Numerische Mathematik manuscript No.
(will be inserted by the editor)

Asymptotic Properties of Monte Carlo Methods in Elliptic PDE-Constrained Optimization under Uncertainty

W. Römisch · **T. M. Surowiec**

the date of receipt and acceptance should be inserted later

Abstract Monte Carlo approximations for random linear elliptic PDE constrained optimization problems are studied. We use empirical process theory to obtain best possible mean convergence rates $O(n^{-\frac{1}{2}})$ for optimal values and solutions, and a central limit theorem for optimal values. The latter allows to determine asymptotically consistent confidence intervals by using resampling techniques. The theoretical results are illustrated with two sets of numerical experiments. The first demonstrates the theoretical convergence rates for optimal values and optimal solutions. This is complemented by a study illustrating the usage of subsampling bootstrap methods for estimating the confidence intervals.

Keywords random elliptic PDE, stochastic optimization, Monte Carlo, central limit theorem, resampling

Mathematics Subject Classification (2010) 49J20, 49J55, 60F17, 65C05, 90C15, 35R60

1 Introduction

PDE-constrained optimization under uncertainty is a rapidly growing field with a number of recent contributions in theory [9,33,34,36], numerical and computational methods [19,20,32,61], and applications [7,8, 11,51]. Nevertheless, a number of open questions remain unanswered, even for the ideal setting including a strongly convex objective function, a closed, bounded and convex feasible set, and a linear elliptic PDE with random inputs.

Broadly speaking, the numerical solution methods available for such problems derive in part from the standard paradigms found in the classical stochastic programming literature: first-order stochastic approximation/stochastic gradient (SG) approaches, [20], versus methods that rely on sampling from the underlying probability measure [61]. The latter approaches are not optimization algorithms per se, rather, they use approximations of the expectations by replacing the underlying probability distribution with a discrete measure. This may be obtained either from available data or using Monte Carlo (MC) type methods. The main advantage of the latter is that we may turn to the wide array of powerful, function-space-based numerical methods for PDE-constrained optimization in a deterministic setting, see e.g., [28].

Nevertheless, as with SG-based methods, the optimal values and solutions of MC-type approximations must also be understood as realizations of a rather complex random process. In this context, it is helpful to think of them as mappings from some space of probability measures into the reals (optimal values) and decision space (solutions). In optimization under uncertainty, stability usually refers to the continuity properties of these mappings with respect to changes in the underlying measure. As the underlying parameter space is not a normed linear space, proving continuity and asymptotic properties as $n \to \infty$ can be a delicate matter. Such statements require techniques not typically employed in PDE-constrained optimization, e.g., empirical process theory or the method of probability metrics. Of course, if we can obtain computable quantitative bounds in n, i.e., convergence rates, then such stability statements can provide us with a priori information regarding the necessary sample size for an MC-based numerical solution method. These can in turn be linked to the PDE discretization error for a comprehensive a priori error estimate.

In a recent paper, we provided a number of qualitative and quantitative stability statements for infinite-dimensional stochastic optimization problems using the method of probability metrics [30]. However, for the PDE-constrained optimization problem provided in [30, Sec. 7], a major open question remained: Can we derive a reasonable rate of convergence for the minimal information metric that supports our numerical

Institute of Mathematics, Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany, E-mail: romisch@math.hu-berlin.de, · Simula Research Laboratory, Kristian Augusts gate 23, 0164 Oslo, Norway. E-mail: thomasms@simula.no.

results? Taking this question as a starting point, we seek to answer this by using deep results from empirical process theory as detailed in [23,64]. The idea to use empirical process theory has been employed in the stochastic programming literature before, cf. some results in [50, Chapters 6–8]. However, it has not been used in situations where the decision spaces are infinite dimensional, which presents an additional challenge. Another approach based on large deviation-type results is employed in the recent preprint [39] in which the author obtained results that are in parts similar those in the present paper. However, the results in [39] cannot be used to derive confidence intervals for the optimal values without further assumptions on the integrands.

Convergence statements based on the underlying smoothness of the uncertainty in the forward-problem have been previously considered in [37]. However, these results do not in fact subsume those found in the present article. There are a number of significant differences. First, the authors in [37] do not place additional constraints on the control variables. Therefore, the first-order optimality conditions are coupled systems of PDEs. Second, the authors use this fact to transform the question of convergence rates of the optimization problem into convergence rates for the PDE system. Following this, they allow the controls to also depend on the uncertainty, but they do not include an additional non-anticipativity constraint of the type: $u(\sigma) = \mathbb{E}_\sigma[u]$ for all parameters σ. Therefore, the rates no longer apply to the original optimization problem. As a consequence, our results appear to be the first of their kind for PDE constrained optimization under uncertainty.

The rest of the article is organized as follows. In Section 2 we introduce the PDE constrained optimization problem (5) with random parameters studied in this article. In addition, we consider a suitable distance measure for probability distributions. After discussing its basic properties, we recall some results from [30] on continuity properties of infima and solutions to the stochastic optimization with respect to such probability metrics. Finally, we prove a new result on Lipschitz continuity for solutions. Section 3 contains the main results on convergence rates for infima and solutions of Monte Carlo approximations and a central limit theorem for infima. All results are consequences of empirical process theory. In Section 4 we shortly describe how subsampling methods can be used to complement the central limit result by deriving confidence intervals for the optimal values. Finally, in Section 5, we provide a detailed example of how the theoretical results manifest themselves in practice. This is split into two main subsections. In the first part, we study the convergence rates predicted by the theory for optimal values and optimal solutions. This is followed by a discussion on the construction of subsampling bootstrapped confidence intervals and finally, a numerical illustration. We close the paper by discussing the limitations and possible extensions of our results.

2 PDE constrained optimization under uncertainty

We start by introducing several function spaces used throughout our study. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space with σ-finite measure and Y be a linear space endowed with a norm $\|\cdot\|_Y$. By $L_p(\Omega, \mathcal{A}, \mu; Y)$, $1 \le p \le \infty$, we denote the linear space of strongly measurable functions $y : \Omega \to Y$ such that the integral $\int_\Omega \|y(\omega)\|_Y^p d\mu$ is finite. As usual we equip this space with the norm

$$\|y\|_p = \left(\int_\Omega \|y(\omega)\|_Y^p d\mu \right)^{\frac{1}{p}}$$

for $p < \infty$ with the usual modification for $p = \infty$. In case $V = \mathbb{R}$ we will omit V and write $L_p(\Omega, \mathcal{A}, \mu)$. If Ω is a subset of some Euclidean space $\mathbb{R}^r$ and $\mu = \lambda$ the Lebesgue measure, we will shortly write $L_p(\Omega)$. If Ω is a subset of a metric space and $\mathcal{A}$ the Borel σ-fileld, we will omit $\mathcal{A}$. For any set Y we denote by $\ell^\infty(Y)$ the linear space of bounded real-valued functions g defined on Y endowed with the norm $\|g\|_\infty = \sup_{y \in Y} |g(y)|$.

Now, let $D \subset \mathbb{R}^m$ be an open bounded domain with Lipschitz boundary and $V = H_0^1(D)$ the usual Sobolev space of (equivalence classes of) functions in $L_2(D)$ that admit square integrable weak derivatives. We endow this space with the inner product $(u, v)_V = \int_D \nabla u \cdot \nabla v \, dx$ and norm $\|u\|_V = \sqrt{(u, u)_V}$. The topological dual is denoted by $V^\star = H^{-1}(D)$ with the usual operator norm $\|\cdot\|_\star$. In addition, we consider the Hilbert space $H = L^2(D)$ with the inner product $(g, h)_H = \int_D g(x) h(x) \, dx$. The dual pairing for $V, V^\star$ is denoted by $\langle \cdot, \cdot \rangle$. Let Ξ be a metric space and $\mathcal{P}(\Xi)$ be the set of all Borel probability measures on Ξ. Fix $P \in \mathcal{P}(\Xi)$. For the parametric PDE, we first define the bilinear form $a(\cdot, \cdot; \xi) : V \times V \to \mathbb{R}$

$$a(u, v; \xi) = \int_D \sum_{i,j=1}^m a_{ij}(x, \xi) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} \, dx \tag{1}$$

for each $\xi \in \Xi$. Here, we impose the condition that the functions $a_{ij} : D \times \Xi \to \mathbb{R}$ are measurable on $D \times \Xi$ and there exist $L > \gamma > 0$ such that

$$\gamma \sum_{i=1}^m y_i^2 \le \sum_{i,j=1}^m a_{ij}(x, \xi) y_i y_j \le L \sum_{i=1}^m y_i^2 \quad (\forall y \in \mathbb{R}^n) \tag{2}$$

for a.e. $x \in D$ and any $\xi \in \Xi$. This implies that each a_{ij} is essentially bounded on $D \times \Xi$ in both arguments with respect to the product measure $\lambda \times P$.

We consider the stochastic optimization problem: Minimize the functional

$$\begin{aligned}\mathcal{J}(u,z) &:= \frac{1}{2}\int_{\Xi}\int_{D} |u(x,\xi) - \widetilde{u}(x)|^2\,dx\,dP(\xi) + \frac{\alpha}{2}\int_{D} |z(x)|^2\,dx \\ &= \frac{1}{2}\mathbb{E}_P[\|u(\cdot) - \widetilde{u}\|_H^2] + \frac{\alpha}{2}\|z\|_H^2 \end{aligned} \tag{3}$$

subject to $(u,z) \in L_2(\Xi,P;V) \times Z_{\mathrm{ad}}$, where $\alpha > 0$, $\widetilde{u} \in H$, Z_{ad} denotes some closed convex bounded subset of H and $u(\cdot)$ solves the random elliptic PDE

$$a(u(\xi),v;\xi) = \int_D (z(x) + g(x,\xi))v(x)\,dx \tag{4}$$

for P-a.e. $\xi \in \Xi$ and all test functions $v \in V$, where $g : D \times \Xi \to \mathbb{R}$ is measurable on $D \times \Xi$ and $g(\cdot,\xi) \in H$ for each $\xi \in \Xi$.

For P-a.e. $\xi \in \Xi$ we define the mapping $A(\xi) : V \to V^\star$ by means of the Riesz representation theorem

$$\langle A(\xi)u, v\rangle = a(u,v;\xi) \quad (u,v \in V).$$

Consequently $A(\xi)$ is linear, uniformly positive definite (with $\gamma > 0$) and uniformly bounded (with $L > 0$) and the random elliptic PDE may be written in operator form

$$A(\xi)u = z + g(\xi) \quad (P\text{-a.e. } \xi \in \Xi).$$

In addition, the inverse mapping $A(\xi)^{-1} : V^\star \to V$ exists, and is linear, uniformly positive definite (with modulus L^{-1}) and uniformly bounded (with constant γ^{-1}). This allows us to rewrite the stochastic optimization problem in reduced form over $z \in Z_{\mathrm{ad}}$ as:

$$\min\left\{F_P(z) = \int_{\Xi} f(z,\xi)\,dP(\xi) : z \in Z_{\mathrm{ad}}\right\} \tag{5}$$

with the integrand

$$\begin{aligned} f(z,\xi) &= \frac{1}{2}\left\|A(\xi)^{-1}(z + g(\xi)) - \widetilde{u}\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \\ &= \frac{1}{2}\left\|A(\xi)^{-1}z - (\widetilde{u} - A(\xi)^{-1}g(\xi))\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \end{aligned} \tag{6}$$

for any $z \in H$ and P-a.e. $\xi \in \Xi$, where $g \in L_2(\Xi,P;H)$ and $A(\xi)^{-1}$ as defined earlier. For each P-a.e. $\xi \in \Xi$ the function $f(\cdot,\xi) : H \to \mathbb{R}$ is convex and continuous. For later use we denote the optimal value of (5) by $v(P)$.

