



Consistency of Monte Carlo Estimators for Risk-Neutral PDE-Constrained Optimization

Johannes Milz¹ 

Accepted: 6 January 2023 / Published online: 7 April 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

We apply the sample average approximation (SAA) method to risk-neutral optimization problems governed by nonlinear partial differential equations (PDEs) with random inputs. We analyze the consistency of the SAA optimal values and SAA solutions. Our analysis exploits problem structure in PDE-constrained optimization problems, allowing us to construct deterministic, compact subsets of the feasible set that contain the solutions to the risk-neutral problem and eventually those to the SAA problems. The construction is used to study the consistency using results established in the literature on stochastic programming. The assumptions of our framework are verified on three nonlinear optimization problems under uncertainty.

Keywords Stochastic programming · Monte Carlo sampling · Sample average approximation · Optimization under uncertainty · PDE-constrained optimization

Mathematics Subject Classification 65C05 · 90C15 · 35R60 · 90C48 · 90C30 · 60H25

1 Introduction

Advances in areas such as computational science and engineering, applied mathematics, software design, and scientific computing have allowed decision makers to optimize complex physics-based systems under uncertainty, such as those modeled using partial differential equations (PDEs) with uncertain inputs. Recent applications in the field of PDE-constrained optimization under uncertainty are, for example, oil field development [60], stellarator coil optimization [81], acoustic wave propagation [82], and shape optimization of electrical engines [39]. In the literature on optimization under uncertainty, several approaches have been proposed for obtaining decisions that

✉ Johannes Milz
milz@ma.tum.de

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

are resilient to uncertainty, such as robust optimization [11] and stochastic optimization [70]. When the parameter vector is modeled as a random vector with known probability distribution, a common approach is to seek decisions that minimize the expected value of a parameterized objective function. The resulting optimization problem is referred to as risk-neutral optimization problem. However, evaluating the risk-neutral problem's objective function would require computing a potentially high-dimensional integral. Furthermore, each evaluation of the parameterized objective function may require the simulation of complex systems of PDEs, adding another challenge to obtaining solutions to risk-neutral PDE-constrained optimization problems.

A common approach for approximating risk-neutral optimization problems is the sample average approximation (SAA) method, yielding the SAA problem. For example, the SAA approach is used in the literature on mathematical programming [38, 70, 73] and on PDE-constrained optimization [29, 43, 66, 81]. The SAA problem's objective function is the sample average of the parameterized objective function computed using samples of the random vector. To assess the quality of the SAA solutions as approximate solutions to the risk-neutral problem, different error measures have been considered, such as the consistency of the SAA optimal value and of SAA solutions [7, 30, 50, 68, 69, 72], nonasymptotic sample size estimates [18, 67, 70, 71, 73], mean and almost sure convergence rates [9], and confidence intervals for SAA optimal values [27].

A number of results on the SAA approach are based on the compactness of either the feasible set or of sets that eventually contain the SAA solutions, such as the consistency properties of SAA solutions and sample size estimates. The analysis of the SAA approach as applied to PDE-constrained optimization problems is complicated by the fact that the feasible sets are commonly noncompact, such as the set of square integrable functions defined on the interval $(0, 1)$ with values in $[-1, 1]$. Moreover, level sets of the SAA objective function may not be contained in a deterministic, compact set as shown in Appendix A. Our approach for establishing consistency is based on that developed in [72, Chap. 5]. While the consistency results in [72, Chap. 5] are established for finite dimensional stochastic programs, the results do not require the compactness of the feasible set. Instead, they are valid provided that the solution set to the stochastic program and those to the SAA problems are eventually contained in a deterministic, compact set.

We establish the consistency of SAA optimal values and SAA solutions to risk-neutral nonlinear PDE-constrained optimization problems. For analyzing the SAA approach, we construct deterministic, compact subsets of the possibly infinite dimensional feasible sets that contain the solutions to risk-neutral PDE-constrained problems and eventually those to the corresponding SAA problems. This observation allows us to study the consistency using the tools developed in the literature on M-estimation [34, 52] and stochastic programming [70, 72]. Our consistency results are inspired by and based on those established in [70, Sects. 2 and 7] and [72, Chap. 5]. For our construction of these compact sets, we use the fact that many PDE-constrained optimization problems involve compact operators, such as compact embeddings. Moreover, we use first-order optimality conditions and PDE stability estimates. The construction is partly inspired by the computations used to establish higher regularity of solutions to deterministic PDE-constrained optimization problems [56, p. 1305], [76, Sect. 2.15]

and a computation made in the author's dissertation [57, Sect. 3.5] which demonstrates that all SAA solutions to certain linear elliptic optimal control problems are contained in a compact set.

The SAA method as applied to risk-neutral strongly convex PDE-constrained optimization has recently been analyzed in [33, 54, 58, 66]. The authors of [62] apply the SAA scheme to the optimal control of ordinary differential equations with random inputs and demonstrate the epiconvergence of the SAA objective function and the weak consistency of SAA critical points in the sense defined in [64, Definition 3.3.6]. The weak consistency implies that accumulation points of SAA critical points are critical points of the optimal control problem [62, p. 13].

Monte Carlo sampling is one approach to approximating expected values in stochastic program's objective functions. For strongly convex elliptic PDE-constrained optimization problems, quasi-Monte Carlo techniques are analyzed in [28]. Further discretization approaches for expectations are, for example, stochastic collocation [74] and low-rank tensor approximations [23]. Besides risk-neutral PDE-constrained optimization, risk-averse PDE-constrained optimization [3, 17, 43, 44, 55], distributionally robust PDE-constrained optimization [41, 59], robust PDE-constrained optimization [5, 39, 51], and PDE-constrained optimization with chance constraints [16, 20, 21, 26, 75] provide approaches to decision making under uncertainty with PDEs.

1.1 Outline

We introduce notation in Sect. 2 and a class of risk-neutral nonlinear PDE-constrained optimization problems and their SAA problems in Sect. 3. Section 3.1 presents a compact subset that contains the solutions to the risk-neutral problem and eventually those to its SAA problems. We study the consistency of SAA optimal values and solutions in Sect. 3.2. Section 4 discusses the application of our theory to three nonlinear PDE-constrained optimization problems under uncertainty. We summarize our contributions, and discuss some limitations of our approach and open research questions in Sect. 5.

2 Notation and Preliminaries

Throughout the paper, the control space U is a real, separable Hilbert space and is identified with its dual, that is, we omit writing the Riesz mapping.

Metric spaces are defined over the real numbers and equipped with their Borel sigma-algebra. We abbreviate "with probability one" by w.p. 1. Let $(\Theta, \mathcal{A}, \mu)$ be probability space. For two complete metric spaces Λ_1 and Λ_2 , a mapping $G : \Lambda_1 \times \Theta \rightarrow \Lambda_2$ is a Carathéodory mapping if $G(\cdot, \theta)$ is continuous for all $\theta \in \Theta$ and $G(v, \cdot)$ is measurable for each $v \in \Lambda_1$. Let Λ be a Banach space. For each $N \in \mathbb{N}$, let $\Upsilon_N : \Theta \rightrightarrows \Lambda$ be a set-valued mapping and let $\Psi \subset \Lambda$ be a set. We say that w.p. 1 for all sufficiently large N , $\Upsilon_N \subset \Psi$ if the set $\{\theta \in \Theta : \exists n(\theta) \in \mathbb{N} \forall N \geq n(\theta) : \Upsilon_N(\theta) \subset \Psi\}$ is contained in \mathcal{A} and occurs w.p. 1, that is, if the limit inferior of the sequence $(\{\theta \in \Theta : \Upsilon_N(\theta) \subset \Psi\})_N$ is contained in \mathcal{A} and occurs w.p. 1 [12,

p. 55]. A mapping $v : \Theta \rightarrow \Lambda$ is strongly measurable if there exists a sequence of simple mappings $v_k : \Theta \rightarrow \Lambda$ such that $v_k(\theta) \rightarrow v(\theta)$ as $k \rightarrow \infty$ for all $\theta \in \Theta$ [35, Definition 1.1.4]. If Λ is separable, then $v : \Theta \rightarrow \Lambda$ is strongly measurable if and only if it is measurable [35, Corollary 1.1.2 and Theorem 1.1.6]. The dual to a Banach space Λ is Λ^* and the norm of Λ is denoted by $\|\cdot\|_\Lambda$. We use $\langle \cdot, \cdot \rangle_{\Lambda^*, \Lambda}$ to denote the dual pairing between Λ^* and Λ . If Λ is a reflexive Banach space, we identify $(\Lambda^*)^*$ with Λ and write $(\Lambda^*)^* = \Lambda$. Let Λ_1 and Λ_2 be real Banach spaces. A linear operator $\Upsilon : \Lambda_1 \rightarrow \Lambda_2$ is compact if the image $\Upsilon(\Lambda_0)$ is precompact in Λ_2 for each bounded set $\Lambda_0 \subset \Lambda_1$ [49, Definition 8.1-1]. The operator $\Upsilon^* : \Lambda_2^* \rightarrow \Lambda_1^*$ is the (Banach space-)adjoint operator of the linear, bounded mapping $\Upsilon : \Lambda_1 \rightarrow \Lambda_2$ and is defined by $\langle \Upsilon^* v_2, v_1 \rangle_{\Lambda_1^*, \Lambda_1} = \langle v_2, \Upsilon v_1 \rangle_{\Lambda_2^*, \Lambda_2}$ [49, Definition 4.5-1]. We use $\Lambda_1 \hookrightarrow \Lambda_2$ to denote a continuous embedding from Λ_1 to Λ_2 , that is, $\Lambda_1 \subset \Lambda_2$ and the embedding operator $\iota : \Lambda_1 \rightarrow \Lambda_2$ defined by $\iota[v] = v$ is continuous [65, Definition 7.15 and Rem. 7.17]. A continuous embedding is compact if the embedding operator is a compact operator [65, Definition 7.25 and Lemma 8.75]. We denote by Df the Fréchet derivative of f , and use the notation $D_x f$ and f_x for partial derivatives with respect to x . Throughout the text, $D \subset \mathbb{R}^d$ is a bounded domain. For $p \in [1, \infty)$, we denote by $L^p(D)$ the Lebesgue space of p -integrable functions defined on D and $L^\infty(D)$ that of essentially bounded functions. The space $H^1(D)$ is the space of all $v \in L^2(D)$ with weak derivatives contained in $L^2(D)^d$, where $L^2(D)^d$ is the Cartesian product of $L^2(D)$ taken d times. We equip $H^1(D)$ with the norm $\|y\|_{H^1(D)} = (\|y\|_{L^2(D)}^2 + \|\nabla y\|_{L^2(D)^d}^2)^{1/2}$. The Hilbert space $H_0^1(D)$ consists of all $v \in H^1(D)$ with zero boundary traces and is equipped with the norm $\|y\|_{H_0^1(D)} = \|\nabla y\|_{L^2(D)^d}$. We define $H^{-1}(D) = H_0^1(D)^*$. We define Friedrichs' constant $C_D \in (0, \infty)$ by $C_D = \sup_{v \in H_0^1(D) \setminus \{0\}} \|v\|_{L^2(D)} / \|v\|_{H_0^1(D)}$. The indicator function $I_{U_0} : U \rightarrow [0, \infty]$ of $U_0 \subset U$ is given by $I_{U_0}(v) = 0$ if $v \in U_0$ and $I_{U_0}(v) = \infty$ otherwise. For a convex, lower semicontinuous, proper function $\chi : U \rightarrow (-\infty, \infty]$, the proximity operator $\text{prox}_\chi : U \rightarrow U$ of χ is defined by (see [10, Definition 12.23])

$$\text{prox}_\chi(v) = \arg \min_{w \in U} \chi(w) + (1/2) \|v - w\|_U^2.$$

3 Risk-Neutral PDE-Constrained Optimization Problem

We consider the risk-neutral PDE-constrained optimization problem

$$\min_{u \in U} \mathbb{E}[J_1(S(u, \xi), \xi)] + \psi(u) + (\alpha/2) \|u\|_U^2 \tag{1}$$

and its sample average approximation

$$\min_{u \in U} \frac{1}{N} \sum_{i=1}^N J_1(S(u, \xi^i), \xi^i) + \psi(u) + (\alpha/2) \|u\|_U^2, \tag{2}$$

where $\alpha > 0$, and ξ^1, ξ^2, \dots are independent identically distributed Ξ -valued random elements defined on a complete probability space (Ω, \mathcal{F}, P) and each ξ^i has the same distribution as that of the random element ξ . Here, ξ maps from a probability space to a complete probability space with sample space Ξ being a complete, separable metric space. We state assumptions on the mappings $J_1 : Y \times \Xi \rightarrow [0, \infty)$, $\psi : U \rightarrow [0, \infty)$, and $S : U \times \Xi \rightarrow Y$ as well as on the control space U and state space Y in Assumptions 1 and 2. Let $F : U \rightarrow [0, \infty]$ be the objective function of (1) and let $\hat{F}_N : U \rightarrow [0, \infty]$ be that of (2). Since ξ^1, ξ^2, \dots are defined on the common probability space (Ω, \mathcal{F}, P) , we can view the function \hat{F}_N as defined on $U \times \Omega$. However, we often omit writing the second argument. We often use ξ to denote a deterministic element in Ξ .

