u^* - u_N^*\|_U \leq \|\nabla F_N(u^*) - \nabla F(u^*)\|_U, \tag{4}$$

valid with probability one. If $\|\nabla_u J(u^*, \xi)\|_U$ is integrable, then $\nabla F_N(u^*)$ is just the empirical mean of $\nabla F(u^*)$ since $F(\cdot)$ and $J(\cdot, \xi)$ for all $\xi \in \Xi$ are convex and Gâteaux differentiable at u^*; see Lemma 3. Hence we can analyze the mean square error $\mathbb{E}[\|u^* - u_N^*\|_U^2]$ and the exponential tail behavior of $\|u^* - u_N^*\|_U$ using standard conditions from the literature on stochastic programming. To obtain a bound on $\mathbb{E}[\|u^* - u_N^*\|_U^2]$, we assume that there exists $\sigma > 0$ with

$$\mathbb{E}[\|\nabla_u J(u^*, \xi) - \nabla F(u^*)\|_U^2] \leq \sigma^2, \tag{5}$$

yielding with (4) the bound

$$\mathbb{E}[\|u^* - u_N^*\|_U^2] \leq \sigma^2/(\alpha^2 N). \tag{6}$$

To derive exponential tail bounds on $\|u^* - u_N^*\|_U$, we further assume the existence of $\tau > 0$ with

$$\mathbb{E}[\exp(\tau^{-2}\|\nabla_u J(u^*, \xi) - \nabla F(u^*)\|_U^2)] \leq e. \tag{7}$$

This condition and its variants are used, for example, in [15, 25, 44]. Using Jensen's inequality, we find that (7) implies (5) with $\sigma^2 = \tau^2$ [44, p. 1584]. Combining (4) and (7) with the exponential moment inequality proven in [48, Theorem 3], we establish the exponential tail bound, our main contribution,

$$\text{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq 2\exp(-\tau^{-2}N\varepsilon^2\alpha^2/3) \quad \text{for all} \quad \varepsilon > 0. \tag{8}$$

This bound solely depends on the characteristics of J but not on properties of the feasible set, $\{ u \in U : \Psi(u) < \infty \}$, other than its convexity. For each $\delta \in (0, 1)$, the exponential tail bound yields, with a probability of at least $1 - \delta$,

$$\|u^* - u_N^*\|_U < \frac{\tau}{\alpha}\sqrt{\frac{3\ln(2/\delta)}{N}}. \tag{9}$$

In particular, if $\varepsilon > 0$ and $N \geq \frac{3\tau^2}{\alpha^2\varepsilon^2}\ln(2/\delta)$, then $\|u^* - u_N^*\|_U < \varepsilon$ with a probability of at least $1 - \delta$, that is, u^* can be estimated reliably via u_N^*.

Requiring $J(\cdot, \xi)$ to be α-strongly convex for each $\xi \in \Xi$ is a restrictive assumption. However, it is fulfilled for the following class of stochastic programs:

$$\min_{u \in U} \{ (1/2)\mathbb{E}[\|K(\xi)u + h(\xi)\|_H^2] + (\alpha/2)\|u\|_U^2 + \Psi(u) \}, \tag{10}$$

where $\alpha > 0$, H and U are real Hilbert spaces, and $K(\xi) : U \to H$ is a bounded, linear operator and $h(\xi) \in H$ for each $\xi \in \Xi$. The control problems governed by affine-linear PDEs with random inputs considered, for example, in [21, 22, 30, 41, 42] can be formulated as instances of (10). In many of these works, the operator $K(\xi)$ is compact for each $\xi \in \Xi$, the expectation function $F_1 : U \to \mathbb{R}$ defined by $F_1(u) = (1/2)\mathbb{E}[\|K(\xi)u + h(\xi)\|_H^2]$ is twice continuously differentiable, and U is infinite-dimensional. In this case, the function F_1 generally lacks strong convexity. This may suggest that the α-strong convexity of the objective function of (10) is solely implied by the function $(\alpha/2)\|\cdot\|_U^2 + \Psi(\cdot)$. The lack of the expectation function's strong convexity is essentially known [6, p. 3]. For example, if the set Ξ has finite cardinality, then the Hessian $\nabla^2 F_1(0)$ is the finite sum of compact operators and hence F_1 lacks strong convexity; see Sect. 6.

A common notion used to analyze the SAA solutions is that of an ε-optimal solution [54, 56, 57].[1] We instead study the tail behavior of $\|u^* - u_N^*\|_U$ since in the literature on PDE-constrained optimization the focus is on studying the proximity of approximate solutions to the "true" ones. For example, when analyzing finite element approximations of PDE-constrained problems, bounds on the error $\|w^* - w_h^*\|_U$ as functions of the discretization parameters h are often established [28, 60], where w^* is the solution to a control problem and w_h^* is the solution to its finite element approximation. The estimate (4) is similar to that established in [28, p. 49] for the variational discretization—a finite element approximation—of a deterministic, linear-quadratic control problem. Since both the variational discretization and the SAA approach yield perturbed optimization problems, it is unsurprising that similar techniques can be used for some parts of the perturbation analysis.

The SAA approach has thoroughly been analyzed, for example, in [4, 7, 50, 54, 56, 57]. Some consistency results for the SAA solutions and finite-sample size estimates require the compactness and total boundedness of the feasible set, respectively. However, in the literature on PDE-constrained optimization, the feasible sets are commonly noncompact; see, e.g., [29, Sect. 1.7.2.3]. Assuming that the function F defined in (3) is α-strongly convex with $\alpha > 0$, Kouri and Shapiro [35, eq. (42)] establish

$$\alpha\|u^* - u_N^*\|_U \leq \|\nabla F_N(u_N^*) - \nabla F(u_N^*)\|_U. \tag{11}$$

The setting in [35] corresponds to Ψ being the indicator function of a closed, convex, nonempty subset of U. In contrast to the estimate (4), the right-hand side in (11) depends on the random control u_N^*. This dependence implies that the right-hand side in (11) is more difficult to analyze than that in (4). However, the convexity assumption on F made in [35] is weaker than ours which requires the function $J(\cdot, \xi)$ be α-strongly convex for all $\xi \in \Xi$. The right-hand side (11) may be analyzed using the approaches developed in [53, Sects. 2 and 4].

For finite-dimensional optimization problems, the number of samples, required to obtain ε-optimal solutions via the SAA approach, can explicitly depend on the problem's dimension [1, 55, Example 1], [25, Proposition 2]. Guigues, Juditsky, and Nemirovski [25] demonstrate that confidence bounds on the optimal value of stochastic, convex, finite-dimensional programs, constructed via SAA optimal values, do not explicitly depend on the problem's dimension. This property is shared by our exponential tail bound.

After the initial version of the manuscript was submitted, we became aware of the papers [52, 61] where assumptions similar to those used to derive (6) and (8) are utilized to analyze the reliability of SAA solutions. For unconstrained minimization in $\mathbb{R}^n$ with $\Psi = 0$, tail bounds for $\|u^* - u_N^*\|_2$ are established in [61] under the assumption that $J(\cdot, \xi)$ is α-strong convex for all $\xi \in \Xi$ and some $\alpha > 0$. Here, $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^n$. Assuming further that $\|\nabla_u J(u^*, \xi)\|_2$ is essentially bounded by $L > 0$, the author establishes

---

[1] A point $\bar{x} \in X$ is an ε-optimal solution to $\inf_{x \in X} f(x)$ if $f(\bar{x}) \leq \inf_{x \in X} f(x) + \varepsilon$.

$$\text{Prob}(\|u^* - u_N^*\|_2 \geq \varepsilon) \leq 2 \exp\left(-\frac{N\alpha^2\varepsilon^2}{2L^2}\left(1 + \frac{\alpha\varepsilon}{3L}\right)^{-1}\right) \tag{12}$$

if $\varepsilon \in (0, L/\alpha]$, and the right-hand side in (12) is zero otherwise [61, Corollary 2]. While (12) is similar to (8) with $\tau = L$, its derivation exploits the essential boundedness of $\|\nabla_u J(u^*, \xi)\|_2$ which is generally more restrictive than (7). The author establishes further tail bounds for $\|u^* - u_N^*\|_2$ under different sets of assumptions on $J(\cdot, \xi)$, and provides exponential tail bounds for $f(u_N^*) - f(u^*)$ assuming that $J(\cdot, \xi)$ is Lipschitz continuous with a Lipschitz constant independent of ξ (see [61, Theorem 5]). For the possibly infinite-dimensional program (1), similar assumptions are used in [52, Theorem 2] to establish a non-exponential tail bound for $f(u_N^*) - f(u^*)$. While tail bounds for $f(u_N^*) - f(u^*)$ are derived in [52, 61], the assumptions used to derive (6) and (8) do not imply bounds on $f(u_N^*) - f(u^*)$.

Hoffhues et al. [30] provide qualitative and quantitative stability results for the optimal value and for the optimal solutions of stochastic, linear-quadratic optimization problems posed in Hilbert spaces, similar to those in (10), with respect to Fortet–Mourier and Wasserstein metrics. These stability results are valid for approximating probability measures other than the empirical one, which is used to define the SAA problem (2). However, the convergence rate $1/N$ for $\mathbb{E}[\|u^* - u_N^*\|_U^2]$, and exponential tail bounds on $\|u^* - u_N^*\|_U$ are not established in [30]. For a class of constrained, linear elliptic control problems, Römisch and Surowiec [49] demonstrate the consistency of the solutions and the optimal value, the convergence rate $1/\sqrt{N}$ for $\mathbb{E}[\|u^* - u_N^*\|_U]$ and for $\mathbb{E}[|f_N(u_N^*) - f(u^*)|]$, and the convergence in distribution of $\sqrt{N}(f_N(u_N^*) - f(u^*))$ to a real-valued random variable. These results are established using empirical process theory and are built on smoothness of the random elliptic operator and right-hand side with respect to the parameters. While our assumptions yield the mean square error bound (6) and the exponential tail bound (8), further conditions may be required to establish bounds on $\mathbb{E}[|f_N(u_N^*) - f(u^*)|]$. A bound on $\mathbb{E}[\|u^* - u_N^*\|_U^2]$ related to (6) is established in [41, Theorem 4.1] for class of linear elliptic control problems.