We will need a few properties of the function $F_P : H \to \mathbb{R}$. They are collected in the following result which is partly proved in [30].

Proposition 1 *For each $P \in \mathcal{P}(\Xi)$ the functional F_P is finite, continuous, and strongly convex on H and, hence, weakly lower semicontinuous on the weakly compact set Z_{ad}. Moreover, there exists an unique minimizer $z(P) \in Z_{\mathrm{ad}}$ of* (5) *and the objective function F_P has quadratic growth around $z(P)$, i.e. we have*

$$\|z - z(P)\|_H^2 \le \frac{8}{\alpha}(F_P(z) - F_P(z(P))) = \frac{8}{\alpha}(F_P(z) - v(P)) \quad (\forall z \in Z_{\mathrm{ad}}). \tag{7}$$

In addition, F_P is Gâteaux differentiable on H with Gâteaux derivative $F_P'(\cdot)$ and the estimate

$$|F_P(z) - F_P(\tilde{z})| \le \sup_{t\in[0,1]} \|F_P'(z + t(\tilde{z} - z))\|\,\|z - \tilde{z}\|_H \tag{8}$$

holds for all $z, \tilde{z} \in H$.

Proof While the first part is proved in [30], it remains to prove the Gâteaux differentiability of F_P and the estimate (8). For any $z, w \in H$ we observe that the real function $h(t) = F_P(z + tw)$ is quadratic for $t \in \mathbb{R}$ and, hence, differentiable. This means that F_P is Gâteaux differentiable at z. Now, we set $w = \tilde{z} - z$ for some $\tilde{z} \in H$. Since a differentiable function $h : [0,1] \to \mathbb{R}$ satisfies the estimate

$$|h(1) - h(0)| \le \sup_{t\in[0,1]} |h'(t)|,$$

we obtain for $h(t) = F_P(z + t(\tilde{z} - z))$ that

$$h'(t) = F_P'(z + t(\tilde{z} - z))(\tilde{z} - z) = \int_{\Xi} f_z'(z + t(\tilde{z} - z), \xi)(\tilde{z} - z)P(d\xi) \tag{9}$$

for all $t \in [0, 1]$, where $f_z'(\cdot, \cdot)$ denotes the partial Gâteaux derivative of f with respect to the first variable. A straightforward evaluation shows that

$$f_z'(z, \xi)(w) = \langle A(\xi)^{-1}(z + g(\xi)) - \tilde{u}, A(\xi)^{-1}w\rangle_H + \alpha\langle z, w\rangle_H$$

holds for all $z, w \in H$, $\xi \in \Xi$. This proves the desired estimate (8) and shows that its right-hand side is finite.□

Motivated by (5) and (7) we consider the pseudo-metric

$$d_{\mathfrak{F}}(P, Q) = \sup_{f \in \mathfrak{F}} \left| \int_{\Xi} f(\xi)\, dP(\xi) - \int_{\Xi} f(\xi)\, dQ(\xi) \right| \tag{10}$$

on $\mathcal{P}(\Xi)$ for studying quantitative stability of (5) with respect to perturbations of the underlying probability distribution P. Here, $\mathfrak{F}$ is a class of real-valued Borel measurable functions on Ξ. The notion of pseudo-metric means that all properties of metrics are satisfied except that $d_{\mathfrak{F}}(P, Q) = 0$ does not imply $P = Q$ in general, unless $\mathfrak{F}$ is sufficiently rich. These are the typical properties required for probability metrics (see [46]). Such distances of probability measures were first introduced and studied in [66]. A number of important probability metrics are of the form $d_{\mathfrak{F}}$, for example, the bounded Lipschitz metric and the Fortet-Mourier metrics for which $\mathfrak{F}$ contains (locally) Lipschitz functions. In both cases the class $\mathfrak{F}$ is rich enough and, in addition, convergence with respect to $d_{\mathfrak{F}}$ implies the weak convergence of probability measures. We recall that a sequence (P_n) in $\mathcal{P}(\Xi)$ converges weakly to P iff

$$\lim_{n\to\infty} \int_{\Xi} f(\xi)\, dP_n(\xi) = \int_{\Xi} f(\xi)\, dP(\xi)$$

holds for all bounded continuous functions $f : \Xi \to \mathbb{R}$. Compared with classical probability metrics we consider here a much smaller class $\mathfrak{F}$ of functions, namely, the collection of all integrands in (5)

$$\mathfrak{F}_{mi} = \{f(z, \cdot) : z \in Z_{\mathrm{ad}}\}. \tag{11}$$

Following [47] we call $d_{\mathfrak{F}_{mi}}$ the problem-based or minimal information (m.i.) distance. Convergence with respect to such distances does not imply weak convergence in general, but the question arises whether it is implied by the weak convergence of probability measures. It is known that a positive answer depends on analytical properties of the class $\mathfrak{F}$. The following result is classical and due to [59].

Lemma 1 *If $\mathfrak{F}$ is uniformly bounded and it holds that*

$$P(\{\xi \in \Xi : \mathfrak{F} \textit{ is not equicontinuous at } \xi\}) = 0,$$

then the set $\mathfrak{F}$ is a so-called P-uniformity class, i.e., weak convergence of (P_n) to P implies

$$\lim_{n\to\infty} d_{\mathfrak{F}}(P_n, P) = 0.$$

The choice (11) of $\mathfrak{F}$ leads to the following result proved in [30].

Theorem 1 *Under the standing assumptions and with the class $\mathfrak{F}_{mi}$ in* (11) *we obtain the estimates*

$$|v(Q) - v(P)| \le d_{\mathfrak{F}_{mi}}(P, Q) \tag{12}$$

$$\|z(Q) - z(P)\|_H \le 2\sqrt{\frac{2}{\alpha}} d_{\mathfrak{F}_{mi}}(P, Q)^{\frac{1}{2}} \tag{13}$$

for the optimal value $v(P)$ and solution $z(P)$ of (5) *if the original probability distribution P is perturbed by any $Q \in \mathcal{P}(\Xi)$.*

Next we collect some properties of the class $\mathfrak{F}_{mi}$ and of its elements under Lipschitz continuity assumptions on the coefficients of the linear elliptic PDE and of its right-hand sides implying that $\mathfrak{F}_{mi}$ is a P-uniformity class (proved in [30]).

Theorem 2 *Assume that all functions $a_{ij}(x, \cdot)$, $i, j = 1, \ldots, m$, and $g(x, \cdot)$ are Lipschitz continuous on Ξ uniformly with respect to $x \in D$, and let $g \in L_\infty(\Xi, P; H)$. Then the family $\mathfrak{F}_{mi} = \{f(z, \cdot) : z \in Z_{\mathrm{ad}}\}$ is uniformly bounded and Lipschitz continuous on Ξ (with a constant not depending on z). In particular, $\mathfrak{F}_{mi}$ is a P-uniformity class.*

We close this section by extending the Hölder stability result (13) in Theorem 1 to Lipschitz stability with respect to a pseudo-metric of the type (10), but with a class $\mathfrak{F}_{di}$ of functions different from $\mathfrak{F}_{mi}$. For deriving the Lipschitz stability result we do not make use of classical work like, e.g., [1,13], but exploit the fact that (5) is formulated as an optimization problem with fixed constraint set. Our methodology exploits the quadratic growth condition of F_P (see Proposition 1) and partly parallels that of [54, Lemma 2.1].

Theorem 3 *Under the standing assumptions the Lipschitz-type estimate*

$$\|z(Q) - z(P)\|_H \leq \frac{8}{\alpha} d_{\mathfrak{F}_{di}}(P, Q) \tag{14}$$

holds for all $P, Q \in \mathcal{P}(\Xi)$, where $\mathfrak{F}_{di}$ denotes the following function class on Ξ

$$\mathfrak{F}_{di} = \left\{ \langle A(\cdot)^{-1}(z + g(\cdot)) - \tilde{u}, A(\cdot)^{-1}h \rangle_H + \alpha\langle z, h\rangle_H : z \in Z_{\mathrm{ad}}, \|h\|_H \leq 1 \right\}. \tag{15}$$

Proof Let $P, Q \in \mathcal{P}(\Xi)$ and $z(P), z(Q) \in Z_{\mathrm{ad}}$ the corresponding solutions to (5). From Proposition 1 we know that F_P has quadratic growth around $z(P)$, i.e.,

$$\begin{aligned}\frac{\alpha}{8}\|z(Q) - z(P)\|_H^2 &\leq F_P(z(Q)) - F_P(z(P)) \\ &\leq (F_P(z(Q)) - F_Q(z(Q))) - (F_P(z(P)) - F_Q(z(P))),\end{aligned}$$

where we added $F_Q(z(P)) - F_Q(z(Q)) \geq 0$ to the right-hand side. Now we consider the function $h : [0,1] \to \mathbb{R}$ given by $h(t) = (F_P - F_Q)(z(P) + t(z(Q) - z(P)))$, $t \in [0,1]$. Due to Proposition 1 F_P and F_Q are Gâteaux differentiable on H with Gâteaux derivatives F_P' and F_Q'. Hence, h is differentiable on $[0,1]$ and it holds that $|h(1) - h(0)| \leq \sup_{t\in[0,1]} |h'(t)|$. This implies

$$\begin{aligned}(F_P - F_Q)(z(Q)) - (F_P - F_Q)(z(P)) &\leq \sup_{z\in Z_{\mathrm{ad}}} |(F_P' - F_Q')(z)(z(Q) - z(P))| \\ &\leq \sup_{z\in Z_{\mathrm{ad}}} \|(F_P' - F_Q')(z)\|\|z(Q) - z(P)\|_H\end{aligned}$$

We obtain after dividing by $\|z(Q) - z(P)\|_H$

$$\begin{aligned}\frac{\alpha}{8}\|z(Q) - z(P)\|_H &\leq \sup_{z\in Z_{\mathrm{ad}}} \|(F_P' - F_Q')(z)\| \\ &\leq \sup_{z\in Z_{\mathrm{ad}}} \sup_{\|h\|_H\leq 1} \left| \int_\Xi f_z'(z,\xi)(h)\, d(P - Q)(\xi) \right|,\end{aligned}$$

where $f_z'(\cdot,\cdot)$ is the partial Gâteaux derivative of f with respect to the first variable given in (9). This completes the proof. □

Remark 1 An inspection of the proof of Theorem 2 (see [30, Section 6]) reveals that the class $\mathfrak{F}_{di}$ of (partial) derivatives of integrands in $\mathfrak{F}_{mi}$ is also a P-uniformity class under the assumptions of Theorem 2.