In the remainder of the section, we impose conditions on the optimization problem (1). Assumptions 1 and 2 ensure that the reduced formulation of the risk-neutral problem (1) and its SAA problem (2) are well-defined.

- Assumption 1** (a) The space U is a real, separable Hilbert space, and Y is a real, separable Banach space.
 (b) The function $J_1 : Y \times \Xi \rightarrow [0, \infty)$ is a Carathéodory function, and $J_1(\cdot, \xi)$ is continuously differentiable for all $\xi \in \Xi$.
 (c) The regularization parameter α is positive, and $\psi : U \rightarrow [0, \infty]$ is proper, convex and lower semicontinuous.

The nonnegativity of J_1 and ψ is fulfilled for many PDE-constrained optimization problems (see Sect. 4). We define the feasible set

$$U_{\text{ad}} = \{u \in U : \psi(u) < \infty\}. \tag{3}$$

- Assumption 2** (a) The operator $E : (Y \times U) \times \Xi \rightarrow Z$ is a Carathéodory mapping, $E(\cdot, \cdot, \xi)$ is continuously differentiable for all $\xi \in \Xi$, and Z is a real, separable Banach space.
 (b) For each $(u, \xi) \in U \times \Xi$, $S(u, \xi) \in Y$ is the unique solution to: find $y \in Y$ with $E(y, u, \xi) = 0$.
 (c) For each $(u, \xi) \in U \times \Xi$, $E_y(S(u, \xi), u, \xi)$ has a bounded inverse.

Assumptions 1 and 2 and the implicit function theorem ensure that $S(\cdot, \xi)$ is continuously differentiable on U for each $\xi \in \Xi$. Let us define $\hat{J}_1 : U \times \Xi \rightarrow [0, \infty)$ by

$$\hat{J}_1(u, \xi) = J_1(S(u, \xi), \xi) \tag{4}$$

and $\hat{J} : U \times \Xi \rightarrow [0, \infty)$ by

$$\hat{J}(u, \xi) = \hat{J}_1(u, \xi) + \psi(u) + (\alpha/2) \|u\|_U^2. \tag{5}$$

Let us fix $\xi \in \Xi$. Assumptions 1 and 2 allow us to use the adjoint approach [32, Sect. 1.6.2] to compute the gradient of the function $\hat{J}_1(\cdot, \xi)$ defined in (4) at each $u \in U$. It yields the gradient

$$\nabla_u \hat{J}_1(u, \xi) = E_u(S(u, \xi), u, \xi)^* z(u, \xi), \tag{6}$$

where for each $(u, \xi) \in U \times \Xi$, $z(u, \xi) \in Z^*$ is the unique solution to the (parameterized) adjoint equation: find $z \in Z^*$ with

$$E_y(S(u, \xi), u, \xi)^* z = -D_y J_1(S(u, \xi), \xi). \tag{7}$$

Assumption 3 The risk-neutral problem (1) has a solution. For each $N \in \mathbb{N}$ and every $\omega \in \Omega$, the SAA problem (2) has a solution.

We refer the reader to [42, Theorem 1] and [44, Proposition 3.12] for theorems on the existence of solutions to risk-averse PDE-constrained optimization problems.

For some $u_0 \in U_{\text{ad}}$ with $\mathbb{E}[\widehat{J}(u_0, \xi)] < \infty$ and a scalar $\rho \in (0, \infty)$, we define the set

$$V_{\text{ad}}^\rho(u_0) = \{u \in U_{\text{ad}} : (\alpha/2) \|u\|_U^2 \leq \mathbb{E}[\widehat{J}(u_0, \xi)] + \rho\}.$$

The existence of such a point u_0 is implied by Assumption 3, for example. Whereas U_{ad} may be unbounded, the set $V_{\text{ad}}^\rho(u_0)$ is bounded. If U_{ad} is bounded and $\rho \in (0, \infty)$ is sufficiently large, then $V_{\text{ad}}^\rho(u_0) = U_{\text{ad}}$. Each solution to the risk-neutral problem (1) is contained in $V_{\text{ad}}^\rho(u_0)$, because $\widehat{J}_1 \geq 0$, $\psi \geq 0$, and $u_0 \in U_{\text{ad}}$.

Assumption 4 allows us to construct compact subsets of the bounded set $V_{\text{ad}}^\rho(u_0)$.

Assumption 4 (a) The linear operator $K : V \rightarrow U$ is compact, V is a real, separable Banach space, and $B_{V_{\text{ad}}^\rho(u_0)} \subset U$ is a bounded, convex neighborhood of $V_{\text{ad}}^\rho(u_0)$.
 (b) The mapping $M : U \times \Xi \rightarrow V$ is a Carathéodory mapping and for all $(u, \xi) \in U \times \Xi$,

$$\nabla_u \widehat{J}_1(u, \xi) = K[M(u, \xi)]. \tag{8}$$

(c) For some integrable random variable $\zeta : \Xi \rightarrow [0, \infty)$,

$$\|M(u, \xi)\|_V \leq \zeta(\xi) \quad \text{for all } (u, \xi) \in B_{V_{\text{ad}}^\rho(u_0)} \times \Xi. \tag{9}$$

Assumption 4 (b) and the gradient formula in (6) yield for all $(u, \xi) \in U \times \Xi$,

$$E_u(S(u, \xi), u, \xi)^* z(u, \xi) = K[M(u, \xi)].$$

Assumption 4 (c) may be verified using stability estimates for the solution operator and adjoint state. If $B_{V_{\text{ad}}^\rho(u_0)}$ would be unbounded, then Assumption 4 (c) may be violated.

Lemma 1 *If Assumptions 1 and 2 hold, then $\widehat{J}_1 : U \times \Xi \rightarrow [0, \infty)$ is a Carathéodory mapping.*

Proof For each $\xi \in \Xi$, the implicit function theorem when combined with Assumption 1 and 2 ensures that the mappings $S(\cdot, \xi)$ is continuously differentiable. In particular, $\widehat{J}_1(\cdot, \xi)$ is continuous. Fix $u \in U$. The measurability of $S(u, \cdot)$ follows from [8, Theorem 8.2.9] when combined with Assumptions 1 and 2. Using the definition of \widehat{J}_1 provided in (4), the measurability of $J_1(u, \cdot)$ and of $S(u, \cdot)$, the separability of Y , and the composition rule [35, Corollary 1.1.11], we find that $\widehat{J}_1(u, \cdot)$ is measurable. \square

We define the expectation function $F_1 : B_{V_{ad}^\rho(u_0)} \rightarrow \mathbb{R}$ and the sample average function $\hat{F}_{1,N} : B_{V_{ad}^\rho(u_0)} \rightarrow \mathbb{R}$ by

$$F_1(u) = \mathbb{E}[\hat{J}_1(u, \xi)] \quad \text{and} \quad \hat{F}_{1,N}(u) = \frac{1}{N} \sum_{i=1}^N \hat{J}_1(u, \xi^i). \tag{10}$$

Lemma 2 *If Assumptions 1-4 hold, then F_1 and $\hat{F}_{1,N}$ are continuously differentiable on $B_{V_{ad}^\rho(u_0)}$ and for each $u \in B_{V_{ad}^\rho(u_0)}$, we have $\nabla F_1(u) = \mathbb{E}[\nabla_u \hat{J}_1(u, \xi)]$ and $\nabla \hat{F}_{1,N}(u) = (1/N) \sum_{i=1}^N \nabla_u \hat{J}_1(u, \xi^i)$.*

We prove Lemma 2 using Lemma 3.

Lemma 3 *If Assumptions 1,2 and 4 hold, then for all $\xi \in \Xi$, the function $\hat{J}_1(\cdot, \xi)$ is continuously differentiable on U , and for all $u \in B_{V_{ad}^\rho(u_0)}$, we have*

$$\hat{J}_1(u, \xi) \leq \hat{J}_1(u_0, \xi) + C_K \zeta(\xi) \|u - u_0\|_U,$$

where $C_K \in [0, \infty)$ is the operator norm of K .

Proof Since K is linear and compact, it is bounded [49, Lemma 8.1-2]. Hence C_K is finite. For each $\xi \in \Xi$, $\hat{J}_1(\cdot, \xi)$ is continuously differentiable on U owing to the implicit function theorem and Assumptions 1 and 2. Since $\psi \geq 0$ and $\hat{J}_1 \geq 0$, we have $u_0 \in V_{ad}^\rho(u_0)$. Using the mean-value theorem, the convexity of $B_{V_{ad}^\rho(u_0)}$, $u_0 \in B_{V_{ad}^\rho(u_0)}$, the formula (8), and the estimate (9), we obtain

$$\begin{aligned} \hat{J}_1(u, \xi) - \hat{J}_1(u_0, \xi) &\leq \sup_{t \in (0,1)} \|\nabla_u \hat{J}_1(u_0 + t(u - u_0), \xi)\|_U \|u - u_0\|_U \\ &\leq C_K \zeta(\xi) \|u - u_0\|_U. \end{aligned}$$

□

Proof (Proof of Lemma 2) Owing to $\psi(u_0) \in [0, \infty)$, $\mathbb{E}[\hat{J}(u_0, \xi)] < \infty$, and $\hat{J}_1 \geq 0$, we have $\mathbb{E}[\hat{J}_1(u_0, \xi)] \in [0, \infty)$. Combined with Lemma 3 and $\mathbb{E}[\zeta(\xi)] < \infty$, we find that F_1 is well-defined on the open set $B_{V_{ad}^\rho(u_0)}$. Moreover, $\hat{J}_1(\cdot, \xi)$ is continuously differentiable on U for all $\xi \in \Xi$. Combined with Assumption 4 and [25, Lemma C.3], we find that F_1 and $\hat{F}_{1,N}$ are Fréchet differentiable on $B_{V_{ad}^\rho(u_0)}$ with the asserted derivatives. Using Assumption 4 and the dominated convergence theorem, we obtain the continuity of the Fréchet derivatives on $B_{V_{ad}^\rho(u_0)}$. □

3.1 Compact Subsets

We define a compact subset of the feasible set U_{ad} that contains the solutions to the risk-neutral problem (1) and eventually those to its SAA problem (2).

Let us define

$$W_{ad}^\rho = V_{ad}^\rho(u_0) \cap \overline{\{\text{prox}_{\psi/\alpha}(-(1/\alpha)K[v]) : v \in V, \|v\|_V \leq \mathbb{E}[\zeta(\xi)] + \rho\}}^{\|\cdot\|_U}, \tag{11}$$

where $\overline{U_0}^{\|\cdot\|_U}$ denotes the $\|\cdot\|_U$ -closure of $U_0 \subset U$.

