

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

W. Römisch^{c,*}, T. M. Surowiec^d

^a*Institute of Mathematics, Humboldt University of Berlin, Unter den Linden 6
10099, Berlin, Germany*

^b*FB12 Mathematik und Informatik, Philipps-Universität
Marburg, Hans-Meerwein-Straße 6, 35032, Marburg, Germany*

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.

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

2000 MSC: 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 [6, 24, 25, 27], numerical and computational methods [16, 17, 23, 46], and applications [4, 5, 8, 39]. 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.

*Corresponding author

Email addresses: romisch@math.hu-berlin.de (W. Römisch),
surowiec@mathematik.uni-marburg.de (T. M. Surowiec)

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, [17], versus methods that rely on sampling from the underlying probability measure [46]. 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., [20].

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 $n \rightarrow \infty$ 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 [21]. However, for the PDE-constrained optimization problem provided in [21, Sec. 7], a major open question remained: Can we derive a reasonable rate of convergence for the minimal information metric that supports our numerical results? Taking this question as a starting point, we seek to answer this by using deep results from empirical process theory as detailed in [18, 49]. The idea to use empirical process theory has been employed in the stochastic programming literature before, cf. some results in [38, 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 [30] in which the author obtained results that are in parts similar those in the present paper. However, the results in [30] 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 [28]. 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 [28] 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 [21] 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. 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 $\|\cdot\|$. By $L_p(\Omega, \mathcal{A}, \mu; Y)$, $1 \leq p \leq \infty$, we denote the linear space of strongly measurable functions $y : \Omega \rightarrow Y$ such

that the integral $\int_{\Omega} \|y(\omega)\|^p d\mu$ is finite. As usual we equip this space with the norm

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

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 -field, 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 $\|g\| = \sup_{y \in Y} |g(y)|$.

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 $\|u\| = \sqrt{(u, u)}$. The topological dual is denoted by $V^* = H^{-1}(D)$ with the usual operator norm $\|\cdot\|_*$. 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^* 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 $a(\cdot, \cdot; \xi) : V \times V \rightarrow \mathbb{R}$

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

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

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

for a.e. $x \in D$ and any $\xi \in \Xi$. This implies that each b_{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

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

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

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

for P -a.e. $\xi \in \Xi$ and all test functions $v \in V$, where $g : D \times \Xi \rightarrow \mathbb{R}$ 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 $A(\xi) : V \rightarrow V^*$ 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 $A(\xi)^{-1} : V^* \rightarrow V$ 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 $z \in Z_{\text{ad}}$ as:

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

with the integrand

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

for any $z \in H$ and $\xi \in \Xi$, where $g \in L_2(\Xi, P; H)$ and $A(\xi)^{-1}$ as defined earlier. For each $\xi \in \Xi$ the function $f(\cdot, \xi) : H \rightarrow \mathbb{R}$ 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 $F_P : H \rightarrow \mathbb{R}$. They are collected in the following result which is partly proved in [21].

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 a unique minimizer $z(P) \in Z_{\text{ad}}$ of (5) and the objective function F_P has quadratic growth around $z(P)$, i.e. we have*

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

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

$$|F_P(z) - F_P(\tilde{z})| \leq \sup_{t \in [0,1]} \|F'_P(z + t(\tilde{z} - z))\| \|z - \tilde{z}\| \quad (8)$$

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

PROOF. While the first part is proved in [21], 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 at $t = 0$. 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 $h : [0, 1] \rightarrow \mathbb{R}$ satisfies the estimate

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

we obtain for $h(t) = F_P(z + t(\tilde{z} - z))$ the desired estimate (8). \square

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

$$d_{\mathfrak{F}}(P, Q) = \sup_{f \in \mathfrak{F}} \left| \int_{\Xi} f(\xi) dP(\xi) - \int_{\Xi} f(\xi) dQ(\xi) \right| \quad (9)$$

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 [34]). Such distances of probability measures were first introduced and studied in [50]. 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

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

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

$$\mathfrak{F}_{mi} = \{f(z, \cdot) : z \in Z_{\text{ad}}\}. \quad (10)$$

Following [35] 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 [45].

Lemma 2. *If \mathfrak{F} is uniformly bounded and it holds that*

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

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

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

The choice (10) of \mathfrak{F} leads to the following result proved in [21].