Besides considering risk-neutral, convex control problems with PDEs which can be expressed as those in Sect. 6, the authors of [40, 42] study the minimization of $u \mapsto \text{Prob}(J(u, \xi) \geq \rho)$, where $\rho \in \mathbb{R}$ and evaluating $J(u, \xi)$ requires solving a PDE. Furthermore, the authors of Marín et al. [40] and Martínez-Frutos and Esparza [42] prove the existence of solutions and use stochastic collocation to discretize the expected values. In [42, Sect. 5.3], the authors adaptively combine a Monte Carlo sampling approach with a stochastic Galerkin finite element method to reduce the computational costs, but error bounds are not established. Stochastic collocation is also used, for example, in [21, 34]. Further approaches to discretize the expected value in (10) are, for example, quasi-Monte Carlo sampling [26] and low-rank tensor approximations [20]. A solution method for (1) is (robust) stochastic approximation. It has thoroughly been analyzed in [38, 44] for finite-dimensional and in [22, 24, 45] for infinite-dimensional optimization problems. For reliable ε-optimal solutions, the sample size estimates established in [44, Proposition 2.2] do not explicitly depend on the problem's dimension.

After providing some notation and preliminaries in Sect. 2, we establish exponential tail bounds for Hilbert space-valued random sums in Sect. 3. Combined with optimality conditions and the integrand's α-strong convexity, we establish exponential tail and mean square error bounds for SAA solutions in Sect. 4. Sect. 5 demonstrates the optimality of the tail bounds. We apply our findings to linear-quadratic control under uncertainty in Sect. 6, and identify a problem class that violates the integrability condition (7). Numerical results are presented in Sect. 7. In Sect. 8, we illustrate that the "dynamics" of finite- and infinite-dimensional stochastic programs can be quite different.

2 Notation and preliminaries

Throughout the manuscript, we assume the existence of solutions to (1) and to (2). We refer the reader to Kouri and Surowiec [36, Proposition 3.12] and Hoffhues et al. [30, Theorem 1] for theorems on the existence of solutions to infinite-dimensional stochastic programs.

The set $\text{dom}\,\Psi = \{\, u \in U : \Psi(u) < \infty \,\}$ is the domain of Ψ. The indicator function $I_{U_0} : U \to \mathbb{R} \cup \{\infty\}$ of a nonempty set $U_0 \subset U$ is defined by $I_{U_0}(u) = 0$ if $u \in U_0$ and $I_{U_0}(u) = \infty$ otherwise. Let $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ be a probability space. A Banach space W is equipped with its Borel σ-field $\mathcal{B}(W)$. We denote by $(\cdot, \cdot)_H$ the inner product of a real Hilbert space H equipped with the norm $\| \cdot \|_H$ given by $\|v\|_H = \sqrt{(v, v)_H}$ for all $v \in H$. For a real, separable Hilbert space H, $\eta : \hat{\Omega} \to H$ is a mean-zero Gaussian random vector if $(v, \eta)_H$ is a mean-zero Gaussian random variable for each $v \in H$ [64, pp. 58–59]. For a metric space V, a mapping $\mathsf{f} : V \times \hat{\Omega} \to W$ is a Carathéodory mapping if $\mathsf{f}(\cdot, \omega)$ is continuous for every $\omega \in \hat{\Omega}$ and $\mathsf{f}(x, \cdot)$ is $\hat{\mathcal{F}}\text{-}\mathcal{B}(W)$-measurable for all $x \in V$.

For two Banach spaces V and W, $\mathscr{L}(V, W)$ is the space of bounded, linear operators from V to W, and $V^* = \mathscr{L}(V, \mathbb{R})$. We denote by $\langle \cdot, \cdot \rangle_{V^*, V}$ the dual pairing of V^* and V. A function $v : \hat{\Omega} \to W$ is strongly measurable if there exists a sequence of simple functions $v_k : \hat{\Omega} \to W$ such that $v_k(\omega) \to v(\omega)$ as $k \to \infty$ for all $\omega \in \hat{\Omega}$ [31, Def. 1.1.4]. An operator-valued function $Y : \hat{\Omega} \to \mathscr{L}(V, W)$ is strongly measurable if the function $\omega \mapsto Y(\omega)x$ is strongly measurable for each $x \in V$ [31, Def. 1.1.27]. Moreover, an operator-valued function $Y : \hat{\Omega} \to \mathscr{L}(V, W)$ is uniformly measurable if there exists a sequence of simple operator-valued functions $Y_k : \hat{\Omega} \to \mathscr{L}(V, W)$ with $Y_k(\omega) \to Y(\omega)$ as $k \to \infty$ for all $\omega \in \hat{\Omega}$. An operator $K \in \mathscr{L}(V, W)$ is compact if the closure of $K(V_0)$ is compact for each bounded set $V_0 \subset V$. For two real Hilbert spaces H_1 and H_2, $K^* \in \mathscr{L}(H_2, H_1)$ is the (Hilbert space-)adjoint operator of $K \in \mathscr{L}(H_1, H_2)$ and is defined by $(Kv_1, v_2)_{H_2} = (v_1, K^*v_2)_{H_1}$ for all $v_1 \in H_1$ and $v_2 \in H_2$ [37, Def. 3.9.1]. For a bounded domain $D \subset \mathbb{R}^d$, $L^2(D)$ is the Lebesgue space of square-integrable functions and $L^1(D)$ is that of integrable functions. The Hilbert space $H_0^1(D)$ consists of all $v \in L^2(D)$ with weak derivatives in $L^2(D)^d$ and with zero boundary traces. We define $H^{-1}(D) = H_0^1(D)^*$.

3 Exponential tail bounds for Hilbert space-valued random sums

We establish two exponential tail bounds for Hilbert space-valued random sums which are direct consequences of known results [47, 48]. Below, (Θ, Σ, μ) denotes a probability space. Proofs are presented at the end of the section.

Theorem 1 *Let H be a real, separable Hilbert space. Suppose that $Z_i : \Theta \to H$ for $i = 1, 2, \ldots$ are independent, mean-zero random variables such that $\mathbb{E}[\exp(\tau^{-2}\|Z_i\|_H^2)] \leq e$ for some $\tau > 0$. Then, for each $N \in \mathbb{N}$, $\varepsilon \geq 0$,*

$$\mathrm{Prob}(\|Z_1 + \cdots + Z_N\|_H \geq N\varepsilon) \leq 2\exp(-\tau^{-2}\varepsilon^2 N/3). \tag{13}$$

If in addition $\|Z_i\|_H \leq \tau$ with probability one for $i = 1, 2, \ldots$, then the upper bound in (13) improves to $2\exp(-\tau^{-2}\varepsilon^2 N/2)$ [47, Theorem 3.5].

As an alternative to the condition $\mathbb{E}[\exp(\tau^{-2}\|Z\|_H^2)] \leq e$ used in Theorem 1 for $\tau > 0$ and a random vector $Z : \Theta \to H$, we can express sub-Gaussianity with $\mathbb{E}[\cosh(\lambda\|Z\|_H)] \leq \exp(\lambda^2\sigma^2/2)$ for all $\lambda \in \mathbb{R}$ and some $\sigma > 0$. While these two conditions are equivalent up to problem-independent constants (see the proof of [11, Lemma 1.6 on p. 9] and Lemma 1), the constant σ can be smaller than τ. For example, if $Z : \Theta \to H$ is a H-valued, mean-zero Gaussian random vector, then the latter condition holds with $\sigma^2 = \mathbb{E}[\|Z\|_H^2]$ [48, Rem. 4]. However, if $H = \mathbb{R}$ then $\tau^2 = 2\sigma^2/(1 - \exp(-2)) \approx 2.31\sigma^2$ [11, p. 9].

Proposition 1 *Let H be a real, separable Hilbert space, and let $Z_i : \Theta \to H$ be independent, mean-zero random vectors such that $\mathbb{E}[\cosh(\lambda\|Z_i\|_H)] \leq \exp(\lambda^2\sigma^2/2)$ for all $\lambda \in \mathbb{R}$ and some $\sigma > 0$ $(i = 1, 2, \ldots)$. Then, for each $N \in \mathbb{N}$, $\varepsilon \geq 0$,*

$$\mathrm{Prob}(\|Z_1 + \cdots + Z_N\|_H \geq N\varepsilon) \leq 2\exp(-\sigma^{-2}\varepsilon^2 N/3).$$

We apply the following two facts to prove Theorem 1 and Proposition 1.