Lemma 4 *If Assumptions 1,2 and 4 hold, then W_{ad}^ρ is a compact subset of U_{ad} .*

Proof We first show that the second set on the right-hand side in (11) is compact. Assumption 4 (c) yields $\mathbb{E}[\zeta(\xi)] < \infty$. Hence the set $\{v \in V : \|v\|_V \leq \mathbb{E}[\zeta(\xi)] + \rho\}$ is bounded. Thus, its image under the compact operator K (see Assumption 4 (a)) is precompact. The operator $\text{prox}_{\psi/\alpha}(-(1/\alpha)\cdot) : U \rightarrow U$ is continuous, as $\text{prox}_{\psi/\alpha}$ is firmly nonexpansive [10, Proposition 12.28]. Since each continuous function maps precompact sets to precompact ones [49, p. 412], the second set on the right-hand side in (11) is compact. This set is a subset of U_{ad} because $\text{prox}_{\psi/\alpha}(U) \subset U_{ad}$. Since $V_{ad}^\rho(u_0)$ is closed, the set W_{ad}^ρ is compact. Owing to $V_{ad}^\rho(u_0) \subset U_{ad}$, we have $W_{ad}^\rho \subset U_{ad}$. \square

For each $\omega \in \Omega$, we define

$$W_{ad}^{[N]}(\omega) = \overline{\{u = \text{prox}_{\psi/\alpha}(-(1/\alpha)\nabla_u \hat{F}_{1,N}(u, \omega)) : u \in V_{ad}^\rho(u_0)\}}^{\|\cdot\|_U}. \tag{12}$$

Lemma 5 *Let Assumptions 1-4 hold. Then the following assertions hold.*

1. *The set of solutions to (1) is contained in W_{ad}^ρ .*
2. *We have w.p. 1 for all sufficiently large N , $W_{ad}^{[N]} \subset W_{ad}^\rho$.*

Proof 1. Let u^* be a solution to (1). Since $\hat{J}_1 \geq 0$ and $\psi \geq 0$, we have $u^* \in V_{ad}^\rho(u_0)$. Lemma 2 ensures that F_1 is continuously differentiable on $B_{V_{ad}^\rho(u_0)}$. Hence $u^* = \text{prox}_{\psi/\alpha}(-(1/\alpha)\nabla F_1(u^*))$ (cf. [63, Proposition 3.5] and [53, p. 2092]). Using Assumption 4, in particular the bound in (9), and [35, Proposition 1.2.2], we have

$$\|\mathbb{E}[M(u^*, \xi)]\|_V \leq \mathbb{E}[\|M(u^*, \xi)\|_V] \leq \mathbb{E}[\zeta(\xi)] < \infty.$$

Combined with (8) and [35, eq. (1.2)], we find that

$$\nabla F_1(u^*) = \mathbb{E}[KM(u^*, \xi)] = K\mathbb{E}[M(u^*, \xi)].$$

Since $u^* = \text{prox}_{\psi/\alpha}(-(1/\alpha)K\mathbb{E}[M(u^*, \xi)])$ and $\rho > 0$, we have $u^* \in W_{ad}^\rho$ (see (11)).

2. The (strong) law of large numbers ensures $(1/N) \sum_{i=1}^N \zeta(\xi^i) \rightarrow \mathbb{E}[\zeta(\xi)]$ w.p. 1 as $N \rightarrow \infty$. Combined with $\rho > 0$, we deduce the existence of an event $\Omega_1 \in \mathcal{F}$ with $P(\Omega_1) = 1$ and for each $\omega \in \Omega_1$, there exists $n(\omega) \in \mathbb{N}$ such that for all $N \geq n(\omega)$, we have

$$\frac{1}{N} \sum_{i=1}^N \zeta(\xi^i(\omega)) \leq \mathbb{E}[\zeta(\xi)] + \rho.$$

Fix $\omega \in \Omega_1$ and let $N \geq n(\omega)$. Let $u \in V_{\text{ad}}^\rho(u_0)$ be arbitrary. Using $V_{\text{ad}}^\rho(u_0) \subset B_{V_{\text{ad}}^\rho(u_0)}$ and Assumption 4, we find that

$$\begin{aligned} \left\| \frac{1}{N} \sum_{i=1}^N M(u, \xi^i(\omega)) \right\|_V &\leq \frac{1}{N} \sum_{i=1}^N \left\| M(u, \xi^i(\omega)) \right\|_V \leq \frac{1}{N} \sum_{i=1}^N \zeta(\xi^i(\omega)) \\ &\leq \mathbb{E}[\zeta(\xi)] + \rho, \end{aligned}$$

where the right-hand side is independent of $u \in V_{\text{ad}}^\rho(u_0)$. Furthermore

$$\nabla \hat{F}_{1,N}(u, \omega) = (1/N) \sum_{i=1}^N \nabla_u \hat{J}_1(u, \xi^i(\omega)) = K \left((1/N) \sum_{i=1}^N M(u, \xi^i(\omega)) \right).$$

We conclude that $u = \text{prox}_{\psi/\alpha}(- (1/\alpha) \nabla \hat{F}_{1,N}(u, \omega)) \in W_{\text{ad}}^\rho$ for each $u \in V_{\text{ad}}^\rho(u_0)$. Since W_{ad}^ρ is closed (see Lemma 4), we have $W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho$. Hence

$$\Omega_1 \subset \{ \omega \in \Omega : \exists n(\omega) \in \mathbb{N} \ \forall N \geq n(\omega); \ W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho \}.$$

The set on the right-hand side is a subset of Ω . Since $\Omega_1 \in \mathcal{F}$, $P(\Omega_1) = 1$ and (Ω, \mathcal{F}, P) is complete, the set on the right-hand side is measurable and hence occurs w.p. 1. □

To establish the measurability of the event “for all sufficiently large N , $W_{\text{ad}}^{[N]} \subset W_{\text{ad}}^\rho$,” we used the fact that (Ω, \mathcal{F}, P) is complete. Since this event equals the limit inferior of the sequence $(\{ \omega \in \Omega : W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho \})_N$, the measurability of the event would also be implied by that of $\{ \omega \in \Omega : W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho \}$ for each $N \in \mathbb{N}$ [12, p. 55]. This approach would require us to show that $\{ \omega \in \Omega : W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho \}$ is measurable for each $N \in \mathbb{N}$, which entails those of $\{ \omega \in \Omega : W_{\text{ad}}^{[N]}(\omega) \subset W_{\text{ad}}^\rho \}$ and $W_{\text{ad}}^{[N]}$. Using [8, Theorem 8.2.8], we can show that $W_{\text{ad}}^{[N]}$ is measurable. However, an application of [8, Theorem 8.2.8] requires (Ω, \mathcal{F}, P) be complete.

3.2 Consistency of SAA Optimal Values and SAA Solutions

We demonstrate the consistency of the SAA optimal value and the SAA solutions. Let ϑ^* and \mathcal{S} be the optimal value and the set of solutions to (1), respectively. Moreover, for each $\omega \in \Omega$, let $\hat{\vartheta}_N^*(\omega)$ and $\hat{\mathcal{S}}_N(\omega)$ be the optimal value and the set of solutions to the SAA problem (2), respectively.

We define the distance $\text{dist}(u, \mathcal{S})$ from $u \in \hat{\mathcal{S}}_N(\omega)$ to \mathcal{S} and the deviation $\mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S})$ between the sets $\hat{\mathcal{S}}_N(\omega)$ and \mathcal{S} by

$$\text{dist}(u, \mathcal{S}) = \inf_{v \in \mathcal{S}} \|u - v\|_U \quad \text{and} \quad \mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S}) = \sup_{u \in \hat{\mathcal{S}}_N(\omega)} \text{dist}(u, \mathcal{S}).$$

Theorem 1 *If Assumptions 1-4 hold, then $\hat{\vartheta}_N^* \rightarrow \vartheta^*$ and $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}) \rightarrow 0$ w.p. 1 as $N \rightarrow \infty$.*

We prepare our proof of Theorem 1, which is based on that of [72, Theorem 5.3].

Lemma 6 *If Assumptions 1, 2 and 4 hold, then the function \hat{J}_1 defined in (4) is a Carathéodory function on $W_{ad}^\rho \times \Xi$. Moreover, $(\hat{J}_1(u, \xi))_{u \in W_{ad}^\rho}$ is dominated by an integrable function.*

Proof Lemma 4 ensures that W_{ad}^ρ is a compact metric space. Since $W_{ad}^\rho \subset U$ and \hat{J}_1 is a Carathéodory function on $U \times \Xi$ (see Lemma 1), the function \hat{J}_1 is a Carathéodory function on $W_{ad}^\rho \times \Xi$. Lemma 3 ensures that for all $u \in W_{ad}^\rho \subset V_{ad}^\rho(u_0)$,

$$\hat{J}_1(u, \xi) \leq \hat{J}_1(u_0, \xi) + C_K \zeta(\xi) \sup_{u \in W_{ad}^\rho} \|u - u_0\|_U.$$

The random variable on the right-hand side is integrable owing to the integrability of ζ (see Assumption 4 (c)), the boundedness of W_{ad}^ρ (see Lemma 4), $C_K \in [0, \infty)$, and $\mathbb{E}[\hat{J}(u_0, \xi)] < \infty$. Combined with $\hat{J}_1 \geq 0$, we find that $(\hat{J}_1(u, \xi))_{W_{ad}^\rho}$ is dominated by an integrable random variable. □

Lemma 7 *If Assumptions 1-4 hold, then for each $N \in \mathbb{N}$, the functions $\hat{\vartheta}_N^*$ and $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S})$ are measurable.*

Proof For each $\omega \in \Omega$, Assumption 3 ensures that $\hat{\mathcal{S}}_N(\omega)$ is nonempty. The function \hat{J}_1 is Carathéodory function on $U \times \Xi$ according to Lemma 6 and ψ is lower semicontinuous according to Assumption 1 (c). Hence $\hat{\vartheta}_N^*$ is measurable [15, Lemma III.39] and the set-valued mapping $\hat{\mathcal{S}}_N$ is measurable [15, p. 86]. Assumption 3 implies that \mathcal{S} is nonempty and, hence, $\text{dist}(\cdot, \mathcal{S})$ is (Lipschitz) continuous [4, Theorem 3.16]. For each $\omega \in \Omega$, $\hat{F}_N(\cdot, \omega)$ is lower semicontinuous and hence $\hat{\mathcal{S}}_N(\omega)$ is closed. Thus $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S})$ is measurable [8, Theorem 8.2.11]. □

Lemma 8 *If Assumptions 1-4 hold, then \hat{F}_N converges to F w.p. 1 uniformly on W_{ad}^ρ .*

Proof We first verify the hypotheses of the uniform law of large numbers established in [52, Corollary 4:1] to demonstrate the uniform almost sure convergence of $\hat{F}_{1,N}$ to F_1 on W_{ad}^ρ .

Lemma 6 ensures that \hat{J}_1 is a Carathéodory function on $W_{ad}^\rho \times \Xi$ and that $(\hat{J}_1(u, \xi))_{u \in W_{ad}^\rho}$ is dominated by an integrable function. Moreover, W_{ad}^ρ is a compact metric space (see Lemma 4). Since ξ^1, ξ^2, \dots are independent identically distributed random elements, the uniform law of large numbers [52, Corollary 4:1] implies that $\hat{F}_{1,N}(\cdot) = (1/N) \sum_{i=1}^N \hat{J}_1(\cdot, \xi^i)$ converges to $F_1(\cdot) = \mathbb{E}[\hat{J}_1(\cdot, \xi)]$ w.p. 1 uniformly on W_{ad}^ρ .

Since U_{ad} is the domain of ψ and $\psi \geq 0$, we have $\psi(u) \in [0, \infty)$ for all $u \in U_{ad}$. Lemma 4 ensures $W_{ad}^\rho \subset U_{ad}$. Hence for all $u \in W_{ad}^\rho$,

$$\begin{aligned} \hat{F}_N(u) - F(u) &= \hat{F}_{1,N}(u) + (\alpha/2) \|u\|_U^2 + \psi(u) - (F_1(u) + (\alpha/2) \|u\|_U^2 + \psi(u)) \\ &= \hat{F}_{1,N}(u) - F_1(u). \end{aligned}$$

Therefore, the assertion follows from the above uniform convergence statement. \square

Lemma 9 demonstrates that the SAA solution set is eventually contained in the compact set W_{ad}^ρ .