Theorem 3. *Under the standing assumptions and with the class \mathfrak{F}_{mi} in (10) we obtain the estimates*

$$|v(Q) - v(P)| \leq d_{\mathfrak{F}_{mi}}(P, Q) \quad (11)$$

$$\|z(Q) - z(P)\|_H \leq 2\sqrt{\frac{2}{\alpha}} d_{\mathfrak{F}_{mi}}(P, Q)^{\frac{1}{2}} \quad (12)$$

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 [21]).

Theorem 4. *Assume that all functions $b_{ij}(x, \cdot)$, $i, j = 1, \dots, m$, 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 $\mathfrak{F}_{mi} = \{f(z, \cdot) : z \in Z_{\text{ad}}\}$ 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 (12) in Theorem 3 to Lipschitz stability with respect to a pseudo-metric of the type (9), 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, 10], 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 [41, Lemma 2.1].

Theorem 5. *Under the standing assumptions the Lipschitz-type estimate*

$$\|z(Q) - z(P)\|_H \leq \frac{8}{\alpha} d_{\mathfrak{F}_{di}}(P, Q) \quad (13)$$

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

$$\mathfrak{F}_{di} = \left\{ \langle A(\cdot)^{-1}(z + g(\cdot)) - \tilde{u}, A(\cdot)^{-1}h \rangle_H + \alpha \langle z, h \rangle_H : z \in Z_{\text{ad}}, \|h\|_H \leq 1 \right\}. \quad (14)$$

PROOF. Let $P, Q \in \mathcal{P}(\Xi)$ and $z(P), z(Q) \in Z_{\text{ad}}$ 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 $h : [0, 1] \rightarrow \mathbb{R}$ given by $h(t) = (F_P - F_Q)(z(P) + t(z(Q) - z(P)))$.

$z(P))$), $t \in [0, 1]$. 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

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

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

$$\begin{aligned} \frac{\alpha}{8} \|z(Q) - z(P)\|_H &\leq \sup_{z \in Z_{\text{ad}}} \|(F'_P - F'_Q)(z)\| \\ &\leq \sup_{z \in Z_{\text{ad}}} \sup_{\|h\|_H \leq 1} \left| \int_{\Xi} f'_z(z, \xi)(h) d(P - Q)(\xi) \right|, \end{aligned}$$

where $f'_z(z, \xi)$ denotes the partial Gâteaux derivative of f with respect to the first variable. A straightforward evaluation shows that

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

holds for all $z \in Z_{\text{ad}}$, $\|h\|_H \leq 1$, $\xi \in \Xi$. This completes the proof. \square

Remark 1. An inspection of the proof of Theorem 4 (see [21, 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 4.

3. Monte Carlo approximations

Let $\xi_1, \xi_2, \dots, \xi_n, \dots$ 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

$$P_n(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{\xi_i(\cdot)} \quad (n \in \mathbb{N}), \quad (15)$$

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.,

$$\min \left\{ \int_{\Xi} f(z, \xi) dP_n(\cdot)(\xi) = \frac{1}{n} \sum_{i=1}^n f(z, \xi_i(\cdot)) : z \in Z_{\text{ad}} \right\}. \quad (16)$$

The optimal value $v(P_n(\cdot))$ of (16) is a real random variable and the solution $z(P_n(\cdot))$ an H -valued random element (see [3, Lemma III.39]).

Qualitative and quantitative results on the asymptotic behavior of optimal values and solutions to (16) are known in finite-dimensional settings (see [13], and the surveys [42] and [31]). 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 2 and Theorems 3 and 4.

Corollary 6. *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

$$\left\{ \mathbb{G}_n(\cdot)f := \sqrt{n}(P_n(\cdot) - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(\xi_i(\cdot)) - Pf) \right\}_{f \in \mathfrak{F}} \quad (17)$$

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

$$\|\mathbb{G}_n(\cdot)\|_{\mathfrak{F}} = \sup_{f \in \mathfrak{F}} |\mathbb{G}_n(\cdot)f| = \sqrt{n} d_{\mathfrak{F}}(P_n(\cdot), P) = \sqrt{n} \sup_{f \in \mathfrak{F}} |P_n(\cdot)f - Pf|. \quad (18)$$

Here, we set $Pf = \int_{\Xi} f(\xi) dP(\xi)$ for any probability distribution P and any $f \in \mathfrak{F}$. Since the supremum in (18) is taken with respect to an uncountable set \mathfrak{F} , it is not necessarily measurable with respect to \mathcal{F} . In our applications to the classes \mathfrak{F}_{mi} and \mathfrak{F}_{di} , however, the supremum (18) can be taken with respect to some countable subset of both classes since the Hilbert space H is separable and all functions f are continuous. Hence, the supremum is measurable.