3 Monte Carlo approximations

Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be independent identically distributed (iid) Ξ-valued random variables on some complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ having the common distribution P, i.e., $P = \mathbb{P}\xi_1^{-1}$. We consider the empirical measures

$$P_n(\cdot) = \frac{1}{n}\sum_{i=1}^n \delta_{\xi_i(\cdot)} \qquad (n \in \mathbb{N}), \tag{16}$$

where δ_ξ denotes the Dirac measure, which places mass 1 at $\xi \in \Xi$ and mass 0 elsewhere. Based on empirical measures we study the sequence of empirical or Monte Carlo approximations of the stochastic program (5) with sample size n, i.e.,

$$\min\left\{ \int_\Xi f(z,\xi)\, dP_n(\cdot)(\xi) = \frac{1}{n}\sum_{i=1}^n f(z, \xi_i(\cdot)) : z \in Z_{\mathrm{ad}} \right\}. \tag{17}$$

The optimal value $v(P_n(\cdot))$ of (17) is a real random variable and the solution $z(P_n(\cdot))$ an H-valued random element (see [6, Lemma III.39]).

Qualitative and quantitative results on the asymptotic behavior of optimal values and solutions to (17) are known in finite-dimensional settings (see [16], and the surveys [55] and [43]). Since the sequence $(P_n(\cdot))$ of empirical measures converges weakly to P $\mathbb{P}$-almost surely, one obtains the following corollary by combining Lemma 1 and Theorems 1 and 2.

Corollary 1 *The sequences $(v(P_n(\cdot)))$ and $(z(P_n(\cdot)))$ of empirical optimal values and solutions converge $\mathbb{P}$-almost surely to the true optimal values and solutions $v(P)$ and $z(P)$, respectively.*

In this section we are mainly interested in quantitative results on the asymptotic behavior of $v(P_n(\cdot))$ and $z(P_n(\cdot))$. This is closely related to uniform convergence properties of the empirical process

$$\left\{\mathbb{G}_n(\cdot)f := \sqrt{n}(P_n(\cdot) - P)f = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(\xi_i(\cdot)) - Pf)\right\}_{f\in\mathfrak{F}} \tag{18}$$

indexed by a class $\mathfrak{F}$ of real-valued measurable functions on Ξ and, hence, to quantitative estimates of

$$\|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}} = \sup_{f\in\mathfrak{F}}|\mathbb{G}_n(\cdot)f| = \sqrt{n}\, d_{\mathfrak{F}}(P_n(\cdot), P) = \sqrt{n}\sup_{f\in\mathfrak{F}}|P_n(\cdot)f - Pf|. \tag{19}$$

Here, we set $Pf = \int_\Xi f(\xi)\mathrm{d}P(\xi)$ for any probability distribution P and any $f \in \mathfrak{F}$. Since the supremum in (19) is taken with respect to an uncountable set $\mathfrak{F}$, it is not necessarily measurable with respect to $\mathcal{F}$. For a detailed discussion of measurability issues we refer to [57]. For the function classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$, however, the situation is very comfortable.

Lemma 2 *For the classes $\mathfrak{F} = \mathfrak{F}_{mi}$ and $\mathfrak{F} = \mathfrak{F}_{di}$ the mapping $\|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}}$ in* (19) *is measurable.*

Proof If there exists a countable set $\mathfrak{F}_0$ contained in a function class $\mathfrak{F}$ such that

$$\|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}} = \|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}_0} \tag{20}$$

holds, the mapping $\|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}}$ from Ω to $\mathbb{R}$ is measurable since it coincides with the supremum of a countable set of measurable functions. For the class $\mathfrak{F} = \mathfrak{F}_{mi}$ we obtain for $f \in \mathfrak{F}_{mi} = \{f(z,\cdot) : z \in Z_{\mathrm{ad}}\}$ that

$$P_n(\cdot)f - Pf = \frac{1}{n}\sum_{i=1}^{n} f(z, \xi_i(\cdot)) - Pf(z,\cdot)$$

holds for some $z \in Z_{\mathrm{ad}}$. Next we restrict the condition $z \in Z_{\mathrm{ad}}$ by $z \in Z_0$ for some countable subset $Z_0 \subseteq Z_{\mathrm{ad}}$ which is dense in Z_{ad} with respect to $\|\cdot\|_H$. Such a countable set Z_0 exists due to the separability of H. Then the continuity of f with respect to the first component implies that (20) is valid for $\mathfrak{F}_0 = \{f(z,\cdot) : z \in Z_0\}$. For the class $\mathfrak{F} = \mathfrak{F}_{di}$ we argue analogously by using $\mathfrak{F}_0 = \{f'_z(z,\cdot)(h) : z \in Z_0, h \in B_0\}$, where $f'_z(\cdot,\cdot)$ is the partial Gâteaux derivative of f with respect to the first variable and B_0 denotes some countable dense subset of the unit ball in H. □

There exist two main approaches to derive quantitative information on the asymptotic behavior of empirical processes. The first consists in the use of concentration inequalities (pioneered in [58] and presented in some detail in [5]) with applications to bounding (19) in probability. The second relies on the notion of Donsker classes of functions with applications to limit theorems. In this paper we study the second approach.

A collection $\mathfrak{F}$ of measurable functions on Ξ is called *P-Donsker* if the empirical process (18) converges in distribution to a tight random variable $\mathbb{G}$ in the space $\ell^\infty(\mathfrak{F})$, where the limit process $\mathbb{G} = \{\mathbb{G}f : f \in \mathfrak{F}\}$ is a zero-mean Gaussian process with the covariance function

$$E_{\mathbb{P}}[\mathbb{G}f_1\mathbb{G}f_2] = P[(f_1 - Pf_1)(f_2 - Pf_2)] \quad (f_1, f_2 \in \mathfrak{F}). \tag{21}$$

The limit $\mathbb{G}$ is sometimes called a P-Brownian bridge process in $\ell^\infty(\mathfrak{F})$.

Remark 2 We will prove that $\mathfrak{F} = \mathfrak{F}_{mi}$ and $\mathfrak{F} = \mathfrak{F}_{di}$ are P-Donsker classes by showing that $\sqrt{n}\mathbb{E}[d_{\mathfrak{F}}(P_n(\cdot), P)]$ is bounded (see Proposition 2). From this we deduce the following mean convergence rates

$$\mathbb{E}[d_{\mathfrak{F}}(P_n(\cdot), P)] = O(n^{-\frac{1}{2}}). \tag{22}$$

Together with Theorems 1 and 3 this then leads to best possible mean convergence rates of Monte Carlo estimates for optimal values and solutions.

Whether $\mathfrak{F}$ satisfies the P-Donsker class property, depends on its size measured in terms of so-called bracketing or metric entropy numbers. To introduce these concepts, let $\mathfrak{F}$ be a subset of the linear normed space $L_p(\Xi, P)$ (for some $p \geq 1$) (of equivalence classes) of measurable functions endowed with the norm

$$\|f\|_{P,p} = (P|f|^p)^{\frac{1}{p}} = \left(\int_\Xi |f(\xi)|^p\, dP(\xi)\right)^{\frac{1}{p}}.$$

Given a pair of functions $l, u \in L_p(\Xi, P)$, $l \leq u$, a bracket $[l, u]$ is defined by $[l, u] = \{f \in L_p(\Xi, P) : l \leq f \leq u\}$. Given $\varepsilon > 0$ the bracketing number $N_{[\,]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{P,p})$ is the minimal number of brackets with $\|l - u\|_{P,p} < \varepsilon$ needed to cover $\mathfrak{F}$. The metric entropy number with bracketing of $\mathfrak{F}$ is defined by

$$H_{[\,]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{P,p}) = \log N_{[\,]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{P,p}).$$

Both numbers are finite if $\mathfrak{F}$ is a totally bounded subset of $L_p(\Xi, P)$. A powerful result on empirical processes is the following (see [62, Theorem A.2]).

Proposition 2 *There exists a universal constant $C > 0$ such that for any class $\mathfrak{F}$ of measurable functions with envelope function $\hat{F}$ (i.e., $|f| \leq \hat{F}$ for every $f \in \mathfrak{F}$) belonging to $L_2(\Xi, P)$ the estimate*

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathfrak{F}}] \leq C \int_0^1 \sqrt{1 + H_{[\,]}(\varepsilon \|\hat{F}\|_{P,2}, \mathfrak{F}, \|\cdot\|_{P,2})}\, d\varepsilon \, \|\hat{F}\|_{P,2} \tag{23}$$

holds. If the integral in (23) *is finite, then the class $\mathfrak{F}$ is P-Donsker.*

Note that the integral in (23) can only be finite if $H_{[\,]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{P,2})$ grows at most like $\varepsilon^{-\beta}$ with $0 < \beta < 2$ for $\varepsilon \to +0$.
Next we discuss the assumption of finiteness of the integral in (23) in case that $\mathfrak{F}$ is a bounded subset of classical linear normed spaces of smooth functions.

Example 1 Let $\Xi \subset \mathbb{R}^d$ be convex, bounded with the property $\Xi \subseteq \mathrm{cl}\,\mathrm{int}\,\Xi$, $k \in \mathbb{N}_0$ and $r \in [0, 1]$. We consider the linear space $C^{k,r}(\Xi)$ of real functions on Ξ having partial derivatives up to order k such that all kth order derivatives are Hölder continuous with exponent r. Next we use the notation $\mathbf{i} = (i_1, \ldots, i_d)$ with $i_j \in \mathbb{N}_0$, $j = 1, \ldots, d$, and $|\mathbf{i}| = \sum_{j=1}^d i_j$. Further, $D^{\mathbf{i}} f$ denotes

$$D^{\mathbf{i}} f = \frac{\partial^{|i|} f}{\partial \xi_1^{i_1} \cdots \partial \xi_d^{i_d}} \qquad (f \in C^{k,r}(\Xi), |\mathbf{i}| \leq k).$$

If the spaces are endowed with the norms

$$\|f\|_{k,0} = \max_{|\mathbf{i}| \leq k} \sup_{\xi} |D^{\mathbf{i}} f(\xi)|$$

$$\|f\|_{k,r} = \max_{|\mathbf{i}| \leq k} \sup_{\xi} |D^{\mathbf{i}} f(\xi)| + \max_{|\mathbf{i}| = k} \sup_{\xi \neq \tilde{\xi}} \frac{|D^{\mathbf{i}} f(\xi) - D^{\mathbf{i}}(\tilde{\xi})|}{\|\xi - \tilde{\xi}\|^r} \quad (r > 0),$$

where the suprema are taken over all $\xi, \tilde{\xi}$ in the interior of Ξ, they become Banach spaces. The metric entropy with bracketing of balls $\mathbb{B}_{k,r}(\rho)$ around the origin with radius ρ in $C^{k,r}(\Xi)$ is computed in [31] with respect to the uniform norm $\|\cdot\|_{0,0} = \|\cdot\|_\infty$. The authors show that there exists a constant $K > 0$ depending only on d, k, r, ρ and the diameter of Ξ such that we have for every $\varepsilon > 0$

$$H_{[\,]}(\varepsilon\rho, \mathbb{B}_{k,r}(\rho), \|\cdot\|_{P,2}) \leq K \varepsilon^{-\frac{d}{k+r}}, \tag{24}$$

where the result from [31] was adapted to the norm in $L_2(\Xi, P)$ (see also [64, Section 2.7.1]). Hence, Proposition 2 can be utilized to show that bounded subsets of $C^{k,r}(\Xi)$ are P-Donsker if $\frac{d}{2} < k + r$. For the situation studied in Theorem 2 with $k = 0$ and $r = 1$ this means that bounded subsets of $C^{0,1}(\Xi)$ are P-Donsker only for $d = 1$. Without imposing stronger smoothness conditions on the coefficients in the bilinear form a (see (1)) and the right-hand side compared to Theorem 2, the integral in (23) will not be finite. Hence, one cannot use Proposition 2 to conclude that $\mathfrak{F}$ is P-Donsker.