Lemma 9 *If Assumptions 1-4 hold, then w.p. 1 for all sufficiently large N , $\hat{\mathcal{S}}_N \subset W_{ad}^\rho$.*

Proof First, we show that w.p. 1 for all sufficiently large N , $\hat{\mathcal{S}}_N \subset V_{ad}^\rho(u_0)$. Lemma 6 ensures that \hat{J}_1 is a Carathéodory function on $U \times \Xi$. Since $\hat{J} \geq 0$, $u_0 \in U_{ad}$, and $\mathbb{E}[\hat{J}(u_0, \xi)] < \infty$, the (strong) law of large numbers ensures

$$\frac{1}{N} \sum_{i=1}^N \hat{J}(u_0, \xi^i) \rightarrow \mathbb{E}[\hat{J}(u_0, \xi)] \quad \text{w.p. 1 as } N \rightarrow \infty.$$

Combined with $\rho > 0$, we deduce the existence of an event $\Omega_1 \in \mathcal{F}$ such that $P(\Omega_1) = 1$ and for each $\omega \in \Omega_1$, there exists $n_1(\omega) \in \mathbb{N}$ such that for all $N \geq n_1(\omega)$, we have

$$\frac{1}{N} \sum_{i=1}^N \hat{J}(u_0, \xi^i(\omega)) \leq \mathbb{E}[\hat{J}(u_0, \xi)] + \rho. \tag{13}$$

Fix $\omega \in \Omega_1$ and let $N \geq n_1(\omega)$. Using $\psi \geq 0$ and $\hat{J}_1 \geq 0$, we have for all $u_N^* = u_N^*(\omega) \in \hat{\mathcal{S}}_N(\omega)$,

$$\begin{aligned} (\alpha/2) \|u_N^*\|_U^2 &\leq \frac{1}{N} \sum_{i=1}^N \hat{J}_1(u_N^*, \xi^i(\omega)) + \psi(u_N^*) + (\alpha/2) \|u_N^*\|_U^2 \\ &\leq \frac{1}{N} \sum_{i=1}^N \hat{J}(u_0, \xi^i(\omega)). \end{aligned}$$

Combined with (13), we find that $\hat{\mathcal{S}}_N(\omega) \subset V_{ad}^\rho(u_0)$.

By construction of $W_{ad}^{[N]}$, we have $\hat{\mathcal{S}}_N(\omega) \cap V_{ad}^\rho(u_0) \subset W_{ad}^{[N]}(\omega)$ for all $\omega \in \Omega$. Indeed, if $u_N^*(\omega) \in \hat{\mathcal{S}}_N(\omega) \cap V_{ad}^\rho(u_0)$, then we have the first-order optimality condition $u_N^*(\omega) = \text{prox}_{\psi/\alpha}(- (1/\alpha) \nabla \hat{F}_{1,N}(u_N^*(\omega), \omega))$. Hence $u_N^*(\omega) \in W_{ad}^{[N]}(\omega)$. Lemma 5 implies that w.p. 1 for all sufficiently large N , $W_{ad}^{[N]} \subset W_{ad}^\rho$. Hence there exists $\Omega_2 \in \mathcal{F}$ with $P(\Omega_2) = 1$ and for each $\omega \in \Omega_2$ there exists $n_2(\omega) \in \mathbb{N}$ such that for all $N \geq n_2(\omega)$, $W_{ad}^{[N]}(\omega) \subset W_{ad}^\rho$. Putting together the pieces, we find that for all $\omega \in \Omega_1 \cap \Omega_2$ and each $N \geq \max\{n_1(\omega), n_2(\omega)\}$, we have $\hat{\mathcal{S}}_N(\omega) \subset W_{ad}^\rho$. Since (Ω, \mathcal{F}, P) is complete and $P(\Omega_1 \cap \Omega_2) = 1$, we have w.p. 1 for all sufficiently large N , $\hat{\mathcal{S}}_N \subset W_{ad}^\rho$. \square

Proof (Proof of Theorem 1) The proof is based on that of [72, Theorem 5.3]. Lemma 5 yields $\mathcal{S} \subset W_{\text{ad}}^\rho$. Lemma 9 ensures that w.p. 1 for all sufficiently large N , $\hat{\mathcal{S}}_N \subset W_{\text{ad}}^\rho$. Hence, we deduce the existence of an event $\Omega_1 \in \mathcal{F}$ with $P(\Omega_1) = 1$ and for each $\omega \in \Omega_1$ there exists $n(\omega) \in \mathbb{N}$ such that for all $N \geq n(\omega)$, $\hat{\mathcal{S}}_N(\omega) \subset W_{\text{ad}}^\rho$. Lemma 8 ensures that $\hat{F}_N(\cdot, \omega)$ converges to $F(\cdot)$ uniformly on W_{ad}^ρ for almost all $\omega \in \Omega$. Therefore, there exists $\Omega_2 \in \mathcal{F}$ with $P(\Omega_2) = 1$ and for each $\omega \in \Omega_2$, $\hat{F}_N(\cdot, \omega)$ converges to $F(\cdot)$ uniformly on W_{ad}^ρ .

We show that $\hat{\vartheta}_N^*(\omega) \rightarrow \vartheta^*$ as $N \rightarrow \infty$ for each $\omega \in \Omega_1 \cap \Omega_2$. Fix $\omega \in \Omega_1 \cap \Omega_2$. Assumption 3 ensures that \mathcal{S} and $\hat{\mathcal{S}}_N(\omega)$ are nonempty for all $N \in \mathbb{N}$. Let $u^* \in \mathcal{S}$ and let $u_N^*(\omega) \in \hat{\mathcal{S}}_N(\omega)$. Then for all $N \geq n(\omega)$, we have $u_N^*(\omega) \in W_{\text{ad}}^\rho$ and hence $|\hat{\vartheta}_N^*(\omega) - \vartheta^*| \leq \sup_{u \in W_{\text{ad}}^\rho} |\hat{F}_N(u, \omega) - F(u)|$ for all $N \geq n(\omega)$ (cf. [37, pp. 194–195]). We deduce $\hat{\vartheta}_N^*(\omega) \rightarrow \vartheta^*$ as $N \rightarrow \infty$.

Next, we show that $\mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S}) \rightarrow 0$ as $N \rightarrow \infty$ for each $\omega \in \Omega_1 \cap \Omega_2$. Fix $\omega \in \Omega_1 \cap \Omega_2$. Since \mathcal{S} is nonempty (see Assumption 3), the function $\text{dist}(\cdot, \mathcal{S})$ is (Lipschitz) continuous [4, Theorem 3.16]. For each $N \geq n(\omega)$, the set $\hat{\mathcal{S}}_N(\omega)$ is closed and $\hat{\mathcal{S}}_N(\omega) \subset W_{\text{ad}}^\rho$. Hence $\hat{\mathcal{S}}_N(\omega)$ is compact for each $N \geq n(\omega)$. Therefore, for each $N \geq n(\omega)$, there exists $u_N = u_N(\omega) \in W_{\text{ad}}^\rho$ with $\text{dist}(u_N, \mathcal{S}) = \mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S})$. Suppose that $\mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S}) \not\rightarrow 0$. We deduce the existence of a subsequence $\mathcal{N} = \mathcal{N}(\omega)$ of $(n(\omega), n(\omega) + 1, \dots)$ such that $\mathbb{D}(u_N, \mathcal{S}) \geq \varepsilon$ for all $N \in \mathcal{N}$ and some $\varepsilon > 0$, and $u_N \rightarrow \bar{u} \in W_{\text{ad}}^\rho$ as $\mathcal{N} \ni N \rightarrow \infty$. Combined with the fact that $\text{dist}(\cdot, \mathcal{S})$ is continuous, we obtain $\bar{u} \notin \mathcal{S}$. Hence $F(\bar{u}) > \vartheta^*$. We have

$$\liminf_{\mathcal{N} \ni N \rightarrow \infty} \hat{F}_N(u_N, \omega) = \lim_{\mathcal{N} \ni N \rightarrow \infty} (\hat{F}_N(u_N, \omega) - F(u_N)) + \liminf_{\mathcal{N} \ni N \rightarrow \infty} F(u_N).$$

The uniform convergence implies that the first term in the right-hand side is zero. Since F is lower semicontinuous on U_{ad} (see Assumption 1 and Lemma 2), $F(\bar{u}) > \vartheta^*$, and $\hat{F}_N(u_N, \omega) = \hat{\vartheta}_N^*(\omega)$, we find that

$$\liminf_{\mathcal{N} \ni N \rightarrow \infty} \hat{\vartheta}_N^*(\omega) = \liminf_{\mathcal{N} \ni N \rightarrow \infty} \hat{F}_N(u_N, \omega) = \liminf_{\mathcal{N} \ni N \rightarrow \infty} F(u_N) \geq F(\bar{u}) > \vartheta^*.$$

This contradicts $\hat{\vartheta}_N^*(\omega) \rightarrow \vartheta^*$ as $N \rightarrow \infty$. Hence $\mathbb{D}(\hat{\mathcal{S}}_N(\omega), \mathcal{S}) \rightarrow 0$ as $N \rightarrow \infty$. Combined with Lemma 7 and the fact that $P(\Omega_1 \cap \Omega_2) = 1$, we obtain the almost sure convergence statements. □

4 Examples

We present three risk-neutral nonlinear PDE-constrained optimization problems and verify the assumptions made in Sect. 3, except Assumption 3 on the existence of solutions in order to keep the section relatively short.

We use the following facts. (i) The Sobolev spaces $H_0^1(D)$ and $H^1(D)$ are separable Hilbert spaces [1, Theorem 3.5]. (ii) If a real Banach space is reflexive and separable, then its dual is separable [1, Theorem 1.14]. (iii) The operator norm of a linear, bounded

operator equals that of its (Banach space-)adjoint operator [49, Theorem 4.5-2]. (iv) If Λ_1 and Λ_2 are real, reflexive Banach spaces and $\Upsilon : \Lambda_1 \rightarrow \Lambda_2$ is linear and bounded, then $(\Upsilon^*)^* = \Upsilon$ [6, p. 390] (see also [65, Theorem 8.57]) because we write $(\Lambda_i^*)^* = \Lambda_i$ for $i \in \{1, 2\}$.

4.1 Boundary Optimal Control of a Semilinear State Equation

We consider the risk-neutral boundary optimal control of a parameterized semilinear PDE. Our model problem is based on the deterministic semilinear boundary control problems studied in [14, 32, 36, 76].

We consider

$$\min_{u \in L^2(\partial D)} (1/2)\mathbb{E} \left[\|S(u, \xi) - y_d\|_{L^2(D)}^2 \right] + (\alpha/2) \|u\|_{L^2(\partial D)}^2 + \psi(u), \tag{14}$$

where ∂D is the boundary of $D \subset \mathbb{R}^2$ and for each $(u, \xi) \in L^2(\partial D) \times \Xi$, the state $S(u, \xi) \in H^1(D)$ is the weak solution to: find $y \in H^1(D)$ with

$$-\nabla \cdot (\kappa(\xi)\nabla y) + g(\xi)y + y^3 = b(\xi) \text{ in } D, \quad \kappa(\xi)\partial_\nu y + \sigma(\xi)y = Bu \text{ on } \partial D, \tag{15}$$

where $\partial_\nu y$ is the normal derivative of y ; see [76, p. 31]. For a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we denote by $L^2(\partial D)$ the space of square integrable functions on ∂D and by $L^\infty(\partial D)$ that of essentially bounded functions [6, p. 263]. The space $L^2(\partial D)$ is a Hilbert space with inner product $(v, w)_{L^2(\partial D)} = \int_{\partial D} v(x)w(x)d\mathbb{H}^{d-1}(x)$, where \mathbb{H}^{d-1} is the $(d - 1)$ -dimensional Hausdorff measure on ∂D [6, Theorem 3.16 and pp. 47, 263 and 267]. The space $L^2(\partial D)$ is separable [61, Theorem 4.1].

We formulate assumptions on the control problem (14).