Theorem 2 (See [48, Theorem 3]) *Let H be a real, separable Hilbert space. Suppose that $Z_i : \Theta \to H$ $(i = 1, \ldots, N \in \mathbb{N})$ are independent, mean-zero random vectors. Then, for all $\lambda \geq 0$,*

$$\mathbb{E}[\cosh(\lambda\|Z_1 + \cdots + Z_N\|_H)] \leq \prod_{i=1}^{N} \mathbb{E}[\exp(\lambda\|Z_i\|_H) - \lambda\|Z_i\|_H].$$

Lemma 1 *If $\sigma > 0$ and $X : \Theta \to \mathbb{R}$ is measurable with $\mathbb{E}[\exp(\sigma^{-2}|X|^2)] \leq e$,*

$$\mathbb{E}[\exp(\lambda|X|) - \lambda|X|] \leq \exp(3\lambda^2\sigma^2/4) \quad \text{for all} \quad \lambda \in \mathbb{R}_+. \tag{14}$$

Proof The proof is based on the proof of [56, Proposition 7.72].

Fix $\lambda \in [0, 4/(3\sigma)]$. For all $s \in \mathbb{R}$, $\exp(s) \leq s + \exp(9s^2/16)$ [56, p. 449]. Using Jensen's inequality and $\mathbb{E}[\exp(|X|^2/\sigma^2)] \leq e$, we obtain

$$\mathbb{E}[e^{\lambda|X|} - \lambda|X|] \leq \mathbb{E}[e^{9\lambda^2|X|^2/16}] \leq \mathbb{E}[e^{|X|^2/\sigma^2}]^{9\lambda^2\sigma^2/16} \leq e^{9\lambda^2\sigma^2/16}. \tag{15}$$

Now, fix $\lambda \geq 4/(3\sigma)$. For all $s \in \mathbb{R}$, Young's inequality yields $\lambda s \leq 3\lambda^2 \sigma^2/8 + 2s^2/(3\sigma^2)$. Combined with Jensen's inequality, $\mathbb{E}[\exp(|X|^2/\sigma^2)] \leq e$, and $2/3 \leq 3\sigma^2\lambda^2/8$, we get

$$\mathbb{E}[e^{\lambda|X|} - \lambda|X|] \leq \mathbb{E}[e^{\lambda|X|}] \leq e^{3\lambda^2\sigma^2/8}\mathbb{E}[e^{2|X|^2/(3\sigma^2)}] \leq e^{3\lambda^2\sigma^2/8+2/3} \leq e^{3\lambda^2\sigma^2/4}.$$

Together with (15), we obtain (14). $\square$

Proof (Proof of Theorem 1) We use a Chernoff-type approach to establish (13). Fix $\lambda > 0$, $\varepsilon \geq 0$, and $N \in \mathbb{N}$. We define $S_N = Z_1 + \cdots + Z_N$. Using $\mathbb{E}[\exp(\tau^{-2}\|Z_i\|_H^2)] \leq e$ and applying Lemma 1 to $X = \|Z_i\|_H$, we find that

$$\prod_{i=1}^N \mathbb{E}[\exp(\lambda\|Z_i\|_H) - \lambda\|Z_i\|_H] \leq \prod_{i=1}^N \exp(3\lambda^2\tau^2/4) = \exp(3\lambda^2\tau^2 N/4).$$

Combined with Markov's inequality, Theorem 2, and $\exp \leq 2\cosh$, we obtain

$$\text{Prob}(\|S_N\|_H \geq N\varepsilon) \leq e^{-\lambda N\varepsilon}\mathbb{E}[e^{\lambda\|S_N\|_H}] \leq 2e^{-\lambda N\varepsilon}\mathbb{E}[\cosh(\lambda\|S_N\|_H)]$$
$$\leq 2e^{-\lambda N\varepsilon + 3\lambda^2\tau^2 N/4}.$$

Minimizing the right-hand side over $\lambda > 0$ yields (13). $\square$

Proof (Proof of Proposition 1) We have $\exp(s) - s \leq \cosh(s\sqrt{3/2})$ for all $s \in \mathbb{R}$. Hence, the assumptions ensure $\mathbb{E}[\exp(\lambda\|Z_i\|_H) - \lambda\|Z_i\|_H] \leq \exp(3\lambda^2\sigma^2/4)$ for all $\lambda \in \mathbb{R}$. The remainder of the proof is as that of Theorem 1. $\square$

4 Exponential tail bounds for SAA solutions

We state conditions that allow us to derive exponential bounds on the tail probabilities of the distance between SAA solutions and their true counterparts. In Sect. 6, we demonstrate that our conditions are fulfilled for many linear-quadratic control problems considered in the literature.

4.1 Assumptions and measurability of SAA solutions

Throughout the manuscript, u^* is assumed to be a solution to (1).

Assumption 1

(a) The space U is a real, separable Hilbert space.
(b) The function $\Psi : U \to \mathbb{R} \cup \{\infty\}$ is convex, proper, and lower-semicontinuous.
(c) The integrand $J : U \times \Xi \to \mathbb{R}$ is a Carathéodory function, and for some $\alpha > 0$, $J(\cdot, \xi)$ is α-strongly convex for each $\xi \in \Xi$.

(d) The function $J(\cdot, \xi)$ is Gâteaux differentiable on a convex neighborhood of dom Ψ for all $\xi \in \Xi$, and $\nabla_u J(u^*, \cdot) : \Xi \to U$ is measurable.

(e) The map $F : U \to \mathbb{R} \cup \{\infty\}$ defined in (3) is Gâteaux differentiable at u^*.

Lemma 2 *Let Assumptions* 1*(a)–(c) hold. If* $u_N^* : \Omega \to U$ *is a solution to* (2)*, then* u_N^* *is the unique solution to* (2) *and is measurable.*

Proof For each $\omega \in \Omega$, the SAA problem's objective function $f_N(\cdot, \omega)$ is strongly convex and hence u_N^* is the unique solution to (2). The function $\inf_{u \in U} f_N(u, \cdot) : \Omega \to \mathbb{R}$ is measurable [12, Corollary VII-2] (see also [12, Lemma III.39]). Hence the multifunction $\arg\inf_{u \in U} f_N(u, \cdot)$ is single-valued and has a measurable selection [5, Theorem 8.2.9]. Therefore $u_N^* : \Omega \to U$ is measurable. $\qquad\square$

We impose conditions on the integrability of $\nabla_u J(u^*, \xi) - \nabla F(u^*)$.

Assumption 2

(a) For some $\sigma > 0$, $\mathbb{E}[\|\nabla_u J(u^*, \xi) - \nabla F(u^*)\|_U^2] \le \sigma^2$.
(b) For some $\tau > 0$, $\mathbb{E}[\exp(\tau^{-2}\|\nabla_u J(u^*, \xi) - \nabla F(u^*)\|_U^2)] \le \mathrm{e}$.

Assumption 2(b) implies Assumption 2(a) with $\sigma^2 = \tau^2$ [44, p. 1584]. Assumption 2(b) and its variants are standard conditions in the literature on stochastic programming [15, p. 679], [25, pp. 1035–1036], [44, Eq. (2.50)]. For example, if $\nabla_u J(u^*, \xi) - \nabla F(u^*)$ is essentially bounded, then Assumption 2(b) is fulfilled. More generally, if $\nabla_u J(u^*, \xi) - \nabla F(u^*)$ is γ-sub-Gaussian, then Assumption 2(b) holds true [17, Theorem 3.4].

4.2 Exponential tail and mean square error bounds

We establish exponential tail and mean square error bounds on $\|u^* - u_N^*\|_U$.

Theorem 3 *Let* u^* *be a solution to* (1) *and let* u_N^* *be a solution to* (2)*. If Assumptions* 1 *and* 2*(a) hold, then*

$$\mathbb{E}[\|u^* - u_N^*\|_U^2] \le \sigma^2/(\alpha^2 N). \tag{16}$$

If in addition Assumption 2*(b) holds, then for all* $\varepsilon > 0$,

$$\mathrm{Prob}(\|u^* - u_N^*\|_U \ge \varepsilon) \le 2\exp(-\tau^{-2}N\varepsilon^2\alpha^2/3). \tag{17}$$

We prepare our proof of Theorem 3.

Lemma 3 *If Assumptions* 1 *and* 2*(a) hold, then* $\mathbb{E}[\nabla_u J(u^*, \xi)] = \nabla F(u^*)$.

Proof Using Assumptions 1(a) and (c)–(e), we have $\mathbb{E}[(\nabla_u J(u^*, \xi), v)_U] = (\nabla F(u^*), v)_U$ for all $v \in U$; cf. [25, p. 1050]. Owing to Assumptions 1(e) and 2(a), the mapping

$\nabla_u J(u^*, \xi)$ is integrable. Hence $\mathbb{E}[(\nabla_u J(u^*, \xi), v)_U] = (\mathbb{E}[\nabla_u J(u^*, \xi)], v)_U$ for all $v \in U$ (cf. [9, p. 78]).   □

Lemma 4 *If Assumption* 1 *holds, then the function* F_N *defined in* (3) *is Gâteaux differentiable on a neighborhood of* dom Ψ *and with probability one,*

$$(\nabla F_N(u_2) - \nabla F_N(u_1), u_2 - u_1)_U \geq \alpha \|u_2 - u_1\|_U^2 \text{ for all } u_1, u_2 \in \text{dom } \Psi. \quad (18)$$

Proof Since, for each $\xi \in \Xi$, $J(\cdot, \xi)$ is α-strongly convex and Gâteaux differentiable on a convex neighborhood V of dom Ψ, the sum rule and the definition of F_N imply its Gâteaux differentiability on V and (18) [45, p. 48].   □