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 [44] and presented in some detail in [2]) with applications to bounding (18) 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 (17) 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

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

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 7). From this we deduce the following mean convergence rates

$$\mathbb{E}[d_{\mathfrak{F}}(P_n(\cdot), P)] = O(n^{-\frac{1}{2}}). \quad (19)$$

Together with Theorems 3 and 5 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 [47, Thm. A.2]).

Proposition 7. *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*

$$\mathbb{E}[\|\mathbb{G}_n\|_{\mathfrak{F}}] \leq C \int_0^1 \sqrt{1 + H_{[]}(\varepsilon \|\hat{F}\|_{P,2}, \mathfrak{F}, \|\cdot\|_{P,2})} d\varepsilon \|\hat{F}\|_{P,2} \quad (20)$$

holds. If the integral in (20) is finite, then the class \mathfrak{F} is P -Donsker.

Note that the integral in (20) can only be finite if $H_{[]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{P,2})$ grows at most like $\varepsilon^{-\beta}$ with $0 < \beta < 2$ for $\varepsilon \rightarrow +0$.

Next we discuss the assumption of finiteness of the integral in (20) 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 $\Xi \subseteq \text{cl int } \Xi$, $k \in \mathbb{N}_0$ 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 k th order derivatives are Hölder continuous with exponent r . Next we use the notation $\mathbf{i} = (i_1, \dots, i_d)$ with $i_j \in \mathbb{N}_0$, $j = 1, \dots, d$, and $|\mathbf{i}| = \sum_{j=1}^d i_j$. Further, $D^{\mathbf{i}}f$ denotes

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

If the spaces are endowed with the norms

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

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 [22] 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 $\varepsilon > 0$

$$H_{[]}(\varepsilon \rho, \mathbb{B}_{k,r}(\rho), \|\cdot\|_{P,2}) \leq K \varepsilon^{-\frac{d}{k+r}}, \quad (21)$$

where the result from [22] was adapted to the norm in $L_2(\Xi, P)$ (see also [49, Section 2.7.1]). Hence, Proposition 7 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 4 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 4, the integral in (20) will not be finite. Hence, one cannot use Proposition 7 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 (19) 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 [21, Section 4]). Namely, it is shown in [9, 15] that the Wasserstein distance W_p of P and $P_n(\cdot)$ has the mean convergence rate

$$\mathbb{E}[W_p(P_n(\cdot), P)] = O(n^{-\frac{1}{d}}) \quad (22)$$

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 [11] 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 [7, Section 4] in a different way.

Theorem 8. *Let $\Xi \subset \mathbb{R}^d$ be a bounded, convex set having the property $\Xi \subseteq \text{cl int } \Xi$ and let $k \in \mathbb{N}$. Let the assumptions of Theorem 4 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))

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

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

obtain for $w, y \in V^*$

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

where $\Delta_A^j(\xi; h) = A(\xi + he_j) - A(\xi)$, $u(h) = A(\xi + he_j)^{-1}w$, $u = A(\xi)^{-1}w$, $v = (A(\xi)^{-1})^*y \in V$ and $(A(\xi)^{-1})^*$ denotes the adjoint mapping to $A(\xi)^{-1}$. While the first summand on the right-hand side converges for $h \rightarrow 0$ to the partial derivative $\frac{\partial}{\partial \xi_j}$ of $\langle A(\cdot)u, v \rangle$ at ξ , the second converges to zero as $(u - u(h))$ 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 $\langle y, A(\xi)^{-1}w \rangle$ 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 $z \in Z_{\text{ad}}$. 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 $z \in Z_{\text{ad}}$. We conclude that both classes \mathfrak{F}_{mi} and \mathfrak{F}_{di} are subsets of $C^{k,0}(\Xi)$. \square

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

$$\langle A(\xi)u, v \rangle = \sum_{i,j=1}^m \int_D b_{ij}(x, \xi) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} dx \quad (23)$$

for all pairs $(u, v) \in V$. Due to the uniform ellipticity condition (2) we know that all functions b_{ij} are essentially bounded. If we assume that all functions $b_{ij}(x, \cdot) : \Xi \rightarrow \mathbb{R}$, $x \in D$, 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 (23).