- $D \subset \mathbb{R}^2$ is a bounded Lipschitz domain.
- $\kappa, g : \Xi \rightarrow L^\infty(D)$ are strongly measurable and there exist $\kappa_{\min}, \kappa_{\max}, g_{\min}, g_{\max} \in (0, \infty)$ such that $\kappa_{\min} \leq \kappa(\xi) \leq \kappa_{\max}$ and $g_{\min} \leq g(\xi) \leq g_{\max}$ for all $\xi \in \Xi$.
- $b : \Xi \rightarrow L^2(D)$ and $\sigma : \Xi \rightarrow L^\infty(\partial D)$ are strongly measurable with $\mathbb{E} \left[\|b(\xi)\|_{L^2(D)}^2 \right] < \infty, \mathbb{E} \left[\|\sigma(\xi)\|_{L^\infty(\partial D)}^2 \right] < \infty$ and $\sigma(\xi) \geq 0$ for all $\xi \in \Xi$.
- $B : L^2(\partial D) \rightarrow L^2(\partial D)$ is a linear, bounded operator.
- $y_d \in L^2(D), \alpha > 0$, and $\psi : L^2(\partial D) \rightarrow [0, \infty]$ is proper, convex, and lower semicontinuous.

Throughout the section, we assume these conditions be satisfied.

We establish Assumption 1. Since the embedding $H^1(D) \hookrightarrow L^2(D)$ is continuous, the function $J_1 : H^1(D) \rightarrow [0, \infty)$ defined by $J_1(y) = (1/2) \|y - y_d\|_{L^2(D)}^2$ is continuously differentiable. We find that Assumption 1 holds true.

We formulate the weak form of (15) as an operator equation; cf. [36, eq. (2)]. We define $E : H^1(D) \times L^2(\partial D) \times \Xi \rightarrow H^1(D)^*$ by

$$\begin{aligned} \langle E(y, u, \xi), v \rangle_{H^1(D)^*, H^1(D)} &= (\kappa(\xi)\nabla y, \nabla v)_{L^2(D)^2} + (g(\xi)y + y^3, v)_{L^2(D)} \\ &\quad + (\sigma(\xi)\tau_{\partial D}[y], \tau_{\partial D}[v])_{L^2(\partial D)} \\ &\quad - (b(\xi), v)_{L^2(D)} - (Bu, \tau_{\partial D}[v])_{L^2(\partial D)}. \end{aligned} \tag{16}$$

where $\tau_{\partial D} : H^1(D) \rightarrow L^2(\partial D)$ is the trace operator. We refer the reader to [6, p. 268] for the definition of $\tau_{\partial D}$. Since D has a Lipschitz boundary, the trace operator $\tau_{\partial D}$ is linear and compact [61, Theorem 6.2].

We verify Assumption 2. Using [32, Theorem 1.15], we find that $E(y, u, \xi) = 0$ has a unique solution $S(u, \xi) \in H^1(D)$ for each $(u, \xi) \in L^2(\partial D) \times \Xi$. Since the embedding $H^1(D) \hookrightarrow L^6(D)$ is continuous [32, Theorem 1.14], we have $y^3 \in L^2(D)$ for each $y \in H^1(D)$ [32, p. 57] and the mapping $L^6(D) \ni y \mapsto y^3 \in L^2(D)$ is continuously differentiable [32, p. 76]. We find that $E(\cdot, \cdot, \xi)$ is continuously differentiable. Now, the Lax–Milgram lemma can be used to show that $E_y(S(u, \xi), u, \xi)$ has a bounded inverse. We show that $E(y, u, \cdot)$ is measurable for each $(y, u) \in H^1(D) \times L^2(\partial D)$. Since $H^1(D)^*$ is separable, it suffices to show that $\xi \mapsto \langle E(y, u, \xi), v \rangle_{H^1(D)^*, H^1(D)}$ is measurable for each fixed $(y, v, u) \in H^1(D)^2 \times L^2(\partial D)$ [35, Theorem 1.1.6]. We define $\phi : L^\infty(D) \rightarrow \mathbb{R}$ by $\phi(v) = (v\nabla y, \nabla v)_{L^2(D)^2}$. Hölder’s inequality ensures that ϕ is (Lipschitz) continuous. Since $\xi \mapsto (\kappa(\xi)\nabla y, \nabla v)_{L^2(D)^2}$ is the composition of the continuous function ϕ with κ , it is measurable [35, Corollary 1.1.11]. Similar arguments can be used to establish the measurability of the other terms in (16). Hence Assumption 2 holds true.

We establish Assumption 4. Fix $(u, \xi) \in L^2(\partial D) \times \Xi$. Choosing $v = S(u, \xi)$ in (16) and using

$$(\kappa(\xi)\nabla y, \nabla y)_{L^2(D)^2} + (g(\xi)y, y)_{L^2(D)} \geq \min\{\kappa_{\min}, g_{\min}\} \|y\|_{H^1(D)}^2$$

valid for all $y \in H^1(D)$, we obtain the stability estimate

$$\min\{\kappa_{\min}, g_{\min}\} \|S(u, \xi)\|_{H^1(D)} \leq \|b(\xi)\|_{L^2(D)} + C_{\tau_{\partial D}} \|Bu\|_{L^2(\partial D)}, \tag{17}$$

where $C_{\tau_{\partial D}}$ is the operator norm of $\tau_{\partial D}$. For each $(u, \xi) \in L^2(\partial D) \times \Xi$, let $z(u, \xi)$ be the unique solution to the adjoint equation: find $z \in H^1(D)$ with

$$\begin{aligned} (\kappa(\xi)\nabla z, \nabla v)_{L^2(D)^2} + (g(\xi)z + 3S(u, \xi)^2z, v)_{L^2(D)} + (\sigma(\xi)\tau_{\partial D}[z], \tau_{\partial D}[v])_{L^2(\partial D)} \\ = -(S(u, \xi) - y_d, v)_{L^2(D)} \quad \text{for all } v \in H^1(D); \end{aligned}$$

cf. [76, eq. (4.54)] and [36, p. 729]. For its solution $z(u, \xi)$, we obtain

$$\min\{\kappa_{\min}, g_{\min}\} \|z(u, \xi)\|_{H^1(D)} \leq \|S(u, \xi) - y_d\|_{L^2(D)}. \tag{18}$$

Since $\tau_{\partial D}^*$ is the adjoint operator of $\tau_{\partial D}$, we have for all $u \in L^2(\partial D)$ and $v \in H^1(D)$, $(Bu, \tau_{\partial D}[v])_{L^2(\partial D)} = \langle \tau_{\partial D}^*Bu, v \rangle_{H^1(D)^*, H^1(D)}$. Combined with $\tau_{\partial D} = (\tau_{\partial D}^*)^*$ and the identity $E_u(S(u, \xi), u, \xi) = -\tau_{\partial D}^*B$ (cf. [32, p. 136]), the gradient formula in (6) yields

$$\nabla \widehat{J}_1(u, \xi) = -B^* \tau_{\partial D}[z(u, \xi)].$$

We choose $K = -B^* \tau_{\partial D}$ and $M(u, \xi) = z(u, \xi)$. The operator $K : H^1(D) \rightarrow L^2(\partial D)$ is compact, as B is linear and bounded and $\tau_{\partial D}$ is linear and compact [61, Theorem 6.2]. Using [8, Theorem 8.2.9] and the measurability of $S(u, \cdot)$ (see Lemma 1), we can show that $z(u, \cdot)$ is measurable for all $u \in L^2(\partial D)$. The implicit function theorem implies that $z(\cdot, \xi)$ is continuous for each $\xi \in \Xi$. Since ψ is proper, there exists $u_0 \in L^2(\partial D)$ with $\psi(u_0) < \infty$. Using Young’s inequality, we have $\widehat{J}_1(u_0, \xi) \leq \|y_d\|_{L^2(D)}^2 + \|S(u_0, \xi)\|_{L^2(D)}^2$. Combined with (17), we find that $\mathbb{E}[\widehat{J}_1(u_0, \xi)] < \infty$ and hence $\mathbb{E}[\widehat{J}(u_0, \xi)] < \infty$. Let $B_{V_{ad}^\rho(u_0)}$ be an open, bounded ball about zero containing $V_{ad}^\rho(u_0)$ and let R_{ad} be its radius. We define

$$\zeta(\xi) = \frac{1}{\min\{\kappa_{\min}, g_{\min}\}} \left(\|y_d\|_{L^2(D)} + \frac{C_{\tau_{\partial D}} C_B R_{ad} + \|b(\xi)\|_{L^2(D)}}{\min\{\kappa_{\min}, g_{\min}\}} \right).$$

where $C_B > 0$ is the operator norm of B . The random variable ζ is integrable. Using the stability estimates (17) and (18), we conclude that Assumption 4 holds true.

4.2 Distributed Control of a Steady Burgers’ Equation

We consider the risk-neutral distributed optimization of a steady Burgers’ equation. Deterministic optimal control problems with the Burgers’ equation are studied, for example, in [19, 78–80]. We refer the reader to [40, 43, 46, 48] for risk-neutral and risk-averse control of the steady Burgers’ equation.

Let us consider

$$\min_{u \in U_{ad}} (1/2)\mathbb{E} \left[\|S(u, \xi) - y_d\|_{L^2(0,1)}^2 \right] + (\alpha/2) \|u\|_{L^2(D_0)}^2, \tag{19}$$

where $D_0 \subset (0, 1)$ is a nonempty domain and for all $(u, \xi) \in L^2(D_0) \times \Xi$, the state $S(u, \xi) \in H^1(0, 1)$ is the weak solution to the steady Burgers’ equation: find $y \in H_0^1(0, 1)$ with

$$-\kappa(\xi)y'' + yy' = b(\xi) + Bu \quad \text{in } (0, 1), \quad y(0) = 0, \quad y(1) = 0,$$

where $b : \Xi \rightarrow L^2(D)$ and $\kappa : \Xi \rightarrow (0, \infty)$. As in [78, p. 78], $B : L^2(D_0) \rightarrow L^2(0, 1)$ is defined by $(Bu)(x) = u(x)$ if $x \in D_0$ and 0 else. We consider homogeneous Dirichlet boundary conditions, as it simplifies the derivation of a state stability estimate.

The weak form of the steady Burgers’ equation has at least one solution $S(u, \xi) \in H_0^1(0, 1)$ for each $(u, \xi) \in L^2(D_0) \times \Xi$ [79, Proposition 3.1]. We assume that the solution $S(u, \xi)$ be unique to ensure that the reduced formulation (19) is well-defined.

A condition sufficient for uniqueness is that $\kappa(\xi)$ is sufficiently large [79, Proposition 3.1]. We formulate the uniqueness as an assumption.

- $\kappa : \Xi \rightarrow \mathbb{R}$ is measurable and there exists $\kappa_{\min}, \kappa_{\max} \in (0, \infty)$ such that $\kappa_{\min} \leq \kappa(\xi) \leq \kappa_{\max}$ for all $\xi \in \Xi$.
- $b : \Xi \rightarrow L^2(0, 1)$ is strongly measurable and there exists $b_{\max} \in (0, \infty)$ such that $\|b(\xi)\|_{L^2(0,1)} \leq b_{\max}$ for all $\xi \in \Xi$.
- For each $(u, \xi) \in L^2(D_0) \times \Xi$, the solution $S(u, \xi) \in H_0^1(0, 1)$ to the weak form of the steady Burgers’ equation is unique.
- $y_d \in L^2(0, 1)$, $U_{\text{ad}} \subset L^2(D_0)$ is nonempty, closed, and convex, and $\alpha > 0$.

Throughout the section, we assume these conditions be satisfied.

Let us verify Assumption 1. The constraints in (19) can be modeled using the indicator function $\psi = I_{U_{\text{ad}}}$. Since U_{ad} is nonempty, closed, and convex, the function $I_{U_{\text{ad}}}$ is proper, convex, and lower semicontinuous [13, Ex. 2.67]. The function $J_1 : H_0^1(D) \rightarrow [0, \infty)$ defined by $J_1(y) = (1/2) \|y - y_d\|_{L^2(D)}^2$ is continuously differentiable. Putting together the pieces, we find that Assumption 1 holds true.