Lemma 5 *Let Assumption* 1 *hold and let* $\omega \in \Omega$ *be fixed. Suppose that* u^* *is a solution to* (1) *and that* $u_N^* = u_N^*(\omega)$ *is a solution to* (2). *Then*

$$(\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \geq 0. \quad (19)$$

Proof Following the proof of [32, Theorem 4.42], we obtain for all $u \in \text{dom } \Psi$,

$$\begin{aligned}(\nabla F(u^*), u - u^*)_U + \Psi(u) - \Psi(u^*) &\geq 0, \\ (\nabla F_N(u_N^*), u - u_N^*)_U + \Psi(u) - \Psi(u_N^*) &\geq 0. \end{aligned} \quad (20)$$

We have $\Psi(u^*), \Psi(u_N^*) \in \mathbb{R}$. Choosing $u = u_N^*$ in the first and $u = u^*$ in the second estimate in (20), and adding the resulting inequalities yields (19).   □

Lemma 6 *Under the hypotheses of Lemma* 5, *we have*

$$\alpha \|u^* - u_N^*\|_U \leq \|\nabla F_N(u^*) - \nabla F(u^*)\|_U. \quad (21)$$

Proof Choosing $u_2 = u^*$ and $u_1 = u_N^*$ in (18), we find that

$$(\nabla F_N(u^*) - \nabla F_N(u_N^*), u^* - u_N^*)_U \geq \alpha \|u^* - u_N^*\|_U^2.$$

Combined with (19), and the Cauchy–Schwarz inequality, we get

$$\begin{aligned}\alpha \|u^* - u_N^*\|_U^2 &\leq (\nabla F_N(u^*) - \nabla F_N(u_N^*), u^* - u_N^*)_U \\ &\quad + (\nabla F_N(u_N^*) - \nabla F(u^*), u^* - u_N^*)_U \\ &\leq \|\nabla F_N(u^*) - \nabla F(u^*)\|_U \|u^* - u_N^*\|_U.\end{aligned}$$

□

Proof (Proof of Theorem 3) Lemma 2 ensures the measurability of $u_N^* : \Omega \to U$. We define $q : \Xi \to U$ by $q(\xi) = \nabla_u J(u^*, \xi) - \nabla F(u^*)$. Assumption 1(c) and (e) ensure that q is well-defined and measurable. Hence, the random vectors $Z_i = q(\xi^i)$ $(i = 1, 2, \ldots)$ are independent identically distributed, and Lemma 3 ensures that they have zero mean. Using the definitions of F and of F_N provided in (3), the Gâteaux differentiability of F at u^* [see Assumption 1(e)], and Lemma 4, we obtain

$$\nabla F_N(u^*) - \nabla F(u^*) = \frac{1}{N} \sum_{i=1}^{N} \left(\nabla_u J(u^*, \xi^i) - \nabla F(u^*) \right) = \frac{1}{N} \sum_{i=1}^{N} Z_i.$$

Now, we prove (16). Combining the above statements with the separability of the Hilbert space U, we get $\mathbb{E}[\| \sum_{i=1}^{N} Z_i \|_U^2] = \sum_{i=1}^{N} \mathbb{E}[\|Z_i\|_U^2]$ [64, p. 79]. For $i = 1, 2, \ldots$, Assumption 2(a) yields $\mathbb{E}[\|Z_i\|_U^2] \le \sigma^2$. Together with the estimate (21), we find that

$$\alpha^2 \mathbb{E}[\|u^* - u_N^*\|_U^2] \le \mathbb{E}[\|\nabla F_N(u^*) - \nabla F(u^*)\|_U^2] \le \sigma^2/N,$$

yielding the mean square error bound (16).

Next, we establish (17). Fix $\varepsilon > 0$. If $\|u^* - u_N^*\|_U \ge \varepsilon$, then the estimate (21) ensures that $\| \sum_{i=1}^{N} Z_i \|_U \ge N\alpha\varepsilon$. For $i = 1, 2, \ldots$, Assumption 2(b) implies that $\mathbb{E}[\exp(\tau^{-2}\|Z_i\|_U^2)] \le e$. Applying Theorem 1, we get

$$\text{Prob}\left(\left\| u^* - u_N^* \right\|_U \ge \varepsilon \right) \le \text{Prob}\left(\left\| \sum_{i=1}^{N} Z_i \right\|_U \ge N\alpha\varepsilon \right)$$

$$\le 2e^{-\tau^{-2}\varepsilon^2\alpha^2 N/3}.$$

Hence the exponential tail bound (17) holds true. $\qquad\square$

5 Optimality of SAA solutions' exponential tail bounds

We show that the dependence of the tail bound (17) on the problem data is essentially optimal for the problem class modeled by Assumptions 1 and 2(b).

Our example is inspired by that analyzed in [55, Example 1]. We consider

$$\min_{u \in L^2(0,1)} \ \mathbb{E}[(\alpha/2)\|u\|_{L^2(0,1)}^2 - (h(\xi), u)_{L^2(0,1)}], \tag{22}$$

where $\alpha > 0$, φ_1, $\varphi_2 \in L^2(0, 1)$ are orthonormal, $h : \mathbb{R}^2 \to L^2(0, 1)$ is given by $h(\xi) = \xi_1 \varphi_1 + \xi_2 \varphi_2$, and ξ_1, ξ_2 are independent, standard Gaussians. The solution u^* to (22) is $u^* = 0$ since $\mathbb{E}[h(\xi)] = 0$, and the SAA solution u_N^* corresponding to (22) is $u_N^* = (1/\alpha)\bar{\xi}_{1,N}\varphi_1 + (1/\alpha)\bar{\xi}_{2,N}\varphi_2$, where $\bar{\xi}_{j,N} = (1/N) \sum_{i=1}^{N} \xi_j^i$ for $j = 1, 2$. The ortho-normality of φ_1, φ_2 yields $\|u_N^*\|_{L^2(0,1)}^2 = (1/\alpha)^2(\bar{\xi}_{1,N})^2 + (1/\alpha)^2(\bar{\xi}_{2,N})^2$. Since $(1/\alpha)\bar{\xi}_{1,N}$ and $(1/\alpha)\bar{\xi}_{2,N}$ are independent, mean-zero Gaussian with variance $N^{-1}\alpha^{-2}$, the random variable $N\alpha^2\|u^* - u_N^*\|_{L^2(0,1)}$ has a chi-square distribution χ_2^2 with two degrees of freedom. Hence, for all $\varepsilon \ge 0$,

$$\text{Prob}(\|u^* - u_N^*\|_{L^2(0,1)} \ge \varepsilon) = \text{Prob}(\chi_2^2 \ge N\alpha^2\varepsilon^2) = e^{-N\alpha^2\varepsilon^2/2}. \tag{23}$$

Since $J(u, \xi) = (\alpha/2)\|u\|_{L^2(0,1)}^2 + (h(\xi), u)_{L^2(0,1)}$ and $F(u) = (\alpha/2)\|u\|_{L^2(0,1)}^2$, we find that $\|\nabla_u J(u^*, \xi) - \nabla F(u^*)\|_{L^2(0,1)}^2 = \|h(\xi)\|_{L^2(0,1)}^2$. Combined with $\|h(\xi)\|_{L^2(0,1)}^2 \sim \chi_2^2$, we obtain $\mathbb{E}[\exp(\tau^{-2}\|h(\xi)\|_{L^2(0,1)}^2)] = e$ for $\tau^2 = 2e/(e - 1)$. Our computations and

the tail bound (23) reveal that the exponential order of the tail bound in (17) is optimal up to the constant $3\tau^2/2 \approx 4.7$.

6 Application to linear-quadratic optimal control

We consider the linear-quadratic optimal control problem

$$\min_{u \in U} \{ (1/2)\mathbb{E}[\|QS(u,\xi) - y_d\|_H^2] + (\alpha/2)\|u\|_U^2 + \Psi(u) \}, \tag{24}$$

where $\alpha > 0$, $Q \in \mathscr{L}(Y,H)$, $y_d \in H$ and H is a real, separable Hilbert space. In this section, U and $\Psi : U \to \mathbb{R} \cup \{\infty\}$ fulfill Assumptions 1(a) and (b), respectively. The parameterized solution operator $S : U \times \Xi \to Y$ is defined as follows. For each $(u,\xi) \in U \times \Xi$, $S(u,\xi)$ is the solution to:

$$\text{find} \quad y \in Y : \quad A(\xi)y + B(\xi)u = g(\xi). \tag{25}$$

The spaces Y and Z are real, separable Banach spaces, $A : \Xi \to \mathscr{L}(Y,Z)$ and $B : \Xi \to \mathscr{L}(U,Z)$, $A(\xi)$ has a bounded inverse for each $\xi \in \Xi$, and $g : \Xi \to Z$.

We can model parameterized affine-linear elliptic and parabolic PDEs with (25), such as the heat equation with random inputs considered in [42, Sect. 3.1.2], and the elliptic PDEs with random inputs considered [19, 41, 59]. When $D \subset \mathbb{R}^d$ is a bounded domain and $U = L^2(D)$, a popular choice has been $\Psi(\cdot) = \gamma\|\cdot\|_{L^1(D)} + I_{U_{ad}}(\cdot)$ for $\gamma \geq 0$, where $U_{ad} \subset U$ is a nonempty, convex, closed set [23, 58]. Further nonsmooth regularizers are considered in [32, Sect. 4.7].

