ORIGINAL PAPER

On quantitative stability in infinite-dimensional optimization under uncertainty

M. Hoffhues[1] · W. Römisch[2] · T. M. Surowiec[1]

Abstract

The vast majority of stochastic optimization problems require the approximation of the underlying probability measure, e.g., by sampling or using observations. It is therefore crucial to understand the dependence of the optimal value and optimal solutions on these approximations as the sample size increases or more data becomes available. Due to the weak convergence properties of sequences of probability measures, there is no guarantee that these quantities will exhibit favorable asymptotic properties. We consider a class of infinite-dimensional stochastic optimization problems inspired by recent work on PDE-constrained optimization as well as functional data analysis. For this class of problems, we provide both qualitative and quantitative stability results on the optimal value and optimal solutions. In both cases, we make use of the method of probability metrics. The optimal values are shown to be Lipschitz continuous with respect to a minimal information metric and consequently, under further regularity assumptions, with respect to certain Fortet-Mourier and Wasserstein metrics. We prove that even in the most favorable setting, the solutions are at best Hölder continuous with respect to changes in the underlying measure. The theoretical results are tested in the context of Monte Carlo approximation for a numerical example involving PDE-constrained optimization under uncertainty.

✉ T. M. Surowiec
surowiec@mathematik.uni-marburg.de

W. Römisch
romisch@math.hu-berlin.de

[1] FB12 Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerwein-Straße 6, 35032 Marburg, Germany

[2] Institute of Mathematics, Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany

1 Introduction

In stochastic optimization, stability usually refers to the continuity properties of optimal values and solution sets as mappings from a set of probability measures, endowed with a suitable distance, into the extended reals and solution space, respectively, see [22]. The distance on the space of probability measures must be selected in order to allow the estimation of differences of the relevant functions, which depend on probability measures. There exists a wide variety of possible distances of probability measures based on various constructions [19,27]. In the present context, distances with ζ-structure introduced first in [27] appear as a natural choice. For a given metric space Ω such a distance is of the form

$$d_{\mathfrak{F}}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathfrak{F}} \left| \int_{\Omega} f(\omega) \, d\mathbb{P}(\omega) - \int_{\Omega} f(\omega) \, d\mathbb{Q}(\omega) \right|, \tag{1}$$

where $\mathfrak{F}$ is a family of Borel measurable functions from Ω to $\overline{\mathbb{R}}$ and $\mathbb{P}$, $\mathbb{Q}$ are Borel probability measures on Ω. Note that the distance $d_{\mathfrak{F}}$ is non-negative, symmetric and satisfies the triangle inequality. It also satisfies $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}) = 0$ and is, thus, a probability metric in the sense of [27]. However, $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{Q}) = 0$ only implies $\mathbb{P} = \mathbb{Q}$, when the family $\mathfrak{F}$ is rich enough. Hence, $d_{\mathfrak{F}}$ is a semi-metric in the usual terminology, in general.

The smallest relevant family $\mathfrak{F}$ of Borel measurable functions in our stability studies contains only those functions which appear in the stochastic optimization problem under consideration. In this case, $d_{\mathfrak{F}}$ may be called the minimal information (m.i.) distance. Stability results with respect to such m.i. distances serve as the starting point (i) to study stability with respect to the weak convergence of probability measures and (ii) to enlarge the family $\mathfrak{F}$ properly by functions sharing essential analytical properties with the original ones. The latter strategy may lead to probability metrics that enjoy desirable properties like dual representations and convergence characterizations.

This method of probability metrics provides quantitative statements on the stability of solutions and optimal values of stochastic programming problems. Nevertheless, the existing theory has not been developed for optimization problems in which the design or decision variables may be infinite-dimensional, as is the case in PDE-constrained optimization under uncertainty. By including infinite-dimensional feasible sets, we introduce a number of complications; in particular, the loss of norm compactness of the feasible set, even in the case of convex, closed, and bounded feasible sets.

After fixing some essential notation in Sect. 2, we state the class of infinite-dimensional stochastic optimization problems for which we study stability in the subsequent sections in Sect. 3. Section 4 contains qualitative results by providing conditions that imply convergence of optimal values and solutions if the underlying sequence of probability distribution converges to a limit distribution in some sense. In Sect. 5 we show that optimal values and solutions even allow Lipschitz or Hölder estimates in terms of the ζ-distance. In Sect. 6 we argue that the stability analysis of the preceding sections applies to certain stochastic PDE-constrained optimization problems. Finally, in Sect. 7, we provide a study of the results in Sect. 6 for the case when Monte Carlo approximations are used.