We define $E : H_0^1(0, 1) \times L^2(D_0) \times \Xi \rightarrow H^{-1}(0, 1)$ by

$$\begin{aligned} \langle E(y, u, \xi), v \rangle_{H^{-1}(D), H_0^1(D)} &= (\kappa(\xi)y', v')_{L^2(0,1)} + (yy', v)_{L^2(0,1)} \\ &\quad - (b(\xi), v)_{L^2(0,1)} - (Bu, v)_{L^2(0,1)}. \end{aligned}$$

Let $\iota : H_0^1(0, 1) \rightarrow L^2(0, 1)$ be the embedding operator of the compact embedding $H_0^1(D) \hookrightarrow L^2(0, 1)$. We have $\langle \iota^*[Bu], v \rangle_{H^{-1}(D), H_0^1(D)} = (Bu, v)_{L^2(0,1)}$ for all $v \in H_0^1(D)$ and $u \in L^2(D_0)$.

We show that Assumption 2 holds true. The operator E is well-defined [78, pp. 76 and 80] and $E(\cdot, \cdot, \xi)$ is twice continuously differentiable for each $\xi \in \Xi$ [78, p. 81]. For each $(u, \xi) \in L^2(D_0) \times \Xi$, $E_y(S(u, \xi), u, \xi)$ has a bounded inverse [46, p. A1866]. Using arguments similar to those in Sect. 4.1, we can show that $E(y, u, \cdot)$ is measurable for each $(y, u) \in H_0^1(D) \times L^2(D_0)$. We conclude that Assumption 2 holds true.

Using the gradient formula (6), $(\iota^*)^* = \iota$, and $E_u(S(u, \xi), u, \xi) = -\iota^*B$, we find that

$$\nabla \widehat{J}_1(u, \xi) = -B^* \iota[z(u, \xi)], \tag{20}$$

where for each $(u, \xi) \in L^2(D_0) \times \Xi$, $z(u, \xi) \in H_0^1(0, 1)$ solves the adjoint equation: find $z \in H_0^1(0, 1)$ with

$$\kappa(\xi)(z', v')_{L^2(0,1)} - (S(u, \xi)z', v)_{L^2(0,1)} = -(S(u, \xi) - y_d, v)_{L^2(0,1)}$$

for all $v \in H_0^1(0, 1)$; cf. [19, 205–206] and [78, p. 83]. Since ι is linear and compact, and B is linear and bounded, the operator $K = -B^* \iota$ is compact [49, Theorem 8.2-5 and p. 427]. We choose $M(u, \xi) = z(u, \xi)$.

We establish Assumption 4. Using [8, Theorem 8.2.9] and the measurability of $S(u, \cdot)$ (see Lemma 1), we can show that $z(u, \cdot)$ is measurable for all $u \in L^2(D_0)$. The

implicit function theorem can be used to show that $z(\cdot, \xi)$ is continuous. Hence z is a Carathéodory mapping. Next, we derive an $H_0^1(0, 1)$ -stability estimate for the state. We have $\|Bu\|_{L^2(0,1)} \leq \|u\|_{L^2(D_0)}$ for all $u \in L^2(D_0)$. Hence the operator norm of B is less than or equal to one. We have $\|v\|_{L^p(0,1)} \leq \|v\|_{H_0^1(0,1)}$ for each $v \in H_0^1(0, 1)$ and $1 \leq p \leq \infty$ [78, Lemma 3.4 on p. 9]. Hence Friedrichs' constant C_D satisfies $C_D \leq 1$. Using integration by parts, we have $(yy', y)_{L^2(0,1)} = 0$ for all $y \in H_0^1(0, 1)$ [78, p. 72]. Choosing $v = S(u, \xi)$ in the weak form of Burgers' equation, we obtain

$$\kappa_{\min} \|S(u, \xi)\|_{H_0^1(0,1)} \leq \|b(\xi)\|_{L^2(0,1)} + \|u\|_{L^2(D_0)}; \tag{21}$$

cf. [78, p. 75]. Next, we establish a stability estimate for $M(u, \xi) = z(u, \xi)$. Combining the $L^\infty(0, 1)$ -stability estimate established in [78, Lemma 3.4 on p. 83] with $(1 + e^{2x})e^x \leq 2e^{3x}$ valid for all $x \geq 0$, we obtain

$$\|z(u, \xi)\|_{L^\infty(0,1)} \leq 2\kappa(\xi)^{-1} e^{3\kappa(\xi)^{-1}\|S(u,\xi)\|_{L^1(0,1)}} \|S(u, \xi) - y_d\|_{L^2(0,1)}. \tag{22}$$

Choosing $v = z(u, \xi)$ in the adjoint equation and using the Hölder and Friedrichs inequalities, and $C_D \leq 1$, we obtain

$$\kappa(\xi) \|z(u, \xi)\|_{H_0^1(0,1)} \leq \|S(u, \xi)\|_{L^2(0,1)} \|z(u, \xi)\|_{L^\infty(0,1)} + \|S(u, \xi) - y_d\|_{L^2(0,1)}. \tag{23}$$

Since U_{ad} is nonempty, there exists $u_0 \in U_{\text{ad}}$. Combined with (21) and the definition of J_1 , we find that $\mathbb{E}[\tilde{J}(u_0, \xi)] < \infty$. Let $B_{V_{\text{ad}}^\rho}(u_0)$ be an open, bounded ball about zero containing $V_{\text{ad}}^\rho(u_0)$ with radius $R_{\text{ad}} > 0$. We define $\zeta_1(\xi) = (1/\kappa_{\min})(\|b(\xi)\|_{L^2(0,1)} + R_{\text{ad}} + \|y_d\|_{L^2(0,1)})$ and

$$\zeta(\xi) = (1/\kappa_{\min})\zeta_1(\xi)((2/\kappa_{\min})\zeta_1(\xi)e^{(3/\kappa_{\min})\zeta_1(\xi)} + 1).$$

Combining (20) and the stability estimates (21), (22) and (23), we conclude that Assumption 4 holds true with ζ being an essentially bounded random variable.

4.3 Distributed Control of a Semilinear State Equation

We consider a distributed control problem with a semilinear state equation based on those considered in [47, Sect. 5] and [45, Sect. 5.2]. Risk-neutral optimization of semilinear PDEs are also studied, for example, in [24, 25]. We refer the reader to [77, Chap. 9] and [76, Chap. 4] for the analysis of deterministic, distributed control problems with semilinear PDEs.

We consider

$$\min_{u \in U_{\text{ad}}} (1/2)\mathbb{E} \left[\|(1 - S(u, \xi))_+\|_{L^2(D)}^2 \right] + (\alpha/2) \|u\|_{L^2(D)}^2, \tag{24}$$

where $(\cdot)_+ = \max\{0, \cdot\}$, $\alpha > 0$, and $U_{\text{ad}} \subset L^2(D)$ is a nonempty, closed, and convex. For each $(u, \xi) \in L^2(D) \times \Xi$, $S(u, \xi) \in H^1(D)$ is the solution to: find $y \in H^1(D)$ with $E(y, u, \xi) = 0$, where the operator $E : H^1(D) \times L^2(D) \times \Xi \rightarrow H^1(D)^*$ is defined by

$$\langle E(y, u, \xi), v \rangle_{H^1(D)^*, H^1(D)} = (\kappa(\xi)\nabla y, \nabla v)_{L^2(D)^2} + (g(\xi)y + y^3, v)_{L^2(D)} - (B(\xi)[u], v)_{L^2(D)} - (b(\xi), v)_{L^2(D)}. \tag{25}$$

Let $\iota : H^1(D) \rightarrow L^2(D)$ be the compact embedding operator of the compact embedding $H^1(D) \hookrightarrow L^2(D)$ [32, Theorem 1.14]. For each $\xi \in \Xi$, we define $B(\xi) = \iota \tilde{B}(\xi) \iota^*$. The operator $\tilde{B}(\xi) : H^1(D)^* \rightarrow H^1(D)$ is the solution operator to a parameterized PDE. For each $(f, \xi) \in H^1(D)^* \times \Xi$, $\tilde{B}(\xi)f \in H^1(D)$ is the solution to: find $w \in H^1(D)$ with

$$(r(\xi)\nabla w, \nabla v)_{L^2(D)^2} + (w, v)_{L^2(D)} = \langle f, v \rangle_{H^1(D)^*, H^1(D)} \quad \text{for all } v \in H^1(D). \tag{26}$$

Since the embedding $H^1(D) \hookrightarrow L^2(D)$ is continuous, the operator ι^* is given by $\langle \iota^*[u], v \rangle_{H^1(D)^*, H^1(D)} = (u, v)_{L^2(D)}$ for all $(u, v) \in L^2(D) \times H^1(D)$ [13, p. 21].

The assumptions stated next ensure the existence and uniqueness of solutions to the PDE defined by the operator in (25) and the well-posedness of the operator $\tilde{B}(\xi)$; see [47, Sects. 3 and 5].

- $D \subset \mathbb{R}^2$ is a bounded Lipschitz domain.
- $\kappa, g : \Xi \rightarrow L^\infty(D)$ are strongly measurable and there exist $\kappa_{\min}, \kappa_{\max}, g_{\min}, g_{\max} \in (0, \infty)$ such that $\kappa_{\min} \leq \kappa(\xi) \leq \kappa_{\max}$ and $g_{\min} \leq g(\xi) \leq g_{\max}$ for all $\xi \in \Xi$.
- $b : \Xi \rightarrow L^2(D)$ and $r : \Xi \rightarrow L^\infty(D)$ are strongly measurable and there exist $b_{\max}, r_{\min}, r_{\max} \in (0, \infty)$ such that $\|b(\xi)\|_{L^2(D)} \leq b_{\max}$ and $r_{\min} \leq r(\xi) \leq r_{\max}$ for all $\xi \in \Xi$.

Throughout the section, we assume these conditions be satisfied.

Assumption 1 is fulfilled since the function $J_1 : H^1(D) \rightarrow [0, \infty)$ defined by $J_1(y) = (1/2) \|(1 - \iota y)_+\|_{L^2(D)}^2$ is continuously differentiable [47, p. 14]. We have $D_y J_1(y) = -\iota^*(1 - \iota[y])_+$. Since $\iota[y] = y$, we have for all $y \in H^1(D)$,

$$\|D_y J_1(y)\|_{H^1(D)^*} \leq \|(1 - y)_+\|_{L^2(D)} \leq \|1\|_{L^2(D)} + \|y\|_{H^1(D)}. \tag{27}$$

For each $\xi \in \Xi$, the operator $E(\cdot, \cdot, \xi)$ is continuously differentiable [47, p. 14] and for each $(u, \xi) \in L^2(D) \times \Xi$, $E_y(S(u, \xi), u, \xi)$ has a bounded inverse [47, p. 9]. Using arguments similar to those in Sect. 4.1, we can show that $E(y, u, \cdot)$ is measurable for each $(y, u) \in H^1(D) \times L^2(D)$. We find that Assumption 2 holds true.