Defining $K(\xi) = -QA(\xi)^{-1}B(\xi)$ and $h(\xi) = QA^{-1}(\xi)g(\xi) - y_d$, the control problem (24) can be written as

$$\min_{u \in U} \{ (1/2)\mathbb{E}[\|K(\xi)u + h(\xi)\|_H^2] + (\alpha/2)\|u\|_U^2 + \Psi(u) \}. \tag{26}$$

We discuss differentiability and the lack of strong convexity of the expectation function $F_1 : U \to \mathbb{R} \cup \{\infty\}$ defined by

$$F_1(u) = (1/2)\mathbb{E}[\|K(\xi)u + h(\xi)\|_H^2]. \tag{27}$$

Assumption 3 The map $K : \Xi \to \mathscr{L}(U,H)$ is strongly measurable and $h : \Xi \to H$ is strongly measurable. For each $u \in U$, $\mathbb{E}[\|K(\xi)^*K(\xi)u\|_U] < \infty$, and $\mathbb{E}[\|h(\xi)\|_H^2]$, $\mathbb{E}[\|K(\xi)^*h(\xi)\|_U] < \infty$.

We define the integrand $J_1 : U \times \Xi \to \mathbb{R}$ by

$$J_1(u,\xi) = (1/2)\|K(\xi)u + h(\xi)\|_H^2. \tag{28}$$

Under the measurability conditions stated in Assumption 3, we can show that J_1 is a Carathéodory function.

Assumption 3 implies that the function F_1 defined in (27) is smooth.

Lemma 7 *If Assumption 3 holds, then F_1 defined in (27) is infinitely many times continuously differentiable, and for all $u, v \in U$,*

$$\nabla F_1(u) = \mathbb{E}[K(\xi)^*(K(\xi)u + h(\xi))] \quad \text{and} \quad \nabla^2 F_1(u)[v] = \mathbb{E}[K(\xi)^* K(\xi)v].$$

Proof The strong measurability of K implies that of $\xi \mapsto K(\xi)^*$ [31, Theorem 1.1.6] and hence that of $\xi \mapsto K(\xi)^* K(\xi)$ [31, Corollary 1.1.29]. Fix $u, v \in U$ and $\xi \in \Xi$. Since $\|K(\xi)u\|_H^2 \leq \|u\|_U \|K(\xi)^* K(\xi)u\|_U$ [37, p. 199], Assumption 3 ensures that F_1 is finite-valued.

Using (28), we find that $\nabla_u J_1(u, \xi) = K(\xi)^*(K(\xi)u + h(\xi))$ and $\nabla_{uu} J_1(u, \xi)[v] = K(\xi)^* K(\xi)v$. Combined with Assumption 3 and [23, Lemma C.3], we obtain that F_1 is Gâteaux differentiable with $\nabla F_1(u) = \mathbb{E}[\nabla_u J_1(u, \xi)]$. Since $\mathbb{E}[\|K(\xi)^* K(\xi)w\|_U] < \infty$ for all $w \in U$, $w \mapsto \mathbb{E}[K(\xi)^* K(\xi)w]$ is linear and bounded [27, Theorem 3.8.2]. Combined with the fact that $J_1(\cdot, \xi)$ is quadratic for all $\xi \in \Xi$, we conclude that F_1 is twice Gâteaux differentiable with $\nabla^2 F_1(u)[v] = \mathbb{E}[K(\xi)^* K(\xi)v]$ and hence infinitely many times continuously differentiable. □

The function F_1 defined in (27) lacks strong convexity under natural conditions; see Lemma 8. In this case, we may deduce that the strong convexity of the objective function of (24) solely comes from the function $(\alpha/2)\| \cdot \|_U^2 + \Psi(\cdot)$, and that the largest strong convexity parameter of $F(\cdot) = F_1(\cdot) + (\alpha/2)\| \cdot \|_U^2$ is $\alpha > 0$.

Assumption 4 The mapping $K : \Xi \to \mathcal{L}(U, H)$ is uniformly measurable, $\mathbb{E}[\|K(\xi)\|_{\mathcal{L}(U,H)}^2] < \infty$, and $K(\xi)$ is compact for all $\xi \in \Xi$. Moreover, the Hilbert space U is infinite-dimensional.

Lemma 8 *If Assumptions 3 and 4 hold, then the expectation function F_1 defined in (27) is not strongly convex.*

Proof We define $T : \Xi \to \mathcal{L}(U, U)$ by $T(\xi) = K(\xi)^* K(\xi)$. The uniform measurability of K implies that of $\xi \mapsto K(\xi)^*$ (cf. [9, Theorem 2.16] and [37, p. 200]) and hence that of T (cf. [31, pp. 12–13]). Since $K(\xi)$ is compact, $T(\xi)$ is compact [37, p. 427]. Moreover, we have $\mathbb{E}[\|T(\xi)\|_{\mathcal{L}(U,U)}] = \mathbb{E}[\|K(\xi)\|_{\mathcal{L}(U,H)}^2]$ [37, Theorem 3.9-4].

We show that $\mathbb{E}[T(\xi)]$ is a compact operator. Let $(v_k) \subset U$ be weakly converging to some $\bar{v} \in U$. Hence there exists $C \in (0, \infty)$ with $\|v_k\|_U \leq C$ for all $k \in \mathbb{N}$ [37, Theorem 4.8-3] which implies $\|T(\xi)v_k\|_U \leq C\|T(\xi)\|_{\mathcal{L}(U,U)}$ for each $\xi \in \Xi$ and $k \in \mathbb{N}$. Since $T(\xi)$ is compact for all $\xi \in \Xi$, we have for each $\xi \in \Xi$, $T(\xi)v_k \to T(\xi)\bar{v}$ as $k \to \infty$ [14, Proposition 3.3.3]. Combined with $\mathbb{E}[\|T(\xi)\|_{\mathcal{L}(U,U)}] < \infty$, the dominated convergence theorem [31, Proposition 1.2.5] yields $\mathbb{E}[T(\xi)v_k] \to \mathbb{E}[T(\xi)\bar{v}]$ as $k \to \infty$. We also have $\mathbb{E}[T(\xi)w] = \mathbb{E}[T(\xi)]w$ for all $w \in U$ [27, p. 85]. Thus $\mathbb{E}[T(\xi)]v_k \to \mathbb{E}[T(\xi)]\bar{v}$ as $k \to \infty$. Since U is reflexive and (v_k) is arbitrary, $\mathbb{E}[T(\xi)]$ is compact [14, Proposition 3.3.3].

Now, we show that F_1 is not strongly convex. Since U is infinite-dimensional, the self-adjoint, compact operator $\mathbb{E}[T(\xi)]$ lacks a bounded inverse [37, p. 428], [27, Theorem 3.8.1]. Hence it is noncoercive [10, Lemma 4.123]. Combined with

$\nabla^2 F_1(0) = \mathbb{E}[T(\xi)]$ (see Lemma 7 and [27, p. 85]), we conclude that F_1 is not strongly convex. □

The compactness of the Hessian of F_1 may also be studied using the theory on spectral decomposition of compact, self-adjoint operators [63, p. 159], or the results on the compactness of covariance operators [63, p. 174].

6.1 Examples

Many instances of the linear-quadratic control problem (24) frequently encountered in the literature are defined by the following data: $\alpha > 0$, $H = U$, Y is a real Hilbert space, $Q \in \mathscr{L}(Y, H)$ is the embedding operator of the compact embedding $Y \hookrightarrow H$, $B \in \mathscr{L}(U, Y^*)$ and $g : \Xi \to Y^*$ is essentially bounded. Moreover $A : \Xi \to \mathscr{L}(Y, Y^*)$ is uniformly measurable and there exist constants $0 < \kappa^*_{\min} \leq \kappa^*_{\max} < \infty$ with $\|A(\xi)\|_{\mathscr{L}(Y,Y^*)} \leq \kappa^*_{\max}$ and $\langle A(\xi)y, y \rangle_{Y^*,Y} \geq \kappa^*_{\min}\|y\|_Y^2$ for all $(y, \xi) \in Y \times \Xi$. The conditions imply that $A(\xi)$ has a bounded inverse for each $\xi \in \Xi$ [37, p. 101] and imply the existence of a solution to (24) when combined with Fatou's lemma; cf. [30, Theorem 1]. Moreover Assumptions 1–4 hold true.

We show that Assumption 2(b) is violated for the class of optimal control problems where the operator A is elliptic and defined by a log-normal random diffusion coefficient [2, 13]. Let Q and B be the embedding operators of the embeddings $H_0^1(0, 1) \hookrightarrow L^2(0, 1)$ and $L^2(0, 1) \hookrightarrow H^{-1}(0, 1)$, respectively. We choose $\Psi = 0$, $U = L^2(0, 1)$, $y_d(\cdot) = \sin(\pi \cdot)/\pi^2$, and $A(\xi) = e^{-\xi}\bar{A}$, where the weak Laplacian operator $\bar{A}$ is defined by $\langle \bar{A}y, v \rangle_{H^{-1}(0,1),H_0^1(0,1)} = (y', v')_{L^2(0,1)}$, and ξ is a standard Gaussian random variable. We have $\mathbb{E}[e^{2\xi}] = e^2$ and $\mathbb{E}[e^\xi] = e^{1/2}$. Since (π^2, y_d) is an eigenpair of $\bar{A}$, we find that $u^* = -\pi^2 e^{1/2}y_d/(e^2 + \pi^4\alpha)$ satisfies the sufficient optimality condition of (24), the normal equation $\alpha u^* + \mathbb{E}[e^{2\xi}]\bar{K}^*\bar{K}u^* = \mathbb{E}[e^\xi]\bar{K}^*y_d$, where $\bar{K} = -Q\bar{A}^{-1}B$. Hence u^* is the solution to (24) for the above data. Using the definition of J_1 provided in (28), we obtain

$$\nabla_u J_1(u^*, \xi) = \frac{e^\xi y_d}{\pi^2} - \frac{e^{1/2+2\xi} y_d}{\pi^2(e^2 + \pi^4\alpha)}.$$

For each $\xi \geq \ln(2(e^2 + \pi^4\alpha)) - 1/2$, $\|\nabla_u J_1(u^*, \xi)\|_{L^2(0,1)} \geq (e^\xi/\pi^2)\|y_d\|_{L^2(0,1)}$. Combined with $\nabla_u J(u^*, \xi) - \nabla F(u^*) = \nabla_u J_1(u^*, \xi) - \nabla F_1(u^*)$, $y_d \in L^2(0, 1)$, and $\mathbb{E}[\exp(s\xi^2/2)] = \infty$ for all $s \geq 1$ [11, p. 9], we conclude that Assumption 2(b) is violated.