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 9. *Let $\Xi \subset \mathbb{R}^d$ be a bounded, convex set having the property $\Xi \subseteq \text{cl int } \Xi$ and let $k \in \mathbb{N}$ be such that $d < 2k$. Assume that all functions $b_{ij}(x, \cdot) : \Xi \rightarrow \mathbb{R}$, $i, j = 1, \dots, m$, and $g(x, \cdot) : \Xi \rightarrow \mathbb{R}$, $x \in D$, 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*

$$\mathbb{E}[|v(P_n(\cdot)) - v(P)|] = O(n^{-\frac{1}{2}}) \quad (24)$$

$$\mathbb{E}[\|z(P_n(\cdot)) - z(P)\|_H] = O(n^{-\frac{1}{2}}) \quad (25)$$

and the sequence $(\sqrt{n}(v(P_n(\cdot)) - v(P)))$ converges in distribution to some real random variable ζ , where $v(P)$ 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 (16), respectively.

PROOF. Our assumptions together with Theorem 8 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 (20) in Proposition 7 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 (24) from Theorem 3 and (25) from Theorem 5. Furthermore, we conclude for $\mathfrak{F} = \mathfrak{F}_{mi}$ from Proposition 7 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 (10) of \mathfrak{F} , we may also write $\{Pf\}_{f \in \mathfrak{F}} = \{Pf(z, \cdot) : z \in Z_{\text{ad}}\}$ and $\{\mathbb{G}f\}_{f \in \mathfrak{F}} = \{\mathbb{G}f(z, \cdot)\}_{z \in Z_{\text{ad}}}$. This means that Pf may be considered as element of the space $\ell^\infty(Z_{\text{ad}})$ of bounded real-valued functions

on Z_{ad} . Correspondingly, $\mathbb{G}f$ may be viewed as random variable in $\ell^\infty(Z_{\text{ad}})$. It remains to utilize the functional delta theorem (see [37]) for the infimal mapping

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

The mapping Φ is finite, concave, hence, directionally differentiable on $\ell^\infty(Z_{\text{ad}})$. In addition, Φ is Lipschitz continuous (with modulus 1) and, hence, Hadamard directionally differentiable (see [40]). The Hadamard directional derivative at $h_0 \in \ell^\infty(Z_{\text{ad}})$ is of the form (see [29] and the survey [37])

$$\Phi'_{h_0}(h) = \liminf_{\varepsilon \downarrow 0} \{h(z) : z \in Z_{\text{ad}}, h_0(z) \leq \Phi(h_0) + \varepsilon\} \quad (h \in \ell^\infty(Z_{\text{ad}})). \quad (26)$$

Then the functional delta theorem [37, Theorem 1] implies that

$$\sqrt{n}(\Phi(P_n(\cdot)f) - \Phi(Pf)) \xrightarrow{d} \Phi'_{Pf}(\mathbb{G}f) = \zeta, \quad (27)$$

where \xrightarrow{d} denotes convergence in distribution of real random variables. Since the Hadamard directional derivative $\Phi'_{Pf}(\cdot)$ is continuous (see [40]), ζ is a real random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. This completes the proof. \square

To establish an extension of Theorem 9 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 $\Xi_j \subseteq \text{cl int } \Xi_j$, $j \in \mathbb{N}$. The idea is to apply Theorem 9 on each subset Ξ_j of \mathbb{R}^d and then to apply the argument in [47, Theorem 1.1] (see also [47, Corollary 2.1]).

Corollary 10. *Let $k \in \mathbb{N}$ be such that $d < 2k$. Assume that all functions $b_{ij}(x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$, $i, j = 1, \dots, m$, and $g(x, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$, $x \in D$, 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*

$$\sum_{j=1}^{\infty} \rho_j P(\Xi_j)^{\frac{1}{2}} < \infty. \quad (28)$$

Then both classes \mathfrak{F}_{mi} and \mathfrak{F}_{di} are P -Donsker and (24), (25) and the central limit theorem for optimal values remain true.

We note that condition (28) represents a quite implicit link between the growth of derivatives of the functions in both classes with the tail behaviour of P .