2 Notation and preliminary results

We assume throughout that Ω is a complete separable metric space, i.e., Polish space, and $\mathcal{F}$ the associated Borel σ-algebra. In addition, we will work exclusively with Borel probability measures $\mathbb{P} : \mathcal{F} \to [0, 1]$. that ensure $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space. In particular, if Ω is finite, then $\mathcal{F}$ must be the power set of Ω. For the abstract portion of our results, we will always assume that Θ is a real separable Hilbert space and $\Theta_{\mathrm{ad}} \subset \Theta$ is a nonempty, closed, and convex set. Given an appropriately chosen integrand $f : \Theta \times \Omega \to \overline{\mathbb{R}}$, we will consider the potentially infinite dimensional stochastic optimization problems:

$$\nu(\mathbb{P}) := \inf_{\theta \in \Theta_{\mathrm{ad}}} \int_{\Omega} f(\theta, \omega) \, \mathrm{d}\mathbb{P}(\omega). \tag{2}$$

Here, we also introduce the notion of optimal value function ν as a function from the space of all Borel probability measures $\mathcal{P}(\Omega)$ into $\overline{\mathbb{R}}$. This potentially extended real-valued function will play a key role in our discussions. If necessary, we will denote the expectation by either $\mathbb{E}$ or if it is not clear in context $\mathbb{E}_{\mathbb{P}}$ to denote the dependence on the measures $\mathbb{P}$.

Given a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a real Banach space W, we recall the definition of the Bochner space $L^p(\Omega, \mathcal{F}, \mathbb{P}; W)$ $p \in [1, \infty)$ as the space of (equivalence classes) of strongly measurable functions v, which map Ω into W and satisfy $\int_{\Omega} \|v(\omega)\|_W^p \, \mathrm{d}\mathbb{P}(\omega) < +\infty$, cf. [13]. If $p = \infty$, then $L^\infty(\Omega, \mathcal{F}, \mathbb{P}; W)$ consists of essentially bounded W-valued strongly measurable functions. In both cases $L^p(\Omega, \mathcal{F}, \mathbb{P}; W)$ is a Banach space with the natural norm(s)

$$\|v\|_{L^p(\Omega, \mathcal{F}, \mathbb{P}; W)} = \begin{cases} \left[\mathbb{E}\|v\|_W^p \right]^{1/p} & \text{, for } p \in [1, \infty), \\ \mathrm{ess\,sup}_{\omega \in \Omega} \|v(\omega)\|_W & \text{, for } p = \infty. \end{cases}$$

In the special case when $W = \mathbb{R}$, we simply write $L^p(\Omega, \mathcal{F}, \mathbb{P})$. As usual norm convergence will be typically denote by $\to$, whereas $\rightharpoonup$ signifies weak convergence and $\overset{*}{\rightharpoonup}$ weak-star convergence.

In our stability analysis, we make use of distances with ζ-structure on $\mathcal{P}(\Omega)$ having the form (1). We will refer to these objects as ζ-distances for brevity. Given a family $\mathfrak{F}$ of Borel measurable functions from Ω into $\overline{\mathbb{R}}$, the ζ-distance $d_{\mathfrak{F}}$ on $(\Omega, \mathcal{F})$ is a highly flexible structure that allows us to define so-called minimal information distances and Fortet-Mourier metrics; each defined in the text below. Properties of ζ-distances like a characterization of its maximal generator and its relation to weak convergence of probability measures can be found in [18,23]. Recall that a sequence of probability measures $\{\mathbb{P}_N\}$ on $(\Omega, \mathcal{F})$ is said to narrowly/weakly converge to the probability measure $\mathbb{P}$ provided

$$\mathbb{E}_{\mathbb{P}_N}[f] \to \mathbb{E}_{\mathbb{P}}[f] \quad \forall f \in C_b^0(\Omega),$$

where $C_b^0(\Omega)$ is the space of all bounded continuous functions on Ω. A family $\mathfrak{F}$ of Borel measurable functions is called a $\mathbb{P}$-uniformity class if

$$\lim_{N \to \infty} d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N) = 0$$

holds for each sequence $\{\mathbb{P}_N\}$ of probability measures converging weakly to $\mathbb{P}$. For example, it is known that $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class if $\mathfrak{F}$ is uniformly bounded and $\mathbb{P}(\{\omega \in \Omega : \mathfrak{F} \text{ is not equicontinuous at } \omega\}) = 0$ [23].