Remark 3 Indeed a convergence rate for the sequence $(\mathbb{E}[d_{\mathfrak{F}}(P_n(\cdot), P)])$ as in (22) cannot be achieved if $\mathfrak{F}$ is the unit ball in $C^{0,1}(\Xi)$ for $d > 1$. Then $d_{\mathfrak{F}}$ coincides with the Wasserstein metric W_1 and the Fortet-Mourier metric ζ_1 of order 1 (see also [30, Section 4]). Namely, it is shown in [12,18] that the Wasserstein distance W_p of P and $P_n(\cdot)$ has the mean convergence rate

$$\mathbb{E}[W_p(P_n(\cdot), P)] = O(n^{-\frac{1}{d}}) \tag{25}$$

if $d > 2$, $p \geq 1$ and sufficiently high moments of P exist. This rate carries over to the mean convergence rate of Fortet-Mourier metrics ζ_p and of the bounded Lipschitz metric β which represents a lower bound of ζ_1 (see also [14] for the mean convergence rate of the sequence $(\mathbb{E}[\beta(P_n(\cdot), P)])$).

Next we derive conditions implying that the functions in $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$, respectively, are sufficiently smooth. A similar result on differentiability of solutions to random PDEs is proved in [10, Section 4] in a different way.

Theorem 4 *Let $\Xi \subset \mathbb{R}^d$ be a bounded convex set having the property $\Xi \subseteq \mathrm{cl\ int}\,\Xi$ and let $k \in \mathbb{N}$. Let the assumptions of Theorem 2 be satisfied and assume that, for all $u, v \in V$, the functions $\langle A(\cdot)u, v\rangle$ and $\langle g(\cdot), v\rangle$ belong to $C^{k,0}(\Xi)$. Then both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ are subsets of $C^{k,0}(\Xi)$.*

Proof The integrands f belonging to $\mathfrak{F}_{mi}$ are of the form (see (6))

$$f(z,\xi) = \frac{1}{2}\,\|u(\xi) - \widetilde{u}\|_H^2 + \frac{\alpha}{2}\|z\|_H^2\,,$$

where $u(\xi) = A(\xi)^{-1}(z + g(\xi))$, $\xi \in \Xi$. We begin by showing that, for $w, y \in V^\star$ the mapping $\xi \mapsto \langle y, A(\xi)^{-1}w\rangle$ has first order partial derivatives at any $\xi \in \mathrm{int}\,\Xi$. We fix $\xi \in \mathrm{int}\,\Xi$, $j \in \{1,\dots,d\}$, and a canonical basis vector $e_j \in \mathbb{R}^d$. Then $\xi + he_j \in \mathrm{int}\,\Xi$ for sufficiently small $|h| > 0$ and we obtain for $w, y \in V^\star$

$$\begin{aligned}\frac{1}{h}\langle y, (A(\xi + he_j)^{-1} - A(\xi)^{-1})w\rangle &= \frac{1}{h}\langle y, A(\xi)^{-1}(A(\xi) - A(\xi + he_j))u(h)\rangle\\ &= \frac{1}{h}(\langle \triangle_A^j(\xi;h)u, v\rangle - \langle (\triangle_A^j(\xi;h))^\star v, u_0 - u_h\rangle),\end{aligned}$$

where $\triangle_A^j(\xi;h) = A(\xi + he_j) - A(\xi)$, $u_h = A(\xi + he_j)^{-1}w$, $u_0 = A(\xi)^{-1}w$, $v = (A(\xi)^{-1})^\star y \in V$ and $(A(\xi)^{-1})^\star$ denotes the adjoint mapping to $A(\xi)^{-1}$. While the first summand on the right-hand side converges for $h \to 0$ to the partial derivative $\frac{\partial}{\partial \xi_j}$ of $\langle A(\cdot)u, v\rangle$ at ξ, the second converges to zero as the left-hand side of the pairing remains bounded and the right-hand side $u_0 - u_h$ converges to zero. Hence, the partial derivative $\frac{\partial}{\partial \xi_j}$ of $\langle y, A(\cdot)^{-1}w\rangle$ exists at ξ and it holds

$$\frac{\partial}{\partial \xi_j}\langle y, A(\xi)^{-1}w\rangle = \frac{\partial}{\partial \xi_j}\langle A(\cdot)u, v\rangle \quad (\text{at } \xi).$$

This identity also shows that $\langle y, A(\cdot)^{-1}w\rangle$ is continuously differentiable. The differentiability of $\langle y, A(\cdot)^{-1}g(\cdot)\rangle$ follows in a straightforward way via the product rule. Hence, we conclude that the partial derivative $\frac{\partial}{\partial \xi_j}f(z,\cdot)$ exists for any $z \in Z_{\mathrm{ad}}$. By the same reasoning we can inductively derive the existence of higher order mixed partial derivatives $D^{\mathbf{i}}f(z,\cdot)$ at ξ for $|\mathbf{i}| \leq k$ and any $z \in Z_{\mathrm{ad}}$. We conclude that both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ are subsets of $C^{k,0}(\Xi)$. □

Remark 4 According to the definition of the mapping $A(\xi) : V \to V^\star$ we have

$$\langle A(\xi)u, v\rangle = \sum_{i,j=1}^{m}\int_D a_{ij}(x,\xi)\frac{\partial u(x)}{\partial x_i}\frac{\partial v(x)}{\partial x_j}\,dx \tag{26}$$

for all pairs $(u, v) \in V$. Due to the uniform ellipticity condition (2) we know that all functions a_{ij} are essentially bounded. If we assume that all functions $a_{ij}(x,\cdot) : \Xi \to \mathbb{R}$, $x \in D$, have continuous mixed partial derivatives up to order k which are in addition all measurable and essentially bounded on $D \times \Xi$, one obtains mixed partial derivatives of $\langle A(\cdot)u, v\rangle$ by differentiating equation (26).
The same is true for $\langle g(\cdot), v\rangle$ if the functions $g(x,\cdot)$, $x \in D$, have continuous mixed partial derivatives up to order k which are all measurable and essentially bounded on $D \times \Xi$.

In order to make use of Example 1 we present conditions implying that both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ are bounded subsets of $C^{k,0}(\Xi)$.

Theorem 5 *Let $\Xi \subset \mathbb{R}^d$ be a bounded, convex set having the property $\Xi \subseteq \mathrm{cl\ int}\,\Xi$ and let $k \in \mathbb{N}$ be such that $d < 2k$. Assume that all functions $a_{ij}(x,\cdot) : \Xi \to \mathbb{R}$, $i,j = 1,\dots,m$, and $g(x,\cdot) : \Xi \to \mathbb{R}$, $x \in D$, have continuous partial derivatives up to order k which are all measurable and essentially bounded on $D \times \Xi$. Then the classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ are P-Donsker, it holds that*

$$\mathbb{E}[|v(P_n(\cdot)) - v(P)|] = O(n^{-\frac{1}{2}}) \tag{27}$$

$$\mathbb{E}[\|z(P_n(\cdot)) - z(P)\|_H] = O(n^{-\frac{1}{2}}) \tag{28}$$

and the sequence $(\sqrt{n}(v(P_n(\cdot)) - v(P)))$ converges in distribution to a normal random variable with mean zero and variance $P(f(z(P)))^2$. Here, $v(P)$ and $z(P)$ are the optimal value and solution of (5), and $v(P_n(\cdot))$ and $z(P_n(\cdot))$ are the optimal value and solution of (17), respectively.

Proof Our assumptions together with Theorem 4 imply that both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ represent bounded subsets of the Banach space $C^{k,0}(\Xi)$. Hence, according to Example 1 the metric entropy with bracketing of any of the two classes satisfies

$$H_{[\,]}(\varepsilon\rho, \mathfrak{F}, \|\cdot\|_{P,2}) \leq K\varepsilon^{-\frac{d}{k}}$$

for some constant $K > 0$. Since Ξ is bounded, the estimate (23) in Proposition 2 implies

$$\mathbb{E}[\sqrt{n} d_{\mathfrak{F}}(P_n(\cdot), P)] \leq \hat{C} \int_0^1 \varepsilon^{-\frac{d}{k}} d\varepsilon$$

for some $\hat{C} > 0$. Since $\frac{d}{k} < 2$, the right-hand side is bounded and we have that

$$\mathbb{E}[d_{\mathfrak{F}}(P_n(\cdot), P)] = O(n^{-\frac{1}{2}})$$

holds for $\mathfrak{F} = \mathfrak{F}_{mi}$ and $\mathfrak{F} = \mathfrak{F}_{di}$. Hence, we obtain (27) from Theorem 1 and (28) from Theorem 3. Furthermore, we conclude for $\mathfrak{F} = \mathfrak{F}_{\mathrm{mi}}$ from Proposition 2 that the empirical process $\{\mathbb{G}_n(\cdot)f = \sqrt{n}(P_n(\cdot) - P)f\}_{f\in\mathfrak{F}}$ converges in distribution to a tight random variable $\{\mathbb{G}f\}_{f\in\mathfrak{F}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in the space $\ell^\infty(\mathfrak{F})$. Due to the structure (11) of $\mathfrak{F}$, we may also write $\{Pf\}_{f\in\mathfrak{F}} = \{Pf(z,\cdot) : z \in Z_{\mathrm{ad}}\}$ and $\{\mathbb{G}f\}_{f\in\mathfrak{F}} = \{\mathbb{G}f(z,\cdot)\}_{z\in Z_{\mathrm{ad}}}$. This means that Pf may be considered as element of the space $\ell^\infty(Z_{\mathrm{ad}})$ of bounded real-valued functions on Z_{ad}. Correspondingly, $\mathbb{G}f$ may be viewed as random variable in $\ell^\infty(Z_{\mathrm{ad}})$. It remains to utilize the functional delta theorem (see [53] or [49]) for the infimal mapping

$$\Phi : \ell^\infty(Z_{\mathrm{ad}}) \to \mathbb{R}, \quad \Phi(h) = \inf_{z\in Z_{\mathrm{ad}}} h(z).$$

The mapping Φ is finite, concave, hence, directionally differentiable on $\ell^\infty(Z_{\mathrm{ad}})$. In addition, Φ is Lipschitz continuous with respect to $\|\cdot\|_\infty$ on Z_{ad} (with modulus 1) and, hence, Hadamard directionally differentiable (see [52]). The Hadamard directional derivative at $h_0 \in \ell^\infty(Z_{\mathrm{ad}})$ is of the form (see [38] and the survey [49])

$$\Phi'_{h_0}(h) = \lim_{\varepsilon\downarrow 0} \inf\{h(z) : z \in S(h_0, \varepsilon)\} \quad (h \in \ell^\infty(Z_{\mathrm{ad}})), \tag{29}$$

where $S(h_0, \varepsilon) = \{z \in Z_{\mathrm{ad}} : h_0(z) \leq \Phi(h_0) + \varepsilon\}$ denotes the ε-solution set of h_0. If h_0 is weakly lower semicontinuous, h_0 attains its infimum on Z_{ad}. Hence, its solution set $S(h_0) = S(h_0, 0)$ is nonempty. In that case, it holds

$$\Phi'_{h_0}(h) \leq \inf\{h(z) : z \in S(h_0)\} \quad (h \in \ell^\infty(Z_{\mathrm{ad}})).$$