Remark 5. Note that Theorem 9 allows to derive asymptotically consistent confidence intervals for optimal values by using resampling techniques such as bootstrapping [19] or subsampling [33]. Since we not know that the Hadamard directional derivative is linear in the direction, the classical bootstrap cannot be used. However, a variant called extended bootstrap in [14] and also subsampling can be used. The subsampling method is more generally applicable than the bootstrap, because only a basic limit theorem like that in Theorem 9 is required.

4. Subsampling

The subsampling method [33] 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 ξ_1, \dots, ξ_n be an iid sample from P . Let P_n be the empirical measure and $v(P_n)$ the corresponding optimal value of (16). Based on the samples $\xi_{n_1}, \dots, \xi_{n_b}$ drawn from $\{1, \dots, n\}$ with cardinality b , we consider the corresponding empirical measure

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

and the optimal value $v(P^*(n_1, \dots, n_b))$. The subsampling method estimates the limit distribution of $\zeta = \Psi'_{Pf}(\mathbb{G}f)$ (see (27)) based on both optimal values. It is justified by the limit theorem [32, Theorem 2.1] which reads in our framework

$$\binom{n}{b}^{-1} \sum_{1 \leq n_1 < \dots < n_b \leq n} \delta_{\{\sqrt{b}(v(P^*(n_1, \dots, n_b)) - v(P_n))\}} \xrightarrow{d} \zeta \quad (29)$$

for $b, n \rightarrow \infty$ and $b/n \rightarrow 0$. The number of summands in (29) 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 $\{1, \dots, n\}$ of cardinality b : Let $N_j^{n,b} \subset \{1, \dots, n\}$ be randomly chosen with cardinality $\#N_j^{n,b} = b$ for

$j = 1, \dots, m$. Then with $P_n^*(N_j^{n,b})$ denoting the empirical measure based on $\{\xi_i : i \in N_j^{n,b}\}$ we have

$$\frac{1}{m} \sum_{j=1}^m \delta_{\{\sqrt{b}(v(P_n^*(N_j^{n,b})) - v(P_n))\}} \xrightarrow{d} \zeta \quad (30)$$

for $b, n, m \rightarrow \infty$ and $b/n \rightarrow 0$ [32, 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 $\xi_1, \xi_2, \dots, \xi_n$ 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

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

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

$$\lim_{n, m \rightarrow \infty} \mathbb{P} \left\{ \sqrt{n}(v(P_n) - v(P)) \leq \zeta_{1-\alpha, m}^* \right\} \geq 1 - \alpha.$$

5. 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 (3), (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

$$\min \left\{ \text{CVaR}_\kappa(f(z, \cdot)) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1 - \kappa} \int_{\Xi} \max\{0, f(z, \xi) - t\} dP(\xi) \right\} : z \in Z_{\text{ad}} \right\} \quad (31)$$

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 9). The classical way for reformulating (31) into a smooth optimization problem was suggested in [36] and leads to

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

and, thus, to an optimization model with random convex constraints. A possible way out consists in the approach of smoothing CVaR as suggested in [24, 26].

Finally, we mention two possible extensions of the results in this paper. The first extension consists in introducing a random mapping $B(\xi) : H \rightarrow V^*$ 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 $z \in Z_{\text{ad}}$, $v \in V$, 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 8 and 9. In our earlier paper [21] and in Section 2 the set Ξ represents a metric space. Hence, the general stability results (Theorems 3 and 5) 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

$$\xi_x = \sum_{j=1}^{\infty} Z_j e_j(x) \quad (x \in D). \quad (32)$$

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 [43] and references therein). A truncated version of (32) 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.

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. Boucheron, G. Lugosi, and P. Massart: *Concentration Inequalities - a Nonasymptotic View of Independence*, Oxford University Press, Oxford, 2013.
- [3] C. Castaing and M. Valadier: *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Mathematics Vol. 580, Springer, Berlin, 1977.
- [4] 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*.
- [5] P. Chen, M. R. Haberman, and O. Ghattas: Optimal design of acoustic metamaterial cloaks under uncertainty, *Journal of Computational Physics* 431 (2021), 110114.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] A. L. Dontchev and W. W. Hager: Lipschitzian stability in nonlinear control and optimization, *SIAM Journal Control and Optimization* 31 (1993), 569–603.