7 Numerical illustration

We empirically verify the results derived in Theorem 3 for finite element discretizations of two linear-quadratic, elliptic optimal control problems, which are instances of (24).

For both examples, we consider $D = (0, 1)^2$, and the mapping Q in (24) is the embedding operator of the compact embedding $H_0^1(D) \hookrightarrow L^2(D)$. Moreover, we define $y_d \in L^2(D)$ by $y_d(x_1, x_2) = (1/6) \exp(2x_1) \sin(2\pi x_1) \sin(2\pi x_2)$ as in [8, p. 511]. For each $(u, \xi) \in L^2(D) \times \Xi$, $y(\xi) = S(u, \xi) \in H_0^1(D)$ solves the weak form of the linear elliptic PDE

$$-\nabla \cdot (\kappa(x, \xi) \nabla y(x, \xi)) = u(x) + r(x, \xi), \quad x \in D, \quad y(x, \xi) = 0, \quad x \in \partial D,$$

where ∂D is the boundary of the domain D. The set Ξ, the parameter $\alpha > 0$, the diffusion coefficient $\kappa : D \times \Xi \to (0, \infty)$ and the random right-hand side $r : D \times \Xi \to \mathbb{R}$ are defined in Examples 1 and 2. Defining $\langle Bu, v \rangle_{H^{-1}(D), H_0^1(D)} = -(u, v)_{L^2(D)}$, and

$$\langle A(\xi)y, v \rangle_{H^{-1}(D), H_0^1(D)} = \int_D \kappa(x, \xi) \nabla y(x) \cdot \nabla v(x) \mathrm{d}x,$$

$$\langle g(\xi), v \rangle_{H^{-1}(D), H_0^1(D)} = \int_D r(x, \xi) v(x) \mathrm{d}x,$$

the weak form of the linear PDE can be written in the form provided in (25).

We approximate the control problem (24) using a finite element discretization. The control space $U = L^2(D)$ is discretized using piecewise constant functions and the state space $Y = H_0^1(D)$ is discretized using piecewise linear continuous functions defined on a triangular mesh on $[0, 1]^2$ with $n \in \mathbb{N}$ being the number of cells in each direction, yielding finite element approximations of (24) and corresponding SAA problems. To simplify notation, we omit the index n when referring to the solutions to these optimization problems. The dimension of the discretized control space is $2n^2$.

Example 1 We define $\alpha = 10^{-3}$, $\Xi = [0.5, 3.5] \times [-1, 1]$, the random right-hand side $r(x, \xi) = \xi_2 \exp(2x_1) \sin(2\pi x_2)$, and $\kappa(\xi) = \xi_1$. The random variables ξ_1 and ξ_2 are independent, and ξ_1 has a truncated normal distribution supported on [0.5, 3.5] with mean 2 and standard deviation 0.25 (cf. [22, p. 2092]), and ξ_2 is uniformly distributed over $[-1, 1]$. We choose $\Psi(\cdot) = \gamma \| \cdot \|_{L^1(D)} + I_{U_{\mathrm{ad}}}(\cdot)$ with $\gamma = 5.5 \cdot 10^{-4}$ and $U_{\mathrm{ad}} = \{ u \in L^2(D) : -1 \leq u \leq 1 \}$, which is nonempty, closed, and convex [29, p. 56]. Furthermore, let $n = 256$.

Since $\kappa(\xi) = \xi_1$ is a real-valued random variable, we can evaluate $\nabla F_1(u)$ and its empirical mean using only two PDE solutions which can be shown by dividing (25) by $\kappa(\xi)$. It allows us to compute the solutions to the finite element approximation of (24) and to their SAA problems with moderate computational effort even though $n = 256$ is relatively large.

We solved the finite element discretization of (24) and the SAA problems using a semismooth Newton method [46, 58, 62] applied to a normal map (cf. [46, Eq. (3.3)]), which provides a reformulation of the first-order optimality conditions as a nonsmooth equation [46, Sect. 3.1]. The finite element discretization was performed using FEniCS [3, 39]. Sparse linear systems were solved using a direct method.
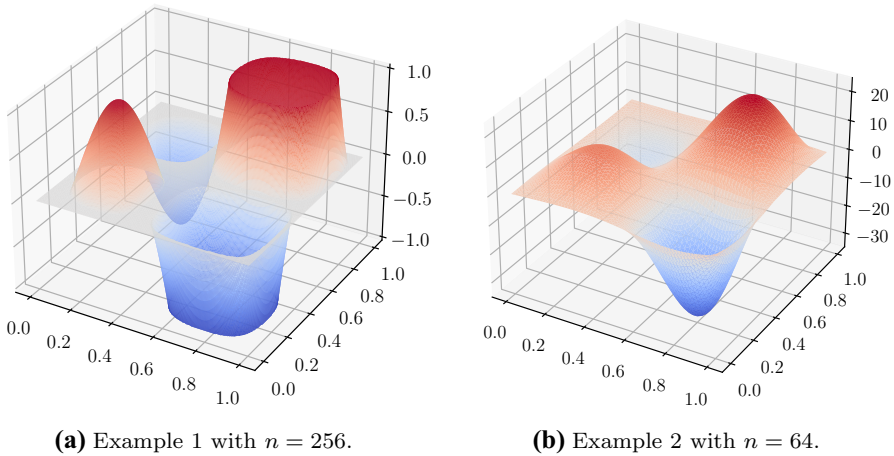
(a) Example 1 with $n = 256$.   **(b)** Example 2 with $n = 64$.

Fig. 1 Reference solutions

Example 2 We define $\alpha = 10^{-4}$, $\Xi = [3, 5] \times [0.5, 2.5]$ and the piecewise constant field κ by $\kappa(x, \xi) = \xi_1$ if $x \in (0, 1) \times (1/2, 1)$ and $\kappa(x, \xi) = \xi_2$ if $x \in (0, 1) \times (1/2, 1)$ (cf. [24, Example 3]). The random variables ξ_1 and ξ_2 are independent and uniformly distributed over $[3, 5]$ and $[0.5, 2.5]$, respectively. Moreover $r = 0$, $\Psi = 0$, and $n = 64$.

To obtain a deterministic reference solution to the finite element approximation of (24), we approximate the probability distribution of ξ by a discrete uniform distribution. It is supported on the grid points of a uniform mesh of Ξ using 50 grid points in each direction, yielding a discrete distribution with 2500 scenarios. Samples for the SAA problems are generated from this discrete distribution.

We used `dolfin-adjoint` [16, 18, 43] with `FEniCs` [3, 39] to evaluate the SAA objective functions and their derivatives, and solved the problems using `moola`'s `NewtonCG` method [18, 51].

Figure 1 depicts the reference solutions for Examples 1 and 2. To generate the surface plots depicted in Fig. 1, the piecewise constant reference solutions were interpolated to the space of piecewise linear continuous functions.

To illustrate the convergence rate $1/\sqrt{N}$ for $\mathbb{E}[\|u^* - u_N^*\|_U]$, we generated 50 independent samples of $\|u^* - u_N^*\|_U$ and computed the sample average. In order to empirically verify the exponential tail bound (17), we use the fact that it is equivalent to a certain bound on the Luxemburg norm of $u^* - u_N^*$. We define the Luxemburg norm $\| \cdot \|_{L_\phi(\Omega;U)}$ of a random vector $Z : \Omega \to U$ by

$$\|Z\|_{L_\phi(\Omega;U)} = \inf_{\nu > 0} \{ \nu : \mathbb{E}[\phi(\|Z\|_U/\nu)] \leq 1 \}, \tag{29}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is given by $\phi(x) = \exp(x^2) - 1$, and $L_\phi(\Omega;U) = L_{\phi(\|\cdot\|_U)}(\Omega;U)$ is the Orlicz space consisting of each random vector $Z : \Omega \to U$ such that there exists

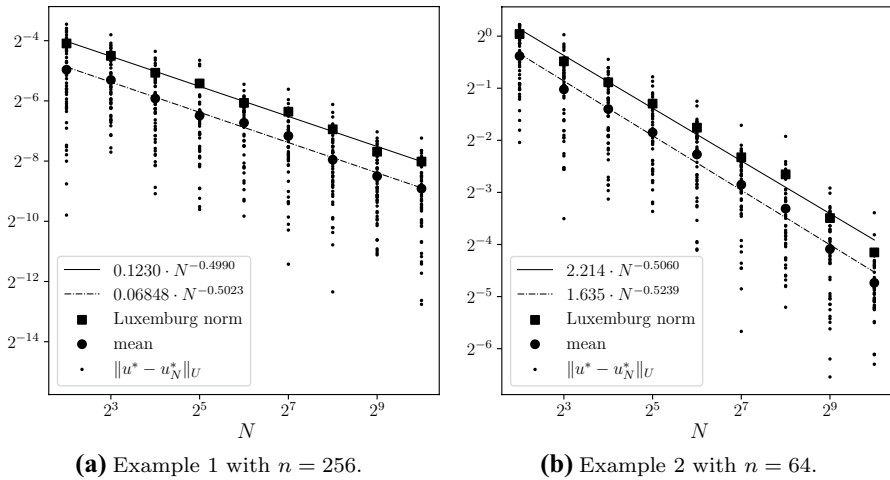**(a)** Example 1 with $n = 256$. **(b)** Example 2 with $n = 64$.