If h is also weakly lower semicontinuous and a sequence (z_n) with $z_n \in S(h_0, \frac{1}{n})$ is chosen such that

$$\inf\left\{h(z) : z \in S(h_0, \frac{1}{n})\right\} \leq h(z_n) \leq \inf\left\{h(z) : z \in S(h_0, \frac{1}{n})\right\} + \frac{1}{n},$$

one can select a subsequence (z_{n_k}) of (z_n) that converges weakly to some $z_0 \in S(h_0)$ and it holds

$$\Phi'_{h_0}(h) = \lim_{k\to\infty} h(z_{n_k}) \geq h(z_0) \geq \inf\{h(z) : z \in S(h_0)\}.$$

Hence, we obtain

$$\Phi'_{h_0}(h) = \inf\{h(z) : z \in S(h_0)\}.$$

Finally, we set $h_0(z) = Pf(z,\cdot)$ and $h(z) = \mathbb{G}f(z,\cdot)$. Since both functions are convex and continuous, they are weakly lower semicontinuous. In addition, we know that the solution set of h_0 is a singleton, namely, $S(h_0) = \{z(P)\}$. Hence, we have

$$\Phi'_{h_0}(h) = h(z(P)).$$

Then the functional delta theorem [53, Theorem 2.1] (or [49, Theorem 1]) implies that

$$\sqrt{n}(\Phi(P_n(\cdot)f) - \Phi(Pf))) \xrightarrow{d} \Phi'_{Pf}(\mathbb{G}f) = \mathbb{G}f(z(P)), \tag{30}$$

where $\xrightarrow{d}$ denotes convergence in distribution of real random variables. Hence, the sequence $(\sqrt{n}(v(P_n(\cdot)) - v(P)))$ converges in distribution to the normal random variable $\mathbb{G}f(z(P))$ (with mean zero and variance given by $P(f(z(P)))^2$ see (21)). This completes the proof. □

Remark 5 Note that in the definition of $h(z)$ at the end of the previous proof, we have $\mathbb{G}(\cdot)f = \int_{\mathbb{R}^d} f(.,\xi)\mathbb{G}(\cdot)(d\xi)$, where $\cdot$ indicates the dependence on ω, which is the space for the corresponding empirical central limit theorem. As such, $\mathbb{G}(\cdot)$ plays the same role as P in Pf.

To establish an extension of Theorem 5 to $\Xi = \mathbb{R}^d$ let $\mathbb{R}^d = \bigcup_{j=1}^{\infty} \Xi_j$ be a partition of $\mathbb{R}^d$, where each set Ξ_j is bounded, convex and has the property $\Xi_j \subseteq \mathrm{cl}\,\mathrm{int}\,\Xi_j$, $j \in \mathbb{N}$. The idea is to apply Theorem 5 on each subset Ξ_j of $\mathbb{R}^d$ and then to apply the argument in [62, Theorem 1.1] (see also [62, Corollary 2.1]).

Corollary 2 *Let $k \in \mathbb{N}$ be such that $d < 2k$. Assume that all functions $a_{ij}(x,\cdot) : \mathbb{R}^d \to \mathbb{R}$, $i,j = 1,\ldots,m$, and $g(x,\cdot) : \mathbb{R}^d \to \mathbb{R}$, $x \in D$, have continuous partial derivatives up to order k which are all measurable on $D \times \Xi$. Moreover, assume that for each $j \in \mathbb{N}$ the restrictions to Ξ_j of all functions in both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ belong to the ball $\mathbb{B}_{k,0}(\rho_j)$ in $C^{k,0}(\Xi_j)$ and that the probability measure P satisfies*

$$\sum_{j=1}^{\infty} \rho_j P(\Xi_j)^{\frac{1}{2}} < \infty. \tag{31}$$

Then both classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ are P-Donsker and (27), (28)*, and the central limit theorem for optimal values remain true.*

We note that condition (31) represents a quite implicit link between the growth of derivatives of the functions in both classes with the tail behaviour of P.

Remark 6 Note that Theorem 5 allows to derive asymptotically consistent confidence intervals for optimal values by using resampling techniques such as bootstrapping [24] or subsampling [45]. Since the Hadamard directional derivative is linear in the direction in our case, the classical bootstrap can be used. The subsampling method is more generally applicable than the bootstrap, since the linearity property is not required.

4 Subsampling

The subsampling method [45] is based on sampling and resampling, but resampling is performed repeatedly without replacement and with a lower sample size $b = b(n) \in \mathbb{N}$, $b \ll n$. For some sufficiently large n, let $\xi_1,\ldots,\xi_n$ be an iid sample from P. Let P_n be the empirical measure and $v(P_n)$ the corresponding optimal value of (17). Based on the samples $\xi_{n_1},\ldots,\xi_{n_b}$ drawn from $\{1,\ldots,n\}$ with cardinality b, we consider the corresponding empirical measure

$$P^*(n_1,\ldots,n_b) = \frac{1}{b}\sum_{i=1}^{b} \delta_{\xi_{n_i}}$$

and the optimal value $v(P^*(n_1,\ldots,n_b))$. The subsampling method estimates the limit distribution of $\zeta = \Psi'_{Pf}(\mathbb{G}f)$ (see (30)) based on both optimal values. It is justified by the limit theorem [44, Theorem 2.1] which reads in our framework

$$\binom{n}{b}^{-1} \sum_{1\le n_1<\ldots<n_b\le n} \delta_{\left\{\sqrt{b}\left(v(P^*(n_1,\ldots,n_b))-v(P_n)\right)\right\}} \xrightarrow{d} \zeta \tag{32}$$

for $b,n \to \infty$ and $b/n \to 0$. The number of summands in (32) becomes extremely large as n and b grow. However, the result remains valid if a number $m = m(n)$ is chosen and the sum over all possible subsets is replaced by the sum over m randomly chosen subsets of $\{1,\ldots,n\}$ of cardinality b: Let $N_j^{n,b} \subset \{1,\ldots,n\}$ be randomly chosen with cardinality $\#N_j^{n,b} = b$ for $j = 1,\ldots,m$. Then with $P_n^*(N_j^{n,b})$ denoting the empirical measure based on $\{\xi_i : i \in N_j^{n,b}\}$ we have

$$\frac{1}{m}\sum_{j=1}^{m} \delta_{\left\{\sqrt{b}\left(v(P_n^*(N_j^{n,b}))-v(P_n)\right)\right\}} \xrightarrow{d} \zeta \tag{33}$$

for $b,n,m \to \infty$ and $b/n \to 0$ [44, Corollary 2.1].

This suggests the following procedure to determine confidence intervals for $v(P)$: Given $n,b,m \in \mathbb{N}$, $b < n$, sufficiently large and a sample $\xi_1,\xi_2,\ldots,\xi_n$ from P. Compute $v(P_n)$. Resample from P_n without replacement with sample size $b < n$ to obtain $\{\xi_i : i \in N_j^{n,b}\}$. Compute $v(P_n^*(N_j^{n,b}))$ and repeat this m times. Let

$$L_m(t) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\left\{\sqrt{b}\left(v(P_n^*(N_j^{n,b}))-v(P_n)\right)\le t\right\}} \quad (t \in \mathbb{R}).$$

Choose $\alpha \in (0,1)$ and calculate the quantile $\zeta^*_{1-\alpha,m} = \inf\{t : L_m(t) \ge 1-\alpha\}$ of L_m. Then we obtain for the asymptotic coverage probability of $v(P)$

$$\lim_{n,m\to\infty} \mathbb{P}\left\{\sqrt{n}(v(P_n) - v(P)) \le \zeta^*_{1-\alpha,m}\right\} \ge 1-\alpha.$$

5 Numerical Experiments

In order to illustrate the theoretical results, we provide several numerical experiments. For simplicity, we restrict the spatial domain to be the two dimensional unit square and construct examples that are flexible in the number of random variables and smoothness in the coefficients.

5.1 Experimental Rate of Convergence

We provide here both plots and experimental rates of convergence using a simple logistic regression scheme.

5.1.1 Example Problem

Set $D = (0,1)^2$ and $\Xi = [0,1]^d$, where $d = 2^q$, $q \in \mathbb{N}_0$. Next, partition the interval $[0,1]$ into d closed intervals D_i $(i = 1,\dots,d)$ of the form

$$D_i = [i/d, (i+1)/d]$$

let χ_i be the associated characteristic function, i.e., $\chi_i(x)$ is 1 if $x \in D_i$ and 0 otherwise.

To define the random coefficients inside the differential operator, we first define the mapping $\hat{a} : D \times \mathbb{R}^d \to \mathbb{R}$ by

$$\hat{a}(x,\eta) = \sum_{i=1}^{d} (\eta_i + 10^{-2}x_2 + 10^{-3})\chi_i(x_1).$$

where $x \in D$ and $\eta \in \mathbb{R}^d$. Next, define $w_k : \mathbb{R}^d \to \mathbb{R}^d$ for each $k = 0,1,2,\dots$ by

$$(w_k(\xi))_j = (\xi_j - 10^{-1})^k \max\{0, \xi_j - 10^{-1}\} \quad j = 1,\dots,d,$$

where $\xi \in \Xi \subset \mathbb{R}^d$. Using w_k, we define the parametric coefficient mapping $a_k : D \times \Xi \to \mathbb{R}$ by setting

$$a_k(x,\xi) = \hat{a}(x, w_k(\xi))$$

For a fixed smoothness parameter $k \in \{0,1,2,\dots,\}$ and every $i,j \in \{1,\dots,m\}$ and $\xi \in \Xi$, we will use

$$a_{ij}(x,\xi) = a_k(x,\xi).$$

Note that if $k = 0$, then $a_{ij}(x,\xi)$ is only Lipschitz in ξ. However, for $k = 1,2,\dots$ and fixed $x \in D$, $a_k(x,\dot{)}$ is in $C^k(\Xi)$. In light of these modelling choices, we consider the bilinear form

$$a_k(u,v;\xi) = \int_D a_k(x,\xi)\nabla u \cdot \nabla v \,\mathrm{d}x,$$

where $u,v \in H_0^1(D)$, $\xi \in \Xi$.

For the right hand side of the differential equation, we introduce an additional parametric dependency in the form

$$g(x,\nu) = \sin(2x_1)\sin(2x_2) + 10^{-2}\nu$$

where $(x,y) \in \Omega$ and $\nu \in \mathbb{R}$. For readability, ν is understood to be part of the vector ξ and Ξ will be henceforth a subset of $\mathbb{R}^{d+1}$. Given $z \in L^2(\Omega)$, we then define the random elliptic PDE in weak form as in the main text:

$$a_k(u(\xi), v;\xi) = \int_D (z(x) + g(x,\xi))v\,dx, \tag{34}$$

where the solution and test functions are assumed to satisfy $u(\xi), v \in H_0^1(\Omega)$. For three randomly chosen vectors $\xi_1,\xi_2,\xi_3 \in \Xi$ and $z \equiv 0$ and we plot the resulting solutions in Figure 1 to highlight the significant spacial variability in the uncontrolled solution.

For this example, we again define the objective function by $\mathcal{J}(u,z)$ as

$$\mathcal{J}(u,z) := \frac{1}{2}\int_\Xi \int_D |u(x,\xi) - \widetilde{u}(x)|^2\,dx\,dP(\xi) + \frac{\alpha}{2}\int_D |z(x)|^2\,dx$$

We set $\alpha > 0$ and $\tilde{u} \equiv 1/2$. Whereas comparatively smaller values of α generally affect the number of semismooth Newton iterations to reach a desired tolerance, the choice of u_d appears to have little effect on the performance. A constant function also allows us to, at least visually, judge the quality of the optimal solution. We define the control constraints by

$$Z_{\mathrm{ad}} := \left\{ v \in L^2(D)\,|{-3/4} \le v \le 3/4 \text{ a.e. } D \right\}.$$

Again, the choice of bounds is merely for illustration.