Finally, we recall that given a σ-algebra $\mathcal{F}$ along with a nominal σ-finite σ-additive positive measure $\mathbb{P}$ on Ω, e.g., a Borel probability measure $\mathbb{P} \in \mathcal{P}(\Omega)$, the dual space of $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ can be identified with the space of all finitely additive signed measures $\mathrm{ba}(\Omega)$ on $\mathcal{F}$ absolutely continuous with respect to $\mathbb{P}$, see e.g., [9].

3 The optimization problem

In order to carry out the stability analysis, we restrict the class of allowable integrands $f(\theta, \omega)$. These restrictions will henceforth be taken as standing assumptions. The particular class considered in this paper is inspired by applications in PDE-constrained optimization under uncertainty in which the PDE is given by a linear elliptic partial differential equation with random coefficients, right-hand side, and/or boundary conditions. We refer the reader to [17] for an overview of the state-of-the-art theory including more general objective functions and risk measures. In addition, many problems in functional data analysis exhibit practically the same form used below, see e.g., [21].

Let V and H be real Hilbert spaces such that V embeds continuously into H, and $\theta_d \in H$. For $\theta \in \Theta$ and $\omega \in \Omega$, let $\Sigma(\omega)\theta = S(\omega)\theta - s(\omega)$, where $S(\omega) : \Theta \to V$ is bounded and linear in θ independently of ω and $s(\omega) \in H$. We then define

$$f(\theta, \omega) := \frac{1}{2}\|\Sigma(\omega)\theta - \theta_d\|_H^2 = \frac{1}{2}\|S(\omega)\theta - (\theta_d + s(\omega))\|_H^2.$$

Furthermore, we assume that for every $\theta \in \Theta$ (or $\theta \in \Theta_{\mathrm{ad}}$) and any $\mathbb{P} \in \mathcal{P}(\Omega)$

$$f(\theta, \cdot) \in L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

This implicitly adds mild regularity assumptions on S and s that are typically fulfilled when S is related to the solution of a parametric elliptic PDE, e.g., $S(\cdot)\theta, s(\cdot) \in L^2(\Omega, \mathcal{F}, \mathbb{P}; V)$. Then for $\alpha > 0$, we consider the optimization problems

$$\inf_{\theta \in \Theta_{\mathrm{ad}}} F(\theta) := \mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_\Theta^2. \tag{3}$$

Theorem 1 *Problem* (3) *admits a unique solution* $\theta_{\mathbb{P}} \in \Theta_{\mathrm{ad}}$ *for every* $\mathbb{P} \in \mathcal{P}(\Omega)$.

Proof For existence, it suffices to prove F is proper, convex, lower-semicontinuous and coercive, cf. e.g., [3, Sec. 3.3]. Since $f(\theta, \cdot) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ for any $\theta \in \Theta_{\mathrm{ad}}$

and $f \geq 0$, F is proper. Convexity follows directly from the $\mathbb{P}$-a.e. convexity of $\theta \mapsto f(\theta, \omega) + \frac{\alpha}{2}\|\theta\|_{\Theta}^2$ and the monotonicity of the expectation $\mathbb{E}_{\mathbb{P}}$. Lower semicontinuity is a result of Fatou's lemma: Let $\theta_k \to \theta$ in Θ. Then since $f \geq 0$ and $f(\theta_k, \cdot) \to f(\theta, \cdot)$ $\mathbb{P}$-a.s. (by the assumptions on S) we have

$$\liminf_k \mathbb{E}_{\mathbb{P}}\left[f(\theta_k)\right] + \frac{\alpha}{2}\|\theta_k\|_{\Theta}^2 \geq \mathbb{E}_{\mathbb{P}}\left[\liminf_k f(\theta_k)\right] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2 = \mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2.$$

Since $\mathbb{E}_{\mathbb{P}}[f(\theta, \cdot)] \geq 0$ for all $\theta \in \Theta_{\text{ad}}$, F is coercive. Given F is proper, convex, and lower semicontinuous, F is weakly lower semicontinuous, as well. Since F is coercive, the level set $\{\theta \in \Theta_{\text{ad}} \,|\, F(\theta) \leq \alpha_0\}$, where $\theta_0 \in \Theta_{\text{ad}}$ and $\alpha_0 := F(\theta_0)$, is weakly sequentially compact. It then follows from the direct method that (3) admits a solution $\theta_{\mathbb{P}}$. Given $\alpha > 0$, F is strictly convex. Hence, $\theta_{\mathbb{P}}$ is unique. □

4 Qualitative stability

In this section, we provide stability results that ensure the approximating optimization problems obtained by replacing $\mathbb{P}$ by another probability measure $\mathbb{Q} \in \mathcal{P}(\Omega)$ will converge in some sense to the original problem. In particular, we show that the solutions $\theta_{\mathbb{Q}}$ will strongly converge to $\theta_{\mathbb{P}}$ provided $\mathbb{Q}$ converges to $\mathbb{P}$ with respect to a properly chosen ζ-distance. This basic result serves as the foundation needed to prove continuity of the solutions with respect to narrow convergence of probability measures. However, in order to do the latter, additional regularity properties will be required on the integrands with respect to ω. These stability results are in some sense more versatile than the quantitative results below. Nevertheless, they do not provide us with a rate of convergence.