Fig. 2 For each example, 50 independent realizations of $\|u^* - u_N^*\|_U$, and the empirical mean error and empirical Luxemburg norm. The convergence rates were computed using least squares

$\nu > 0$ with $\mathbb{E}[\phi(\|Z\|_U/\nu)] < \infty$; cf. [33, Sect. 6.2]. The exponential tail bound (17) implies

$$\|u^* - u_N^*\|_{L_\phi(\Omega;U)} \leq \frac{3\sqrt{3}\tau}{\alpha\sqrt{N}}, \tag{30}$$

and (30) ensures $\text{Prob}(\|u^* - u_N^*\|_U \geq \varepsilon) \leq 2e^{-\tau^{-2}N\varepsilon^2\alpha^2/27}$ for all $\varepsilon > 0$. These two statements follow from [11, Theorem 3.4 on p. 56] when applied to the real-valued random variable $\|u^* - u_N^*\|_U$. To empirically verify the convergence rate $1/\sqrt{N}$ for $\|u^* - u_N^*\|_{L_\phi(\Omega;U)}$, we approximated the expectation in (29) using the same samples used to estimate $\mathbb{E}[\|u^* - u_N^*\|_U]$.

Figure 2 depicts 50 realizations of the errors $\|u^* - u_N^*\|_U$, the empirical approximations of $\mathbb{E}[\|u^* - u_N^*\|_U]$ and of the Luxemburg norm $\|u^* - u_N^*\|_{L_\phi(\Omega;U)}$ as well as the corresponding convergence rates. The rates were computed using least squares. The empirical convergences rates depicted in Fig. 2 are close to the theoretical rate $1/\sqrt{N}$ for $\mathbb{E}[\|u^* - u_N^*\|_U]$ and $\|u^* - u_N^*\|_{L_\phi(\Omega;U)}$; see (16) and (30).

8 Discussion

We have considered convex stochastic programs posed in Hilbert spaces where the integrand is strongly convex with the same parameter for each random element's realization. We have established exponential tail bounds for the distance between SAA solutions and the true ones. For this problem class, tail bounds are optimal up to problem-independent, moderate constants. We have applied our findings to stochastic linear-quadratic control problems, a subclass of the above problem class.

We conclude the paper by illustrating that the "dynamics" of finite- and infinite-dimensional stochastic programs can be quite different. We consider

$$\min_{\|x\|_2 \leq 1} \mathbb{E}[\|x\|_2^2 - 2x^T\zeta], \tag{31}$$

where ζ is an $\mathbb{R}^n$-valued, mean-zero Gaussian random vector with covariance matrix $\sigma^2 I$ and $\sigma^2 > 0$. This corresponds to the choice $m = 1$ in [55, Example 1]. For $\delta \in (0, 0.3)$ and $\varepsilon \in (0, 1)$, at least $N > n\sigma^2/\varepsilon = \mathbb{E}[\|\zeta\|_2^2]/\varepsilon$ samples are required for the corresponding SAA problem's optimal solution to be an ε-optimal solution to (31) with a probability of at least $1 - \delta$ [55, Example 1].

The infinite-dimensional analogue of (31) is given by

$$\min_{\|u\|_{\ell^2(\mathbb{N})} \leq 1} \mathbb{E}[\|u\|_{\ell^2(\mathbb{N})}^2 - 2(u, \xi)_{\ell^2(\mathbb{N})}], \tag{32}$$

where ξ is an $\ell^2(\mathbb{N})$-valued, mean-zero Gaussian random vector, and $\ell^2(\mathbb{N})$ is the standard sequence space. For each $\varepsilon \in (0, 1)$, the SAA solution u_N^* corresponding to (32) is an ε-optimal solution to (32) if and only if we have $\|(1/N)\sum_{i=1}^N \xi^i\|_{\ell^2(\mathbb{N})}^2 \leq \varepsilon$. Combined with Remark 4 in [48], we find that $N \geq (3/\varepsilon)\ln(2/\delta)\mathbb{E}[\|\xi\|_{\ell^2(\mathbb{N})}^2]$ samples are sufficient in order for u_N^* to be an ε-optimal solution to (32), with a probability of at least $1 - \delta \in (0, 1)$.

Let us compare the stochastic program (31) with (32). Whereas $\mathbb{E}[\|\zeta\|_2^2] = n\sigma^2 \to \infty$ as $n \to \infty$ and $\mathbb{E}[|\zeta_k|^2] = \sigma^2$ $(1 \leq k \leq n)$, we have $\mathbb{E}[\|\xi\|_{\ell^2(\mathbb{N})}^2] < \infty$ due to the Landau–Shepp–Fernique theorem and $\mathbb{E}[|\xi_k|^2] \to 0$ as $k \to \infty$ [64, p. 59]. We find that the "overall level-of-randomness" for the finite-dimensional problem (31) depends on its dimension n, while that for the infinite-dimensional analogue (32) is fixed.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

References

1. Agarwal, A., Bartlett, P.L., Ravikumar, P., Wainwright, M.J.: Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. IEEE Trans. Inf. Theory **58**(5), 3235–3249 (2012). https://doi.org/10.1109/TIT.2011.2182178

2. Ali, A.A., Ullmann, E., Hinze, M.: Multilevel Monte Carlo analysis for optimal control of elliptic PDEs with random coefficients. SIAM/ASA J. Uncertain. Quantif. **5**(1), 466–492 (2017). https://doi.org/10.1137/16M109870X

3. Alnæs, M.S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M.E., Wells, G.N.: The FEniCS project version 1.5. Arch. Numer. Software **3**(100), 9–23 (2015). https://doi.org/10.11588/ans.2015.100.20553

4. Artstein, Z., Wets, R.J.B.: Consistency of minimizers and the SLLN for stochastic programs. J. Convex Anal. **2**(1/2), 1–17 (1995)

5. Aubin, J.P., Frankowska, H.: Set-Valued Analysis. Mod. Birkhäuser Class. Springer, Boston (2009). https://doi.org/10.1007/978-0-8176-4848-0

6. Bach, F., Moulines, E.: Non-asymptotic analysis of stochastic approximation algorithms for machine learning. In: Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger K.Q. (eds.) Advances in neural information processing systems, pp. 451–459. Curran Associates, Inc., Red Hook (2011). https://papers.nips.cc/paper/4316-non-asymptotic-analysis-of-stochastic-approximation-algorithms-for-machine-learning

7. Banholzer, D., Fliege, J., Werner, R.: On rates of convergence for sample average approximations in the almost sure sense and in mean. Math. Program. (2019). https://doi.org/10.1007/s10107-019-01400-4

8. Bergounioux, M., Haddou, M., Hintermüller, M., Kunisch, K.: A comparison of a Moreau-Yosida-based active set strategy and interior point methods for constrained optimal control problems. SIAM J. Optim. **11**(2), 495–521 (2000). https://doi.org/10.1137/S1052623498343131

9. Bharucha-Reid, A.T.: Random Integral Equations. Math. Sci. Eng., vol. 96. Academic Press, New York (1972)

10. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Ser. Oper. Res. Springer, New York (2000). https://doi.org/10.1007/978-1-4612-1394-9

11. Buldygin, V.V., Kozachenko, Yu, V.: Metric Characterization of Random Variables and Random Processes. Transl. Math. Monogr., vol. 188. American Mathematical Society, Providence (2000)

12. Castaing, C., Valadier, M.: Convex Analysis and Measurable Multifunctions. Lecture Notes in Math., vol. 580. Springer, Berlin (1977)

13. Charrier, J.: Strong and weak error estimates for elliptic partial differential equations with random coefficients. SIAM J. Numer. Anal. **50**(1), 216–246 (2012). https://doi.org/10.1137/100800531

14. Conway, J.B.: A Course in Functional Analysis. Grad. Texts in Math., 2nd edn., vol. 96. Springer, New York (1990). https://doi.org/10.1007/978-1-4757-4383-8

15. Duchi, J.C., Bartlett, P.L., Wainwright, M.J.: Randomized smoothing for stochastic optimization. SIAM J. Optim. **22**(2), 674–701 (2012). https://doi.org/10.1137/110831659

16. Farrell, P.E., Ham, D.A., Funke, S.W., Rognes, M.E.: Automated derivation of the adjoint of high-level transient finite element programs. SIAM J. Sci. Comput. **35**(4), C369–C393 (2013). https://doi.org/10.1137/120873558

17. Fukuda, R.: Exponential integrability of sub-Gaussian vectors. Prob. Theory Relat. Fields **85**(4), 505–521 (1990). https://doi.org/10.1007/BF01203168

18. Funke, S.W., Farrell, P.E.: A framework for automated PDE-constrained optimisation (2013). arXiv:1302.3894

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

20. Garreis, S., Ulbrich, M.: Constrained optimization with low-rank tensors and applications to parametric problems with PDEs. SIAM J. Sci. Comput. **39**(1), A25–A54 (2017). https://doi.org/10.1137/16M1057607

21. Ge, L., Wang, L., Chang, Y.: A sparse grid stochastic collocation upwind finite volume element method for the constrained optimal control problem governed by random convection diffusion equations. J. Sci. Comput. **77**(1), 524–551 (2018). https://doi.org/10.1007/s10915-018-0713-y