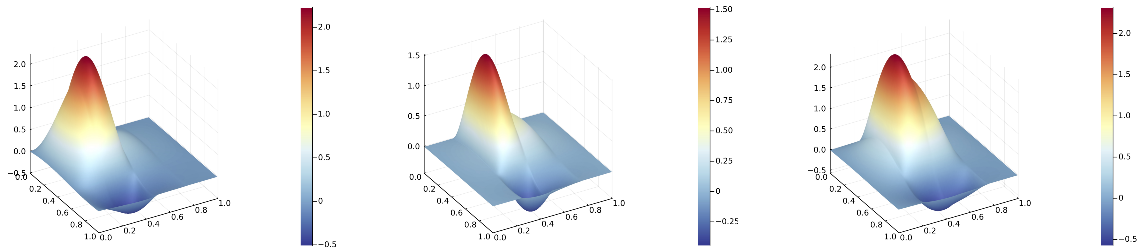


Fig. 1: Uncontrolled, random states: Three realizations of $u(\xi)$ computed by setting $z \equiv 0$ in (34).

5.1.2 Solution Algorithm

We describe the solution algorithm used in the statistical experiments here. For fixed $n \in \mathbb{N}$, we let $\xi_1, \xi_2, \ldots$ be iid Ξ-valued random variables on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with common distribution $P = \mathbb{P} \circ \xi_1^{-1}$. For the numerical experiments, the random parameters in a_k are assumed to be uniformly distributed on $[0,1]^d$ and the single random variable in g is uniformly distributed on $[-1,1]$.

For the numerical solution of (34), we use a standard H^1-conforming P1-finite element discretization, where the domain D is triangulated by a uniform mesh rule. The theoretical optimality conditions indicate that the optimal solutions $z(P)$, $z(P_n)$ share the same regularity as the adjoint variable, which incidentally is also in $H^1_0(D)$. Therefore, we discretize the control variables in the same way as the state and adjoint state.

The favorable structure of the optimization problem lends itself nicely to the application of a semismooth Newton/primal-dual active set strategy to compute $z(P)$ and $z(P_n)$, cf. [29] or [60] . We recall several details of this method here for the fully continuous setting.

The unique optimal solution z satisfies the nonsmooth equation in $L^2(D)$

$$z - \mathrm{Proj}_{Z_{\mathrm{ad}}}(z - c(\mathbb{E}_P[\Lambda(z)] + \alpha z)) = 0. \tag{35}$$

Here, $c > 0$ is any positive constant, $\mathrm{Proj}_{Z_{\mathrm{ad}}}$ is the metric projection on Z_{ad}, and $\Lambda(z)$ is the adjoint state mapping. For fixed ξ, $\lambda(\xi) := \Lambda(z,\xi)$ satisfies

$$\lambda(\xi) = A^{-*}(\xi)(-\mathcal{J}'_u(A^{-1}(\xi)(z + g(\xi)), z)) = -A^{-1}(\xi)(A^{-1}(\xi)(z + g(\xi)) - \tilde{u}).$$

Under the standing assumptions, we can rewrite (35) as

$$\mathcal{F}(z) := z - \min\{3/4, \max\{-3/4, z - c(\mathbb{E}_P[\Lambda(z)] + \alpha z)\}\} = 0. \tag{36}$$

It is straightforward to argue then that $\mathbb{E}_P[\Lambda(\cdot)]$ is smooth from $L^2(D)$ into $L^p(D)$ with $p > 2$. Moreover, taking $c = \alpha^{-1}$, we can remove z from inside the projection operator and finally, we can argue that the nonsmooth superposition operator composed with $\mathbb{E}_P[\Lambda(\cdot)]$ satisfies the necessary “norm-gap” condition that allows us to apply a semismooth Newton method in function space. Given $z_k \in L^2(D)$ such that $\|\mathcal{F}(z_k)\|_{L^2} \neq 0$, the step calculation for solving (36) amounts to finding $\delta z \in L^2(D)$ such that

$$\mathcal{G}(z_k)\delta z = -\mathcal{F}(z_k),$$

where (pointwise a.e.) we have

$$[\mathcal{G}(z_k)\delta z](x) = \begin{cases} \delta z(x) & \text{if } z_k(x) - c(\mathbb{E}_P[\Lambda(z)](x) + \alpha z_k(x)) > 3/4 \\ \delta z(x) & \text{if } z_k(x) - c(\mathbb{E}_P[\Lambda(z)](x) + \alpha z_k(x)) < -3/4 \\ c(\mathbb{E}_P[\Lambda'(z_k)\delta z] + \alpha \delta z) & \text{else.} \end{cases}$$

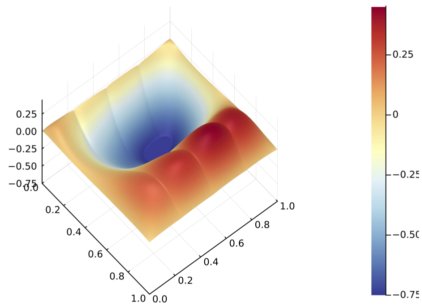
Consequently, the structure of the residual mapping $\mathcal{F}(z)$ provides the following reduced iteration:

1. Given $z_k \in L^2(D)$ such that $\|\mathcal{F}(z_k)\|_{L^2} > 0$ compute

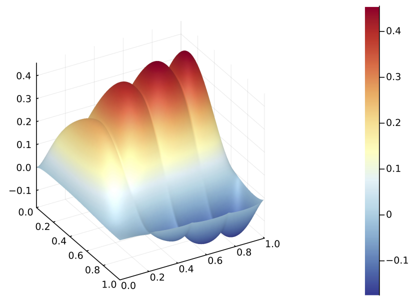
$$\begin{aligned} \mathcal{A}^1_k &:= \{x \in D \,|\, z_k(x) - c(\mathbb{E}_P[\Lambda(z)](x) + \alpha z_k(x)) > 3/4\} \\ \mathcal{A}^2_k &:= \{x \in D \,|\, z_k(x) - c(\mathbb{E}_P[\Lambda(z)](x) + \alpha z_k(x)) < -3/4\} \\ \mathcal{I}_k &:= D \setminus (\mathcal{A}^1_k \cup \mathcal{A}^2_k) \end{aligned}$$

2. Set

$$\begin{aligned} \delta z_k &:= 3/4 - z_k \quad && \text{on } \mathcal{A}^1_k, \\ \delta z_k &:= -3/4 - z_k \quad && \text{on } \mathcal{A}^2_k. \end{aligned}$$



(a) Optimal control $z(P_n)$.



(b) Average controlled state.

Fig. 2: Optimal solution and average states: Figure 2a shows the optimal control computed for a random sample of size $n = 500$ on a uniform mesh with 16129 degrees of freedom. Figure 2b shows the effect this optimal control $z(P_n)$ on the state variables $u(\xi)$ for a new sample of size 500 by computing $\frac{1}{n}\sum_{i=1}^{n} A(\xi_i)^{-1}(z(P_n) + g(\xi_i))$.

3. Solve for $\delta z_k|_{\mathcal{I}_k}$:

$$\mathbb{E}_P[\Lambda'(z_k)\delta z|_{\mathcal{I}_k}]|_{\mathcal{I}_k} + \alpha\delta z|_{\mathcal{I}_k} = -[\mathbb{E}_P[\Lambda(z_k)] + \alpha z_k]|_{\mathcal{I}_k} - [\mathbb{E}_P[\Lambda'(z_k)\delta z|_{\mathcal{A}_k^1\cup\mathcal{A}_k^2}]|_{\mathcal{I}_k}$$

4. Set $z_{k+1} := z_k + t_k\delta z_k$ for some damping factor $t_k > 0$ (e.g. $t_k = 1$).

In Figure 2, we have plotted the result applying this scheme to our problem with $\alpha = 1$, $n = 500$ and a mesh defined by 128×128 grid. This corresponds to 16129 degrees of freedom for the control variables z and approximately 8 million degrees of freedom for the state variables associated with the 500 elliptic PDEs. The average controlled state is much closer to the desired state of $\tilde{u} \equiv 0.5$ than observed in the uncontrolled states in Figure 1. We also note that the active set $\mathcal{A}_k^2$ is nonempty and appears to have positive Lebesgue measure, thus indicating that the equation (36) is indeed nonsmooth at the solution.

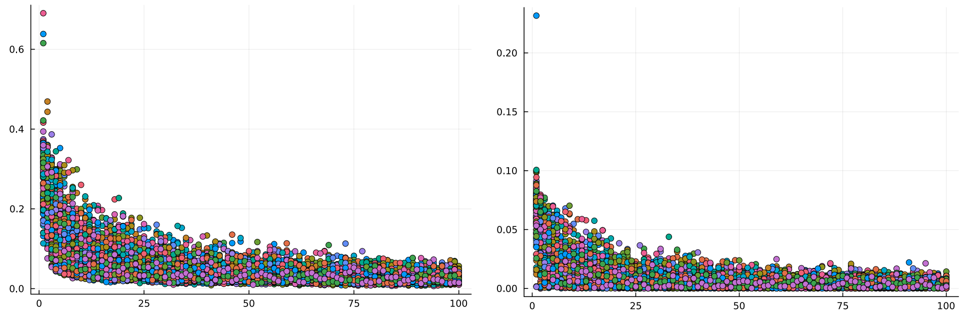
5.1.3 Details of the Implementation

The algorithm was implemented in the programming language Julia, [3], using the finite element package Gridap, [2]. Julia's multithreading capabilities allow us to easily parallelize what are usually the most expensive parts of a PDE-constrained optimization code, e.g., state, adjoint, and sensitivity equations, as the equations for individual samples ξ_i do not need to communicate. The n-instances of the random stiffness matrices for any given run of the algorithm are assembled and pre-factorized using a numeric factorization. The corresponding objects, which are repeatedly used in the optimization algorithm, are then cached. The generalized second-order derivatives needed to solve for the Newton step are only implicitly available. We made use of the efficient computation of Hessian-vector products in our setting as detailed in [28, Chap. 1.6.5] and applied a standard CG algorithm, [27], to solve the linear system for the step calculation δz_k. We (heuristically) chose $c = 10^{-3}$ as the scaling factor for determining the active and inactive sets and the damping factor $t_k = 1.01$ or 1.00. For the stopping criterion, we used an absolute and relative tolerance of 10^{-8} for the discrete L^2-norm of the residual. With these parameters, the semismooth Newton method was very robust and converged in no more than 10 iterations on average, often much less.