We verify Assumption 4. For each $(u, \xi) \in L^2(D) \times \Xi$, the adjoint state $z(u, \xi) \in H^1(D)$ is the solution to: find $z \in H^1(D)$ with

$$(\kappa(\xi)\nabla z, \nabla v)_{L^2(D)^2} + (g(\xi)z + 3S(u, \xi)^2 z, v)_{L^2(D)} = ((1 - S(u, \xi))_+, v)_{L^2(D)}$$

for all $v \in H^1(D)$. Choosing $v = z(u, \xi)$ and using (27), we obtain the stability estimate

$$\min\{\kappa_{\min}, g_{\min}\} \|z(u, \xi)\|_{H^1(D)} \leq \|1\|_{L^2(D)} + \|S(u, \xi)\|_{H^1(D)}. \tag{28}$$

Moreover, for all $f \in H^1(D)$ and $u \in L^2(D)$, we have the stability estimates (cf. [47, Sects. 3 and 5])

$$\begin{aligned} \min\{r_{\min}, 1\} \|\tilde{B}(\xi)f\|_{H^1(D)} &\leq \|f\|_{H^1(D)^*}, \\ \min\{\kappa_{\min}, g_{\min}\} \|S(u, \xi)\|_{H^1(D)} &\leq \|\iota^*[B(\xi)u] + \iota^*[b(\xi)]\|_{H^1(D)^*}. \end{aligned} \tag{29}$$

Using calculus for adjoint operators [49, p. 235] and $\iota = (\iota^*)^*$, we find that $(\iota^*B(\xi))^* = \iota\tilde{B}(\xi)^*\iota^*$. Consequently, the gradient formula (6) yields

$$\nabla \widehat{J}_1(u, \xi) = -\iota[\tilde{B}(\xi)^*\iota^*\iota z(u, \xi)].$$

We choose $K = -\iota$ and $M(u, \xi) = \tilde{B}(\xi)^*\iota^*\iota z(u, \xi)$. Using the implicit function theorem and [47, Proposition 4.3], we find that z is a Carathéodory mapping. Combined with (29), we obtain that $M(\cdot, \xi)$ is continuous for each $\xi \in \Xi$. Fix $f, v \in H^1(D)^*$. Using [8, Theorem 8.2.9], we can show that $\xi \mapsto \tilde{B}(\xi)f$ is measurable. Hence $\xi \mapsto \langle v, \tilde{B}(\xi)f \rangle_{H^1(D)^*, H^1(D)}$ is measurable [35, Theorem 1.1.6]. Since $\langle \tilde{B}(\xi)^*v, f \rangle_{H^1(D), H^1(D)^*} = \langle v, \tilde{B}(\xi)f \rangle_{H^1(D)^*, H^1(D)}$ for all $\xi \in \Xi$, the mapping $\xi \mapsto \tilde{B}(\xi)^*v$ is measurable [35, Theorem 1.1.6]. Since $H^1(D)$ is separable, $\xi \mapsto \tilde{B}(\xi)^*$ is strongly measurable [35, Theorem 1.1.6]. Combined with the composition rules [35, Proposition 1.1.28 and Corollary 1.1.29], we can show that $M(u, \cdot)$ is measurable.

Using (29) and the fact that U_{ad} is nonempty, we find that there exists $u_0 \in U_{\text{ad}}$ with $\mathbb{E}[\widehat{J}(u_0, \xi)] < \infty$. Let $B_{V_{\text{ad}}^\rho(u_0)}$ be an open, bounded ball about zero containing $V_{\text{ad}}^\rho(u_0)$ and let R_{ad} be its radius. We define the random variable

$$\zeta(\xi) = \frac{1}{\min\{\kappa_{\min}, g_{\min}\}} \left(\|1\|_{L^2(D)} + \frac{\|b(\xi)\|_{L^2(D)} + \frac{1}{\min\{r_{\min}, 1\}} R_{\text{ad}}}{\min\{\kappa_{\min}, g_{\min}\}} \right).$$

Our assumptions and Hölder’s inequality ensure that ζ is integrable.

Combined with the stability estimates (28) and (29), we conclude that Assumption 4 holds true.

5 Discussion

The analysis of the SAA approach for PDE-constrained optimization under uncertainty is complicated by the fact that the feasible sets are generally noncompact, stopping us from directly applying the consistency results developed in the literature on M-estimation and stochastic programming. Inspired by the consistency results in [70,

72], we constructed compact subsets of the feasible set that contain the solutions to the stochastic programs and eventually those to their SAA problems, allowing us to establish consistency of the SAA optimal values and SAA solutions. To construct such compact sets, we combined the adjoint approach, optimality conditions, and PDE stability estimates. We applied our framework to three risk-neutral nonlinear PDE-constrained optimization problems.

We comment on four limitations of our approach. First, our construction of the compact sets exploits the positivity of the regularization parameter α , limiting our approach at first to PDE-constrained optimization problems with strongly convex control regularization. However, we can add $(\alpha/2) \|\cdot\|_{L^2(D)}^2$ with $\alpha > 0$ to the objective function, allowing us to establish the consistency of regularized SAA solutions. If U_{ad} is contained in a ball with radius $r_{\text{ad}} > 0$, $\varepsilon > 0$, and $\alpha = 2\varepsilon/r_{\text{ad}}^2$, then solutions to the regularized SAA problem provide ε -optimal solutions to the non-regularized SAA problem (2).¹ Second, the analysis developed here demonstrates the consistency of SAA optimal values and SAA optimal solutions, but not of SAA critical points. Since the risk-neutral PDE-constrained optimization problems considered here are generally nonconvex, a consistency analysis of SAA critical points would be desirable. However, even though risk-neutral nonlinear PDE-constrained optimization problems and their SAA problems are generally nonconvex, significant progress has been made in establishing convexity properties of nonlinear PDE-constrained optimization problems [22, 31] and in developing verifiable conditions that can be used to certify global optimality of critical points [2]. Third, the construction of the compact subsets performed in Sect. 3 exploits the fact that the feasible set (3) of the SAA problems is the same as that of the risk-neutral problem. Therefore, our approach does not allow for a consistency analysis for SAA problems defined by random constraints, such as those resulting from sample-based approximations of expectation constraints [72, pp. 168–170]. Fourth, our analysis does not apply to risk-averse PDE-constrained optimization problems, as it exploits smoothness of the expectation function. However, our approach may be generalized to allow for the consistency analysis of risk-averse PDE-constrained programs, such as those defined via the superquantile/conditional value-at-risk [43].

Acknowledgements JM thanks Professor Alexander Shapiro for valuable discussions about the SAA approach. The authors thank the two anonymous referees for their helpful comments and suggestions.

Funding The project was partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 188264188/GRK1754.

Declarations

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

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

Appendix A: Lack of Inf-Compactness

Besides the feasible set’s lack of compactness, the set of ε -optimal solutions to the SAA problem and the level sets of the SAA problem’s objective function may be noncompact for risk-neutral PDE-constrained optimization problems. We illustrate this observation on the semilinear PDE-constrained problem

$$\min_{u \in U_{ad}} (1/2)\mathbb{E} \left[\|S(u, \xi)\|_{L^2(D)}^2 \right] + (\alpha/2) \|u\|_{L^2(D)}^2, \tag{30}$$

where $\alpha > 0$, $U_{ad} = \{u \in L^2(D) : \|u\|_{L^2(D)} \leq 2\}$, $D \subset \mathbb{R}^2$ is a bounded Lipschitz domain, and Ξ is as in Sect. 3. For each $(u, \xi) \in L^2(D) \times \Xi$, the state $S(u, \xi) \in H_0^1(D)$ is the solution to the semilinear PDE: find $y \in H_0^1(D)$ with

$$(\kappa(\xi)\nabla y, \nabla v)_{L^2(D)^2} + (y^3, v)_{L^2(D)} = (u, v)_{L^2(D)} \text{ for all } v \in H_0^1(D). \tag{31}$$

We assume that $\kappa : \Xi \rightarrow L^\infty(D)$ is strongly measurable and that there exists $\kappa_{min} > 0$ with $\kappa(\xi) \geq \kappa_{min}$ for all $\xi \in \Xi$. The SAA problem of (30) is given by

$$\min_{u \in U_{ad}} \frac{1}{2N} \sum_{i=1}^N \|S(u, \xi^i)\|_{L^2(D)}^2 + (\alpha/2) \|u\|_{L^2(D)}^2, \tag{32}$$

where ξ^1, ξ^2, \dots , are as in Sect. 3.

Let \hat{F}_N be the objective function of (32) and let C_D be Friedrichs’ constant of the domain D . For each $(u, \xi) \in L^2(D) \times \Xi$, we have the stability estimate (cf. [24, eqns. (2.11)])

$$\|S(u, \xi)\|_{H_0^1(D)} \leq (C_D/\kappa_{min}) \|u\|_{L^2(D)}. \tag{33}$$

The optimal value of the risk-neutral problem (30) and those of the corresponding SAA problems (32) are zero, as $S(0, \xi) = 0$ for all $\xi \in \Xi$ and $0 \in U_{ad}$. We define $\varepsilon_{max} = (C_D^2/\kappa_{min})^2 + \alpha$. Let $\varepsilon > 0$ satisfy $\varepsilon \leq \varepsilon_{max}$. We define $V_\varepsilon = \{u \in L^2(D) : \|u\|_{L^2(D)}^2 \leq \frac{2\varepsilon}{(C_D^2/\kappa_{min})^2 + \alpha}\}$. It holds that $V_\varepsilon \subset U_{ad}$. For each $u \in V_\varepsilon$, the stability estimate (33) and Friedrichs’ inequality yield

$$\begin{aligned} & \frac{1}{2N} \sum_{i=1}^N \|S(u, \xi^i)\|_{L^2(D)}^2 + (\alpha/2) \|u\|_{L^2(D)}^2 \\ & \leq (1/2)(C_D^2/\kappa_{min})^2 \|u\|_{L^2(D)}^2 + (\alpha/2) \|u\|_{L^2(D)}^2 \\ & \leq 0 + \varepsilon. \end{aligned}$$

Hence each $u \in V_\varepsilon$ is an ε -optimal solution to the SAA problem (32). The set V_ε is a closed ball about zero with positive radius because $\varepsilon > 0$. Since $L^2(D)$ is infinite dimensional, this set is noncompact [49, Theorem 2.5-5]. Therefore, as long as $\tilde{\varepsilon} > 0$,

the level sets of the SAA objective function, $\{u \in U_{\text{ad}}: \hat{F}_N(u) \leq \tilde{\varepsilon}\}$, are noncompact, as they contain the noncompact set V_ε with $\varepsilon = \min\{\tilde{\varepsilon}, \varepsilon_{\text{max}}\}$. In this case, an inf-compactness condition (see [72, p. 166]) is violated.

References

1. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
2. Ahmad Ali, A., Deckelnick, K., Hinze, M.: Global minima for semilinear optimal control problems. *Comput. Optim. Appl.* **65**(1), 261–288 (2016). <https://doi.org/10.1007/s10589-016-9833-1>
3. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: Mean-variance risk-averse optimal control of systems governed by PDEs with random parameter fields using quadratic approximations. *SIAM/ASA J. Uncertain. Quantif.* **5**(1), 1166–1192 (2017). <https://doi.org/10.1137/16M106306X>
4. Aliprantis, C.D., Border, K.C.: *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd edn. Springer, Berlin (2006). <https://doi.org/10.1007/3-540-29587-9>
5. Alla, A., Hinze, M., Kolvenbach, P., Lass, O., Ulbrich, S.: A certified model reduction approach for robust parameter optimization with PDE constraints. *Adv. Comput. Math.* **45**(3), 1221–1250 (2019). <https://doi.org/10.1007/s10444-018-9653-1>
6. Alt, H.W.: *Linear Functional Analysis: An Application-Oriented Introduction*. Universitext. Springer, London (2016). <https://doi.org/10.1007/978-1-4471-7280-2>. Translated from the German edition by Robert Nürnberg
7. Artstein, Z., Wets, R.J.B.: Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2**(1/2), 1–17 (1995)
8. Aubin, J.P., Frankowska, H.: *Set-Valued Analysis*. Mod. Birkhäuser Class. Springer, Boston (2009). <https://doi.org/10.1007/978-0-8176-4848-0>
9. Banholzer, D., Fliege, J., Werner, R.: On rates of convergence for sample average approximations in the almost sure sense and in mean. *Math. Program.* **191**(1, Ser. B), 307–345 (2022). <https://doi.org/10.1007/s10107-019-01400-4>
10. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books Mathematics. Springer, New York (2011). <https://doi.org/10.1007/978-1-4419-9467-7>
11. Ben-Tal, A., El Ghaoui, L., Nemirovski, A.: *Robust Optimization*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton (2009)
12. Billingsley, P.: *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, Hoboken (2012)
13. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer, New York (2000). <https://doi.org/10.1007/978-1-4612-1394-9>
14. Casas, E., Tröltzsch, F.: Second order analysis for optimal control problems: improving results expected from abstract theory. *SIAM J. Optim.* **22**(1), 261–279 (2012). <https://doi.org/10.1137/110840406>
15. Castaing, C., Valadier, M.: *Convex Analysis and Measurable Multifunctions*. Lecture Notes in Mathematics, vol. 580. Springer, Berlin (1977)
16. Chen, P., Ghattas, O.: Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters. *SIAM/ASA J. Uncertain. Quantif.* **9**(4), 1381–1410 (2021). <https://doi.org/10.1137/20M1381381>
17. Conti, S., Held, H., Pach, M., Rumpf, M., Schultz, R.: Risk averse shape optimization. *SIAM J. Control Optim.* **49**(3), 927–947 (2011). <https://doi.org/10.1137/090754315>
18. Cucker, F., Smale, S.: On the mathematical foundations of learning. *Bull. Am. Math. Soc. (N.S.)* **39**(1), 1–49 (2002). <https://doi.org/10.1090/S0273-0979-01-00923-5>
19. de los Reyes, J.C., Kunisch, K.: A comparison of algorithms for control constrained optimal control of the Burgers equation. *Calcolo* **41**(4), 203–225 (2004). <https://doi.org/10.1007/s10092-004-0092-7>
20. Farshbaf-Shaker, M.H., Henrion, R., Hömberg, D.: Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. *Set-Valued Var. Anal.* **26**(4), 821–841 (2018). <https://doi.org/10.1007/s11228-017-0452-5>
21. Farshbaf-Shaker, M.H., Gugat, M., Heitsch, H., Henrion, R.: Optimal Neumann boundary control of a vibrating string with uncertain initial data and probabilistic terminal constraints. *SIAM J. Control Optim.* **58**(4), 2288–2311 (2020). <https://doi.org/10.1137/19M1269944>