- [11] R. M. Dudley: The speed of mean Glivenko-Cantelli convergence, *The Annals of Mathematical Statistics* 40 (1969), 40–50.
- [12] R. M. Dudley: *Real Analysis and Probability*, Chapman & Hall, New York, 1989.
- [13] 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.
- [14] A. Eichhorn and W. Römisch: Stochastic integer programming: Limit theorems and confidence intervals, *Mathematics of Operations Research* 32 (2007), 118–135.
- [15] 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.
- [16] S. Garreis, T. M. Surowiec, and M. Ulbrich: An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures, *SIAM Journal on Optimization* 31 (2021), 1–29.
- [17] 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.
- [18] E. Giné: Empirical processes and applications: An overview, *Bernoulli* 2 (1996), 1–28.
- [19] E. Giné and J. Zinn: Bootstrapping general empirical measures, *The Annals of Probability* 18 (1990), 851–869.
- [20] M. Hinze, R. Pinnau, S. Ulbrich and M. Ulbrich: *Optimization with PDE constraints*, Springer, New York, 2009.
- [21] M. Hoffhues, W. Römisch and T. M. Surowiec: On quantitative stability of infinite-dimensional optimization under uncertainty, *Optimization Letters* (to appear).
- [22] 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.

- [23] 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.
- [24] 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.
- [25] 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.
- [26] D. P. Kouri and T. M. Surowiec: Epi-regularization of risk measures, *Mathematics of Operations Research* 45 (2019), 774–795.
- [27] D. P. Kouri and T. M. Surowiec: Risk-averse optimal control of semi-linear elliptic PDEs, *ESAIM: Control, Optimization and Calculus of Variations* 26 (2020) 53.
- [28] 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.
- [29] P. Lachout: Sensitivity of stochastic programs by means of the infimum functional, Proceedings of Prague Stochastics 2006, Prague, 495–504.
- [30] J. Milz: Sample average approximations of strongly convex stochastic programs in Hilbert spaces, Preprint arXiv:2104.05114v1.
- [31] 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.
- [32] D. N. Politis and J. P. Romano: Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics* 22 (1994), 2031–2050.

- [33] D. N. Politis, J. P. Romano, and M. Wolf: *Subsampling*, Springer Series in Statistics, Springer, New York, 1999.
- [34] S. T. Rachev: *Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester, 1991.
- [35] S. T. Rachev and W. Römisch, *Quantitative stability in stochastic programming: The method of probability metrics*, *Mathematics of Operations Research* **27** (2002), 792–818.
- [36] R. T. Rockafellar and S. Uryasev: Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* **26** (2002), 1443–1471.
- [37] W. Römisch: Delta method, infinite dimensional, in: *Encyclopedia of Statistical Sciences* (S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic eds.), Wiley, 2006.
- [38] A. Ruszczyński and A. Shapiro (eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, Amsterdam, 2003.
- [39] C. Schillings, S. Schmidt, and V. Schulz: Efficient shape optimization of certain and uncertain aerodynamic design, *Computers & Fluids* **46** (2011), 78–87.
- [40] A. Shapiro: On concepts of directional differentiability, *Journal Optimization Theory and Applications* **66** (1990), 477–487.
- [41] A. Shapiro: Perturbation analysis of optimization problems in Banach spaces, *Numerical Functional Analysis and Optimization* **13** (1992), 97–116.
- [42] 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.
- [43] I. Steinwart: Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties, *Potential Analysis* **51** (2019), 361–395.

- [44] M. Talagrand: Concentration of measure and isoperimetric inequalities in product spaces, *Publications Mathématiques de l'I.H.E.S.* 81 (1995), 73–205.
- [45] F. Topsøe: On the connection between P -continuity and P -uniformity in weak convergence, *Probability Theory and its Applications* 12 (1967), 281–290.
- [46] 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.
- [47] A. W. van der Vaart: New Donsker classes, *The Annals of Probability* 24 (1996), 2128–2140.
- [48] A. W. van der Vaart: *Asymptotic Statistics*, Cambridge University Press, Cambridge, 1998.
- [49] A. W. van der Vaart and J. A. Wellner: *Weak Convergence and Empirical Processes*, Springer, New York, 1996.
- [50] V. M. Zolotarev: Probability metrics, *Theory of Probability and its Applications* 28, 278–302.