5.1.4 Results

We use the following strategy to compute experimental convergence rates for the optimal solutions and optimal values. Fixing a mesh, we choose a maximum sample size, in our case $n = 500$, and solve the optimization problem using the solution method detailed above. This yields a "true" solution $z(P_n)$ and optimal value $v(P_n)$. We select smaller sample sizes $m = 1, \ldots, M$, with $M = 100$ and resolve the optimization problems to obtain $z(P_m)$ and $v(P_m)$. For each m, we repeat the experiment 100 times and generate a sample of solutions and optimal values $\left\{(z(P_{m,j}), v(P_{m,j})\right\}_{j=1}^{100}$. The values of ξ used in these experiments are "out of sample" in that we do not use subsamples of the same data that was used to compute $z(P_n)$ and $v(P_n)$. We then compute and save the values $\|z(P_{m,j}) - z(P_n)\|_{L^2}$ and $|v(P_{m,j}) - v(P_n)|$. We plot these results in Figure 3, where the horizontal axis corresponds to $m = 1, \ldots, 100$ and each dot represents $\|z(P_{m,j}) - z(P_n)\|_{L^2}$ (left subfigure) and $|v(P_{m,j}) - v(P_n)|$ (right subfigure), respectively. Finally, we use a simple application of logistic regression on these values to compute experimental rates of convergence, this yielded:

$$\|z(P_m) - z(P_n)\|_{L^2(\Omega)} \in O(m^{-0.53656}) \text{ and } |v(P_m) - v(P_n)| \in O(m^{-0.66035}),$$



(a) Statistics of $\|z(P_m) - z(P_n)\|_{L^2(\Omega)}$. (b) Statistics of $|v(P_m) - v(P_n)|$

Fig. 3: Stability statistics: Experimental convergence rates of the optimal solutions and optimal values. A coarse uniform mesh was chosen that corresponded to 900 degrees of freedom was used. Figure 3a exhibits an experimental rate of $\|z(P_m) - z(P_n)\|_{L^2(\Omega)} \in O(m^{-0.53656})$. Figure 3b exhibits an experimental rate of $|v(P_m) - v(P_n)| \in O(m^{-0.66035})$.

which is within the theoretical bounds. It is important to mention that there are many factors that would cause a deviation from the theoretical value of $-1/2$, among them we note: the discretization error due to the finite element method, the use of an iterative solver to compute the steps δz_k, the size of n and m used to approximate the convergence rate, and that each subproblem for both $z(P_n)$ and $z(P_{m,j})$ is a random realization of the true problem.

5.2 Subsampling Bootstrapped Confidence Intervals

5.2.1 Background and Computational Method

As discussed, it is possible to use subsampling techniques to provide experimental confidence intervals for the optimal values. The theory of subsampling makes very few assumptions and can be widely applied. We briefly review the numerical procedure and its justification. Afterwards, we provide several numerical experiments to give the reader a better idea about the behavior of the sample size n, subsample size b, and number of subsamples m in practice.

According to Theorem 5 and Corollary 2, we can argue that

$$\sqrt{n}(v(P_n) - v(P)) \xrightarrow{d} \mathcal{N}(0, P(f(z(P)))^2).$$

Since $v(P)$ and $P(f(z(P)))^2$ are not available to us, it is essentially impossible to make use of this in practice. The subsampling-based bootstrapping estimates $v(P_n^*(N_j^{n,b}))$ can be used instead based on several direct applications of the calculus of stochastic convergence. We refer the reader to [4] or [63, Chap. 1 and 2] for details.

We sketch the arguments in [45, Chap. 2], especially [45, Thm 2.2.1], which rigorously justifies the procedure. To this aim, assume that n, b are chosen such that $\sqrt{b}/\sqrt{n} \to 0$. From a sampling perspective, $\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P))$ also converges in distribution to $\mathcal{N}(0, P(f(z(P)))^2)$ as $b \to +\infty$. Moreover, as observed in [4, pp. 26-27 and pg. 28 Problem 1.], we can deduce

$$-\sqrt{b}(v(P_n) - v(P)) = -\frac{\sqrt{b}}{\sqrt{n}}(\sqrt{n}(v(P_n) - v(P))) \xrightarrow{d} 0,$$

where $-\sqrt{b}/\sqrt{n}$ can be thought of as a (degenerate) random variable that converges in distribution to zero. In fact, the previous statement holds for the stronger notion of convergence in probability. Consequently, we have

$$\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P_n)) = \sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P)) - \sqrt{b}(v(P_n) - v(P)) \xrightarrow{d} \mathcal{N}(0, P(f(z(P)))^2).$$

This follows from the fact that $\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P)) \to 0$ in distribution, $-\sqrt{b}(v(P_n) - v(P)) \to 0$ in probability, and [4, pg. 28 Problem 1.] or alternatively, using $-\sqrt{b}(v(P_n) - v(P)) \to 0$ in distribution and the continuous mapping theorem.

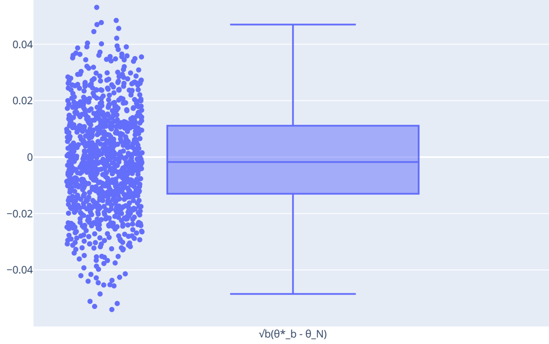


Fig. 4: Box plot for the subsampling bootstrapped statistic $\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P_n))$ for a mesh of size 32 × 32 using $n = 2000$, $b = 1000$ $j = 1, \dots, m$ with $m = 1000$, $\alpha = 0.05$.

Next, let F denote the cumulative distribution associated with some random variable $U \sim \mathcal{N}(0, P(f(z(P)))^2)$ and fix $\alpha \in (0, 1)$, a confidence level. In addition, let $L_{n,b}$ be the empirical cumulative distribution associated with $\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P_n))$ for some sample $j = 1, \dots, m$. For large n,

$$F^{-1}(\alpha/2) < \sqrt{n}(v(P_n) - v(P)) < F^{-1}(1 - \alpha/2)$$

holds with probability $1 - \alpha$. Similarly,

$$L_{n,b}^{-1}(\alpha/2) < \sqrt{n}(v(P_n) - v(P)) < L_{n,b}^{-1}(1 - \alpha/2),$$

holds, as well (with probability close to $1 - \alpha$), see [45, Thm 2.2.1 (iii)]. Combining these observations, we can now compute $1 - \alpha$ confidence intervals $[\underline{c}_{n,b,\alpha}, \overline{c}_{n,b,\alpha}]$ on $v(P)$ by setting

$$\underline{c}_{n,b,\alpha} = v(P_n) - n^{-1/2} L_{n,b}^{-1}(\alpha/2)$$
$$\overline{c}_{n,b,\alpha} = v(P_n) - n^{-1/2} L_{n,b}^{-1}(1 - \alpha/2).$$

5.2.2 Results

We illustrate the computational method presented above here. The theoretical results are asymptotic and we require a large amount of computational power to generate the subsampling bootstrapped statistics. For example, a single run of the semismooth Newton algorithm, which only requires eight to ten Newton steps, may need several million PDE solves. To understand this fact, we note that each CG iteration requires, in a general setting, four PDE solves per sample plus two initial PDE solves (per sample) for the computation of the nonlinear residual. Our CG implementation typically reached a tolerance of 10^{-8} after 25-30 iterations. Consequently, we only show results for a rather coarse mesh of size 32 × 32. For the sampling-based quantities, we set $n = 2000$, $b = 1000$, $m = 1000$ and we consider a confident level of 95 ($\alpha = 0.05$ in the preceding computation). This provides the following lower and upper 95% confidence bounds

$$\underline{c}_{n,b,\alpha} = 0.089148 \text{ and } \overline{c}_{n,b,\alpha} = 0.090720.$$

The individual statistics are plotted in Figure 4. In addition, we resolve the original problem 100 times for samples of size $n = 2000$ to compute“hit-or-miss” statistics based on how often the new optimal values fall between the confidence bounds. In this setting, the out-of-sample optimal values were within the bounds 84 out of 100 times. We would also like to note that our experiments indicated a deeper connection between n, b, m and the mesh size. For the values of n, b, m used here, we observed more “hits” for coarser meshes. The quality quickly degraded as the mesh became finer. A deeper study of this relationship goes beyond the scope of the current text, but will be the source of future work.

6 Discussion and conclusions

In this paper we studied Monte Carlo methods for solving a stochastic optimization problem with linear quadratic risk-neutral objective function, a linear elliptic PDE with random coefficients and convex control constraints. Based on empirical process theory we were able to show that both optimal values and solutions converge in mean with the best possible convergence rate $O(n^{-\frac{1}{2}})$ if the coefficients of the PDE are sufficiently smooth. The required degree of smoothness is related to the finite dimension of the random parameter. In addition, the optimal values satisfy a central limit result which enables the derivation of confidence intervals by resampling.

Our methodology is no longer successful if the optimization model (5) contains random convex control constraints that correspond to state constraints in the original stochastic optimization problem (35), (4). It also fails if the risk-neutral expectation in the objective is replaced by some convex risk measure. Although such risk measures preserve convexity, they typically introduce nonsmoothness as, for example, in the case of so-called Conditional or Average Value-at-risk CVaR. In this case, problem (5) would be of the form

$$\min\left\{\mathrm{CVaR}_\kappa(f(z,\cdot))=\inf_{t\in\mathbb{R}}\left\{t+\frac{1}{1-\kappa}\int_\Xi \max\{0,f(z,\xi)-t\}\,dP(\xi)\right\}:z\in Z_{\mathrm{ad}}\right\} \tag{37}$$

for some $\kappa\in(0,1)$ and f defined in (6). Hence, the corresponding minimal information distance is based on a class $\mathfrak{F}$ of functions that is no longer smooth as needed for our main result (Theorem 5). The classical way for reformulating (37) into a smooth optimization problem was suggested in [48] and leads to

$$\min\left\{t+\frac{1}{1-\kappa}\int_\Xi y(\xi)\,dP(\xi):y(\xi)\geq f(z,\xi)-t,\;y(\xi)\geq 0,\;t\in\mathbb{R}_+,\;z\in Z_{\mathrm{ad}}\right\}$$

and, thus, to an optimization model with random convex constraints. A possible way out consists in the approach of smoothing CVaR as suggested in [33,35].

Finally, we mention two possible extensions of the results in this paper. The first extension consists in introducing a random mapping $B(\xi):H\to V^\star$ and by replacing z in (4) by $B(\xi)z$. If one requires that the function $\langle B(\cdot)z,v\rangle$ is sufficiently smooth on Ξ for all $z\in Z_{\mathrm{ad}}$, $v\in V$, the function classes $\mathfrak{F}_{mi}$ and $\mathfrak{F}_{di}$ have to be modified, but the main results carry over.
A second extension concerns the finite dimensionality of Ξ in Theorems 4 and 5. In our earlier paper [30] and in Section 2 the set Ξ represents a metric space. Hence, the general stability results (Theorems 1 and 3) enable the use of probability measures on infinite dimensional spaces. For example, this allows to consider the Karhunen-Loève expansion of a centered stochastic process $\{\xi_x\}_{x\in D}$ with probability distribution P on $\Xi=L_2(D)$, finite second moments and continuous covariance function $K(x,y)=\mathbb{E}[\xi_x\,\xi_y]$, $x,y\in D$, which is of the form

$$\xi_x=\sum_{j=1}^{\infty}Z_je_j(x)\qquad(x\in D). \tag{38}$$

Here, $(e_j)_{j\in\mathbb{N}}$ is an orthogonal system in $H=L_2(D)$ and $(Z_j)_{j\in\mathbb{N}}$ is a sequence of centered, uncorrelated real random variables (see [56] and references therein). A truncated version of (38) with d summands can then be used for Monte Carlo sampling and the truncation error be estimated by the distance $d_{\mathfrak{F}}$ or possible upper bounds.