22. Gahururu, D., Hintermüller, M., Stengl, S.M., Surowiec, T.M.: Generalized Nash equilibrium problems with partial differential operators: Theory, algorithms, and risk aversion. In: Hintermüller, M., Herzog, R., Kanzow, C., Ulbrich, M., Ulbrich, S. (eds.) *Non-Smooth and Complementarity-Based Distributed Parameter Systems: Simulation and Hierarchical Optimization*. Internat. Ser. Numer. Math., vol. 172. Birkhäuser, Cham (2022). https://doi.org/10.1007/978-3-030-79393-7_7
23. 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>
24. Garreis, S., Ulbrich, M.: A fully adaptive method for the optimal control of semilinear elliptic PDEs under uncertainty using low-rank tensors. Preprint, Technische Universität München, München (2019). <http://go.tum.de/204409>
25. 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>
26. Geletu, A., Hoffmann, A., Schmidt, P., Li, P.: Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. *ESAIM Control Optim. Calc. Var.* **26**, Paper No. 70, 28 (2020). <https://doi.org/10.1051/cocv/2019077>
27. 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>
28. 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>
29. Haber, E., Chung, M., Herrmann, F.: An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optim.* **22**(3), 739–757 (2012). <https://doi.org/10.1137/11081126X>
30. Hess, C.: Epi-convergence of sequences of normal integrands and strong consistency of the maximum likelihood estimator. *Ann. Stat.* **24**(3), 1298–1315 (1996). <https://doi.org/10.1214/aos/1032526970>
31. Hintermüller, M., Stengl, S.M.: On the convexity of optimal control problems involving non-linear PDEs or VIs and applications to Nash games. Preprint (2020). <https://doi.org/10.20347/WIAS.PREPRINT.2759>
32. 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>
33. 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>
34. Huber, P.J.: The behavior of maximum likelihood estimates under nonstandard conditions. In: Le Cam, L.M., Neyman, J. (eds.) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1: Statistics, pp. 221–233. University of California Press, Berkeley (1967)
35. Hytönen, T., van Neerven, J., Veraar, M., Weis, L.: *Analysis in Banach Spaces: Martingales and Littlewood-Paley Theory*. *Ergeb. Math. Grenzgeb.* (3) 63. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-48520-1>
36. Kahlbacher, M., Volkwein, S.: Estimation of diffusion coefficients in a scalar Ginzburg-Landau equation by using model reduction. In: Kunisch, K., Of, G., Steinbach, O. (eds.) *Numerical Mathematics and Advanced Applications*, pp. 727–734. Springer, Berlin (2008). https://doi.org/10.1007/978-3-540-69777-0_87
37. Kaniovski, Yu.M., King, A.J., Wets, R.J.B.: Probabilistic bounds (via large deviations) for the solutions of stochastic programming problems. *Ann. Oper. Res.* **56**(1), 189–208 (1995). <https://doi.org/10.1007/BF02031707>
38. Kleywegt, A.J., Shapiro, A., Homem-de Mello, T.: The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* **12**(2), 479–502 (2002). <https://doi.org/10.1137/S1052623499363220>
39. Kolvenbach, P., Lass, O., Ulbrich, S.: An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty. *Optim. Eng.* **19**(3), 697–731 (2018). <https://doi.org/10.1007/s11081-018-9388-3>
40. Kouri, D.P.: A multilevel stochastic collocation algorithm for optimization of PDEs with uncertain coefficients. *SIAM/ASA J. Uncertain. Quantif.* **2**(1), 55–81 (2014). <https://doi.org/10.1137/130915960>

41. Kouri, D.P.: A measure approximation for distributionally robust PDE-constrained optimization problems. *SIAM J. Numer. Anal.* **55**(6), 3147–3172 (2017). <https://doi.org/10.1137/15M1036944>
42. 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
43. Kouri, D.P., Surowiec, T.M.: Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM J. Optim.* **26**(1), 365–396 (2016). <https://doi.org/10.1137/140954556>
44. 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>
45. Kouri, D.P., Surowiec, T.M.: Epi-regularization of risk measures. *Math. Oper. Res.* **45**(2), 774–795 (2020). <https://doi.org/10.1287/moor.2019.1013>
46. 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>
47. Kouri, D.P., Surowiec, T.M.: Risk-averse optimal control of semilinear elliptic PDEs. *ESAIM Control. Optim. Calc. Var.* (2020). <https://doi.org/10.1051/cocv/2019061>
48. Kouri, D.P., Surowiec, T.M.: A primal–dual algorithm for risk minimization. *Math. Program.* **193**(1, Ser. A), 337–363 (2022). <https://doi.org/10.1007/s10107-020-01608-9>
49. Kreyszig, E.: *Introductory Functional Analysis with Applications*. Wiley, New York (1978)
50. Lachout, P., Liebscher, E., Vogel, S.: Strong convergence of estimators as ϵ_n -minimisers of optimisation problems. *Ann. Inst. Stat. Math.* **57**(2), 291–313 (2005). <https://doi.org/10.1007/BF02507027>
51. Lass, O., Ulbrich, S.: Model order reduction techniques with a posteriori error control for nonlinear robust optimization governed by partial differential equations. *SIAM J. Sci. Comput.* **39**(5), S112–S139 (2017). <https://doi.org/10.1137/16M108269X>
52. Le Cam, L.M.: On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Stat.*, vol. 1, pp. 277–329 (1953). <https://hdl.handle.net/2027/wu.89045844305>
53. Mannel, F., Rund, A.: A hybrid semismooth quasi-Newton method for nonsmooth optimal control with PDEs. *Optim. Eng.* **22**(4), 2087–2125 (2021). <https://doi.org/10.1007/s11081-020-09523-w>
54. 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>
55. Martínez-Frutos, J., Esparza, F.P.: *Optimal Control of PDEs Under Uncertainty: An Introduction with Application to Optimal Shape Design of Structures*. SpringerBriefs Math. Springer, Cham (2018). <https://doi.org/10.1007/978-3-319-98210-6>
56. Meidner, D., Vexler, B.: A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part II: Problems with control constraints. *SIAM J. Control Optim.* **47**(3), 1301–1329 (2008). <https://doi.org/10.1137/070694028>
57. Milz, J.: *Topics in PDE-Constrained Optimization under Uncertainty and Uncertainty Quantification*. Dissertation, Technische Universität München, München (2021)
58. Milz, J.: Sample average approximations of strongly convex stochastic programs in Hilbert spaces. *Optim. Lett.* (2022). <https://doi.org/10.1007/s11590-022-01888-4>
59. Milz, J., Ulbrich, M.: An approximation scheme for distributionally robust PDE-constrained optimization. *SIAM J. Control Optim.* **60**(3), 1410–1435 (2022). <https://doi.org/10.1137/20M134664X>
60. Nasir, Y., Volkov, O., Durlafsky, L.J.: A two-stage optimization strategy for large-scale oil field development. *Optim. Eng.* (2021). <https://doi.org/10.1007/s11081-020-09591-y>
61. Nečas, J.: *Direct Methods in the Theory of Elliptic Equations*. Springer Monogr. Math. Springer, Heidelberg (2012). <https://doi.org/10.1007/978-3-642-10455-8>
62. Phelps, C., Royset, J., Gong, Q.: Optimal control of uncertain systems using sample average approximations. *SIAM J. Control Optim.* **54**(1), 1–29 (2016). <https://doi.org/10.1137/140983161>
63. 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/resolver.pl?urn:nbn:de:bvb:91-diss-20150420-1241413-1-4>
64. Polak, E.: *Optimization: Algorithms and Consistent Approximations*. Appl. Math. Sci., vol. 124. Springer, New York (1997). <https://doi.org/10.1007/978-1-4612-0663-7>

65. Renardy, M., Rogers, R.C.: An Introduction to Partial Differential Equations, 2nd edn. Texts Appl. Math. 13. Springer, New York, NY (2004). <https://doi.org/10.1007/b97427>
66. Römisch, W., Surowiec, T.M.: Asymptotic properties of Monte Carlo methods in elliptic PDE-constrained optimization under uncertainty. Preprint (2021). <https://arxiv.org/abs/2106.06347>
67. 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>
68. Shapiro, A.: Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* **30**(1), 169–186 (1991). <https://doi.org/10.1007/BF02204815>
69. 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>
70. Shapiro, A.: Monte Carlo sampling methods. In: *Stochastic Programming, Handbooks in Oper. Res. Manag. Sci.*, vol. 10, pp. 353–425. Elsevier, Amsterdam (2003). [https://doi.org/10.1016/S0927-0507\(03\)10006-0](https://doi.org/10.1016/S0927-0507(03)10006-0)
71. 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>
72. Shapiro, A., Dentcheva, D., Ruszczyński, A.: *Lectures on Stochastic Programming: Modeling and Theory*, 2nd edn. MOS-SIAM Ser. Optim. SIAM, Philadelphia (2014). <https://doi.org/10.1137/1.9781611973433>
73. 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
74. Tiesler, H., Kirby, R.M., Xiu, D., Preusser, T.: Stochastic collocation for optimal control problems with stochastic PDE constraints. *SIAM J. Control Optim.* **50**(5), 2659–2682 (2012). <https://doi.org/10.1137/110835438>
75. Tong, S., Subramanyam, A., Rao, V.: Optimization under rare chance constraints. *SIAM J. Optim.* **32**(2), 930–958 (2022). <https://doi.org/10.1137/20M1382490>
76. Tröltzsch, F.: *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. Grad. Stud. Math., vol. 112. AMS, Providence (2010). <https://doi.org/10.1090/gsm/112>. Translated by J. Sprekels
77. 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>
78. Volkwein, S.: Mesh-Independence of an Augmented Lagrangian-SQP Method in Hilbert Spaces and Control Problems for the Burgers Equation. Dissertation, Technical University of Berlin, Berlin (1997). <https://imsc.uni-graz.at/volkwein/diss.ps>
79. Volkwein, S.: Application of the augmented Lagrangian-SQP method to optimal control problems for the stationary Burgers equation. *Comput. Optim. Appl.* **16**(1), 57–81 (2000). <https://doi.org/10.1023/A:1008777520259>
80. Volkwein, S.: Mesh-independence for an augmented Lagrangian-SQP method in Hilbert spaces. *SIAM J. Control Optim.* **38**(3), 767–785 (2000). <https://doi.org/10.1137/S0363012998334468>
81. Wechsung, F., Giuliani, A., Landreman, M., Cerfon, A.J., Stadler, G.: Single-stage gradient-based stellarator coil design: stochastic optimization. *Nucl. Fusion* **62**(7), 076034 (2022). <https://doi.org/10.1088/1741-4326/ac45f3>
82. Yang, H., Gunzburger, M.: Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling. *Comput. Methods Appl. Mech. Eng.* **319**, 217–239 (2017). <https://doi.org/10.1016/j.cma.2017.02.030>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.