22. Geiersbach, C., Pflug, GCh.: Projected stochastic gradients for convex constrained problems in Hilbert spaces. SIAM J. Optim. **29**(3), 2079–2099 (2019). https://doi.org/10.1137/18M1200208

23. Geiersbach, C., Scarinci, T.: Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces. Comput. Optim. Appl. **78**(3), 705–740 (2021). https://doi.org/10.1007/s10589-020-00259-y

24. Geiersbach, C., Wollner, W.: A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty. SIAM J. Sci. Comput. **42**(5), A2750–A2772 (2020). https://doi.org/10.1137/19M1263297

25. Guigues, V., Juditsky, A., Nemirovski, A.: Non-asymptotic confidence bounds for the optimal value of a stochastic program. Optim. Methods Softw. **32**(5), 1033–1058 (2017). https://doi.org/10.1080/10556788.2017.1350177

26. Guth, P.A., Kaarnioja, V., Kuo, F.Y., Schillings, C., Sloan, I.H.: A quasi-Monte Carlo method for optimal control under uncertainty. SIAM/ASA J. Uncertain. Quantif. **9**(2), 354–383 (2021). https://doi.org/10.1137/19M1294952

27. Hille, E., Phillips, R.S.: Functional Analysis and Semi-Groups. Colloq. Publ., vol. 31. American Mathematical Society, Providence (1957)

28. Hinze, M.: A variational discretization concept in control constrained optimization: the linear-quadratic case. Comput. Optim. Appl. **30**(1), 45–61 (2005). https://doi.org/10.1007/s10589-005-4559-5

29. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE Constraints. Math. Model. Theory Appl., vol. 23. Springer, Dordrecht (2009). https://doi.org/10.1007/978-1-4020-8839-1

30. Hoffhues, M., Römisch, W., Surowiec, T.M.: On quantitative stability in infinite-dimensional optimization under uncertainty. Optim. Lett. **15**(8), 2733–2756 (2021). https://doi.org/10.1007/s11590-021-01707-2

31. Hytönen, T., van Neerven, J., Veraar, M., Weis, L.: Analysis in Banach Spaces: Martingales and Littlewood–Paley Theory. Ergeb. Math. Grenzgeb. (3), vol. 63. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48520-1

32. Ito, K., Kunisch, K.: Lagrange Multiplier Approach to Variational Problems and Applications. Adv. Des. Control, vol. 15. SIAM, Philadelphia (2008). https://doi.org/10.1137/1.9780898718614

33. Kosmol, P., Müller-Wichards, D.: Optimization in Function Spaces: With Stability Considerations in Orlicz Spaces. De Gruyter Ser. Nonlinear Anal. Appl., vol. 13. De Gruyter, Berlin (2011). https://doi.org/10.1515/9783110250213

34. Kouri, D.P., Heinkenschloss, M., Ridzal, D., van Bloemen Waanders, B.: A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. SIAM J. Sci. Comput. **35**(4), A1847–A1879 (2013). https://doi.org/10.1137/120892362

35. Kouri, D.P., Shapiro, A.: Optimization of PDEs with uncertain inputs. In: Antil, H., Kouri, D.P., Lacasse, M.D., Ridzal, D. (eds.) Frontiers in PDE-Constrained Optimization, IMA Vol. Math. Appl., vol. 163, pp. 41–81. Springer, New York (2018). https://doi.org/10.1007/978-1-4939-8636-1_2

36. Kouri, D.P., Surowiec, T.M.: Existence and optimality conditions for risk-averse PDE-constrained optimization. SIAM/ASA J. Uncertain. Quantif. **6**(2), 787–815 (2018). https://doi.org/10.1137/16M1086613

37. Kreyszig, E.: Introductory Functional Analysis with Applications. Wiley, New York (1978)

38. Lan, G.: First-order and Stochastic Optimization Methods for Machine Learning. Springer Ser. Data Sci. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39568-1

39. Logg, A., Mardal, K.A., Wells, G.N. (eds.): Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book. Lect. Notes Comput. Sci. Eng., vol. 84. Springer, Berlin (2012). https://doi.org/10.1007/978-3-642-23099-8

40. Marín, F.J., Martínez-Frutos, J., Periago, F.: Control of random PDEs: an overview. In: Doubova, A., González-Burgos, M., Guillén-González, F., Marín Beltrán, M. (eds.) Recent advances in PDEs: analysis, numerics and control: in honor of Prof. Fernández-Cara's 60th Birthday, pp. 193–210. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-97613-6_10

41. Martin, M., Krumscheid, S., Nobile, F.: Complexity analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters. ESAIM Math. Model. Numer. Anal. **55**(4), 1599–1633 (2021). https://doi.org/10.1051/m2an/2021025

42. Martínez-Frutos, J., Esparza, F.P.: Optimal Control of PDEs under Uncertainty: An Introduction with Application to Optimal Shape Design of Structures. Springer Briefs Math. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98210-6

43. Mitusch, S.K., Funke, S.W., Dokken, J.S.: dolfin-adjoint 2018.1: automated adjoints for FEniCS and Firedrake. J. Open Source Softw. **4**(38), 1292 (2019). https://doi.org/10.21105/joss.01292

44. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM J. Optim. **19**(4), 1574–1609 (2009). https://doi.org/10.1137/070704277

45. Nemirovsky, A.S., Yudin, D.B.: Problem Complexity and Method Efficiency in Optimization. Wiley-Interscience Series in Discrete Mathematics. Wiley, Chichester (1983). Translated by E. R. Dawson

46. Pieper, K.: Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems. Dissertation, Technische Universität München, München (2015). http://nbn-resolving.de/urn:nbn:de:bvb:91-diss-20150420-1241413-1-4

47. Pinelis, I.: Optimum bounds for the distributions of martingales in Banach spaces. Ann. Probab. **22**(4), 1679–1706 (1994). https://doi.org/10.1214/aop/1176988477

48. Pinelis, I.F., Sakhanenko, A.I.: Remarks on inequalities for large deviation probabilities. Theory Probab. Appl. **30**(1), 143–148 (1986). https://doi.org/10.1137/1130013

49. Römisch, W., Surowiec, T.M.: Asymptotic properties of Monte Carlo methods in elliptic PDE-constrained optimization under uncertainty (2021). arXiv:2106.06347

50. Royset, J.O.: Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. Math. Program. **184**, 289–318 (2020). https://doi.org/10.1007/s10107-019-01413-z

51. Schwedes, T., Ham, D.A., Funke, S.W., Piggott, M.D.: Mesh Dependence in PDE-Constrained Optimisation. SpringerBriefs Math. Planet Earth. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59483-5

52. Shalev-Shwartz, S., Shamir, O., Srebro, N., Sridharan, K.: Learnability, stability and uniform convergence. J. Mach. Learn. Res. **11**, 2635–2670 (2010)

53. Shapiro, A.: Asymptotic behavior of optimal solutions in stochastic programming. Math. Oper. Res. **18**(4), 829–845 (1993). https://doi.org/10.1287/moor.18.4.829

54. Shapiro, A.: Monte Carlo sampling methods. In: Stochastic Programming, Handbooks in Oper. Res. Manag. Sci., vol. 10, pp. 353–425. Elsevier (2003). https://doi.org/10.1016/S0927-0507(03)10006-0

55. Shapiro, A.: Stochastic programming approach to optimization under uncertainty. Math. Program. **112**(1), 183–220 (2008). https://doi.org/10.1007/s10107-006-0090-4

56. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory. MOS-SIAM Ser. Optim., 2nd edn. SIAM, Philadelphia (2014). https://doi.org/10.1137/1.9781611973433

57. Shapiro, A., Nemirovski, A.: On complexity of stochastic programming problems. In: Jeyakumar, V., Rubinov, A. (eds.) Continuous Optimization: Current Trends and Modern Applications, Appl. Optim., vol. 99, pp. 111–146. Springer, Boston (2005). https://doi.org/10.1007/0-387-26771-9_4

58. Stadler, G.: Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices. Comput. Optim. Appl. **44**(2), 159–181 (2009). https://doi.org/10.1007/s10589-007-9150-9

59. Sun, T., Shen, W., Gong, B., Liu, W.: A priori error estimate of stochastic Galerkin method for optimal control problem governed by stochastic elliptic PDE with constrained control. J. Sci. Comput. **67**(2), 405–431 (2015). https://doi.org/10.1007/s10915-015-0091-7

60. Tröltzsch, F.: On finite element error estimates for optimal control problems with elliptic PDEs. In: Lirkov, I., Margenov, S., Waśniewski, J. (eds.) Large-Scale Scientific Computing, Lecture Notes in Comput. Sci., vol. 5910, pp. 40–53. Springer, Berlin (2010). https://doi.org/10.1007/978-3-642-12535-5_4

61. Tsybakov, A.B.: Error bounds for the method of minimization of empirical risk. Probl. Peredachi Inf. **17**, 50–61 (1981). (in Russian). http://mi.mathnet.ru/ppi1380

62. Ulbrich, M.: Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces. MOS-SIAM Ser. Optim. SIAM, Philadelphia (2011). https://doi.org/10.1137/1.9781611970692

63. Vakhania, N.N., Tarieladze, V.I., Chobanyan, S.A.: Probability Distributions on Banach spaces. Math. Appl. (Soviet Ser.), vol. 14. D. Reidel Publishing Co., Dordrecht (1987). https://doi.org/10.1007/978-94-009-3873-1. Translated from the Russian and with a preface by Wojbor A. Woyczynski
64. Yurinsky, V.: Sums and Gaussian Vectors. Lecture Notes in Math, vol. 1617. Springer, Berlin (1995)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.