The literature on uncertainty quantification is abound with many techniques whose origins can be traced to the original Monte Carlo approach. For example, there are multi-fidelity methods, see e.g. [41,42] and references therein, and multi-level Monte Carlo, see e.g., [21,22] and the references therein, as well as the pioneering work by S. Heinrich et al., e.g. [25,26]. These methods are all useful for increasing the scale, efficiency, and application to more complex examples of the numerical experiments in Section 5. The techniques themselves are often used for the approximation of random and parametric partial differential equations or functionals of the solutions, not typically for optimization problems or their solutions/optimal values. In some sense, the estimator $\sqrt{b}(v(P_n^*(N_j^{n,b}))-v(P_n))$ can be thought of as a multifidelity estimator for the true optimal value as it uses a) finite-dimensional approximations of the underlying partial differential equation and b) two samples to estimate the underlying probability measures ($P_n^*(N_j^{n,b})$ and P_n). The fine mesh, large sample estimate $v(P_n)$ is a “high fidelity” approximation and the many subsampled estimates $v(P_n^*(N_j^{n,b}))$ are “low fidelity” approximations. Using the subsampling bootstrapping analysis above, we are able to derive a central limit theorem for the centered approximation of $v(P)$. An interesting future research direction would be to investigate to what extent one can use subsampling bootstrapping techniques in the context of other multifidelity approaches, e.g. for the problems considered in [40].

Acknowledgement

The project benefited from the Experimental Infrastructure for the Exploration of Exascale Computing (eX3), supported by the Research Council of Norway under contract No. 270053. The authors would especially like to thank Luk Burchard for his help, discussions, and suggestions for using eX3 for the statistical experiments.

References

1. W. Alt: Local stability of solutions to differentiable optimization problems in Banach spaces, *Journal Optimization Theory Applications* 70 (1991), 443–466.
2. S. Badia and F. Verdugo, Gridap: An extensible Finite Element toolbox in Julia, *Journal Of Open Source Software*. **5**, 2520 (2020).
3. J. Bezanson, A. Edelman, S. Karpinski, & V. Shah, Julia: A fresh approach to numerical computing. *SIAM Review*. **59**, 65-98 (2017).
4. P. Billingsley: *Convergence of Probability Measures*, Wiley Series in Probability and Mathematical Statistics, Vol. 9, Wiley, New York, 1968.
5. S. Boucheron, G. Lugosi, and P. Massart: *Concentration Inequalities - a Nonasymptotic View of Independence*, Oxford University Press, Oxford, 2013.
6. C. Castaing and M. Valadier: *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics Vol. 580, Springer, Berlin, 1977.
7. A. Chaudhuri, M. Norton, and B. Kramer: Risk-based design optimization via probability of failure, conditional value-at-risk and buffered probability of failure, in *AIAA Scitech 2020 Forum*.
8. P. Chen, M. R. Haberman, and O. Ghattas: Optimal design of acoustic metamaterial cloaks under uncertainty, *Journal of Computational Physics* 431 (2021), 110114.
9. P. Chen, A. Quarteroni, and G. Rozza: Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations, *SIAM Journal on Numerical Analysis* 51 (2013), 2700–2722.
10. A. Cohen, R. DeVore and C. Schwab: Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs, *Foundations of Computational Mathematics* 10 (2010), 615–646.
11. S. Conti, H. Held, M. Pach, M. Rumpf and R. Schultz: Risk-averse shape optimization, *SIAM Journal on Control and Optimization* 49 (2011), 927–947.
12. S. Dereich, M. Scheutzow and R. Schottstedt: Constructive quantization: Approximation by empirical measures, *Annales de l'Institut Henri Poincaré – Probabilités et Statistiques* 49 (2013), 1183–1203.
13. A. L. Dontchev and W. W. Hager: Lipschitzian stability in nonlinear control and optimization, *SIAM Journal Control and Optimization* 31 (1993), 569–603.
14. R. M. Dudley: The speed of mean Glivenko-Cantelli convergence, *The Annals of Mathematical Statistics* 40 (1969), 40–50.
15. R. M. Dudley: *Real Analysis and Probability*, Chapman & Hall, New York, 1989.
16. J. Dupačová and R. J-B Wets: Asymptotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems, *The Annals of Statistics* 16 (1988), 1517–1549.
17. A. Eichhorn and W. Römisch: Stochastic integer programming: Limit theorems and confidence intervals, *Mathematics of Operations Research* 32 (2007), 118–135.
18. N. Fournier and A. Guillin: On the rate of convergence in Wasserstein distance of the empirical measure, *Probability Theory and Related Fields* 162 (2015), 707–738.
19. S. Garreis, T. M. Surowiec, and M. Ulbrich: An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures, *SIAM Journal on Optimization* 31 (2021), 1–29.
20. C. Geiersbach and W. Wollner: A stochastic gradient method with mesh refinement for PDE-constrained optimization under uncertainty, *SIAM Journal on Scientific Computing* 42 (2020), A2750–A2772.
21. Giles, M. Multilevel Monte Carlo path simulation. *Oper. Res.*. **56**, 607-617 (2008)
22. Giles, M. Multilevel Monte Carlo methods. *Acta Numer.*. **24** 259-328 (2015)
23. E. Giné: Empirical processes and applications: An overview, *Bernoulli* 2 (1996), 1–28.
24. E. Giné and J. Zinn: Bootstrapping general empirical measures, *The Annals of Probability* 18 (1990), 851–869.
25. Heinrich, S. Monte Carlo complexity of global solution of integral equations. *J. Complexity*. **14**, 151-175 (1998),
26. Heinrich, S. & Sindambiwe, E. Monte Carlo complexity of parametric integration. *J. Complexity*. **15** 317-341 (1999),
27. M.R. Hestenes and E. Stiefel, Methods of Conjugate Gradients for Solving Linear Systems, *Journal of Research of the National Bureau of Standards.* 49 (6) (1952) 409-436.
28. M. Hinze, R. Pinnau, S. Ulbrich and M. Ulbrich: *Optimization with PDE constraints*, Springer, New York, 2009.
29. M. Hintermüller, K. Ito, and K. Kunisch, The primal-dual active set strategy as a semismooth Newton method, *SIAM J. Optim.* 13(3), 865–888 (2003)
30. M. Hoffhues, W. Römisch and T. M. Surowiec: On quantitative stability of infinite-dimensional optimization under uncertainty, *Optimization Letters* (to appear).
31. A. N. Kolmogorov and V. M. Tikhomirov: Epsilon-entropy and epsilon-capacity of sets in function spaces, *American Mathematical Society Translation Series* 2 17 (1961), 277–364.
32. D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders: A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty, *SIAM Journal on Scientific Computing* 35 (2013), A1847–A1870.
33. D. P. Kouri and T. M. Surowiec: Risk-averse PDE-constrained optimization using the conditional value-at-risk, *SIAM Journal on Optimization* 26 (2016), 365–396.
34. D. P. Kouri and T. M. Surowiec: Existence and optimality conditions for risk-averse PDE-constrained optimization, *SIAM/ASA Journal on Uncertainty Quantification* 6 (2018), 787–815.
35. D. P. Kouri and T. M. Surowiec: Epi-regularization of risk measures, *Mathematics of Operations Research* 45 (2019), 774–795.
36. D. P. Kouri and T. M. Surowiec: Risk-averse optimal control of semilinear elliptic PDEs, *ESAIM: Control, Optimization and Calculus of Variations* 26 (2020) 53.
37. A. Kunoth and C. Schwab: Analytic regularity and gpc approximation for control problems constrained by linear parametric elliptic and parabolic PDEs, *SIAM Journal Control and Optimization* 51 (2013), 2442–2471.

38. P. Lachout: Sensitivity of stochastic programs by means of the infimum functional, Proceedings of Prague Stochastics 2006, Prague, 495–504.
39. J. Milz: Sample average approximations of strongly convex stochastic programs in Hilbert spaces, Preprint arXiv:2104.05114v1.
40. Ng, L. & Willcox, K. Multifidelity approaches for optimization under uncertainty. *Internat. J. Numer. Methods Engrg.*. **100**, 746-772 (2014).
41. Peherstorfer, B., Willcox, K. & Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.*. **60**, 550-591 (2018),
42. Peherstorfer, B., Gunzburger, M. & Willcox, K. Convergence analysis of multifidelity Monte Carlo estimation. *Numer. Math.*. **139**, 683-707 (2018),
43. G. Ch. Pflug: Stochastic optimization and statistical inference, in *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks of Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003, 426–482.
44. D. N. Politis and J. P. Romano: Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics* 22 (1994), 2031–2050.
45. D. N. Politis, J. P. Romano, and M. Wolf: *Subsampling*, Springer Series in Statistics, Springer, New York, 1999.
46. S. T. Rachev: *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
47. S. T. Rachev and W. Römisch, *Quantitative stability in stochastic programming: The method of probability metrics*, *Mathematics of Operations Research* **27** (2002), 792–818.
48. R. T. Rockafellar and S.Uryasev: Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* 26 (2002), 1443–1471.
49. W. Römisch: Delta method, infinite dimensional, in: *Encyclopedia of Statistical Sciences* (S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic eds.), Wiley, 2006.
50. A. Ruszczyński and A. Shapiro (eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.
51. C. Schillings, S. Schmidt, and V. Schulz: Efficient shape optimization of certain and uncertain aerodynamic design, *Computers & Fluids* 46 (2011), 78–87.
52. A. Shapiro: On concepts of directional differentiability, *Journal Optimization Theory and Applications* 66 (1990), 477–487.
53. A. Shapiro: Asymptotic analysis of stochastic programs, *Annals of Operations Research* 30 (1991), 169–186.
54. A. Shapiro: Perturbation analysis of optimization problems in Banach spaces, *Numerical Functional Analysis and Optimization* 13 (1992), 97–116.
55. A. Shapiro: Monte Carlo sampling methods, in *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks of Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003, 363–425.
56. I. Steinwart: Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties, *Potential Analysis* 51 (2019), 361–395.
57. M. Talagrand: Measurability problems fot empirical processes, *The Annals of Probability* 15 (1987), 204–212.
58. M. Talagrand: Concentration of measure and isoperimetric inequalities in product spaces, *Publications Mathématiques de l'I.H.E.S.* 81 (1995), 73–205.
59. F. Topsøe: On the connection between P-continuity and P-uniformity in weak convergence, *Probability Theory and its Applications* 12 (1967), 281–290.
60. M. Ulbrich *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, SIAM, Philadelphia, 2011
61. A. van Barel and S. Vandewalle: Robust optimization of PDEs with random coefficients using multilevel Monte Carlo method, *SIAM/ASA Journal on Uncertainty Quantification* 7 (2019), 174–202.
62. A. W. van der Vaart: New Donsker classes, *The Annals of Probability* 24 (1996), 2128–2140.
63. A. W. van der Vaart: *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
64. A. W. van der Vaart and J. A. Wellner: *Weak Convergence and Empirical Processes*, Springer, New York, 1996.
65. F. Verdugo and S, Badia, The software design of Gridap: A Finite Element package based on the Julia JIT compiler, *Computer Physics Communications*. **276** pp. 108341 (2022).
66. V. M. Zolotarev: Probability metrics, *Theory of Probability and its Applications* 28, 278–302.