Theorem 2 *In the context of Theorem 1, suppose we are given a sequence $\{\mathbb{P}_N\}$ with $\mathbb{P}_N \in \mathcal{P}(\Omega)$ and a probability measure $\mathbb{P} \in \mathcal{P}(\Omega)$ such that $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \to 0$, where $\mathfrak{F}$ is any class of measurable functions from Ω into $\overline{\mathbb{R}}$ large enough to contain $f(\theta, \cdot)$ for any $\theta \in \{\theta_N : N \in \mathbb{N}\} \cup \{\theta_{\mathbb{P}}\}$ with $\theta_N := \theta_{\mathbb{P}_N}$. Then $\theta_N \to \theta_{\mathbb{P}}$ strongly in Θ as $N \to +\infty$.*

Remark 1 The obvious candidate for the set $\mathfrak{F}$ would be to choose the collection of all possible integrands $f(\theta, \cdot) : \Omega \to \mathbb{R}$ indexed by $\theta \in \Theta_{\text{ad}}$. In terms of the associated ζ-distance, this would result in what is referred to in [19,20,22] as the minimal information metric.

Proof We first show $\{\theta_N\}$ is uniformly bounded in Θ. Indeed, we have

$$\frac{\alpha}{2}\|\theta_N\|_{\Theta}^2 \leq \mathbb{E}_{\mathbb{P}_N}\left[f(\theta_N)\right] + \frac{\alpha}{2}\|\theta_N\|_{\Theta}^2 \leq \mathbb{E}_{\mathbb{P}_N}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2 \quad \theta \in \Theta_{\text{ad}}. \quad (4)$$

For any fixed $\theta \in \Theta_{\text{ad}}$, it follows from the hypotheses that

$$\mathbb{E}_{\mathbb{P}_N}[f(\theta)] = \mathbb{E}_{\mathbb{P}_N}[f(\theta)] - \mathbb{E}_{\mathbb{P}}[f(\theta)] + \mathbb{E}_{\mathbb{P}}[f(\theta)] \leq d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) + \mathbb{E}_{\mathbb{P}}[f(\theta)]. \quad (5)$$

Substituting this into (4) we obtain the bound

$$\frac{\alpha}{2}\|\theta_N\|_\Theta^2 \le d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) + F(\theta) \quad \theta \in \Theta_{\mathrm{ad}}.$$

Since $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \to 0$, $\{\theta_N\}$ is bounded in Θ. Therefore, there exists a $\widehat{\theta} \in \Theta_{\mathrm{ad}}$ and a weakly convergent subsequence $\{\theta_{N_\ell}\}$ such that $\theta_{N_\ell} \rightharpoonup \widehat{\theta}$ as $\ell \to +\infty$.

For fixed $\mathbb{P}$, it follows from the proof of Theorem 1 that $\mathbb{E}_{\mathbb{P}}[f(\cdot)] : \Theta \to \mathbb{R}$ is weakly lower semicontinuous. Therefore,

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[f\left(\widehat{\theta}\right)\right] + \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 &\le \liminf_\ell \mathbb{E}_{\mathbb{P}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_\Theta^2 \\
&\le \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_\Theta^2 + \mathbb{E}_{\mathbb{P}}\left[f\left(\theta_{N_\ell}\right)\right] - \mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right]\right] \\
&\le \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_\Theta^2 + d_{\mathfrak{F}}\left(\mathbb{P}_{N_\ell}, \mathbb{P}\right)\right] \\
&= \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_\Theta^2\right].
\end{aligned}
\tag{6}
$$

It then follows from (6), the optimality of θ_{N_ℓ}, and (5) that for any $\theta \in \Theta_{\mathrm{ad}}$ we have:

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] + \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 &\le \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_\Theta^2\right] \\
&\le \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_\Theta^2\right] \\
&\le \liminf_\ell d_{\mathfrak{F}}(\mathbb{P}_{N_\ell}, \mathbb{P}) + \mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_\Theta^2 \\
&= \mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_\Theta^2.
\end{aligned}
\tag{7}
$$

Hence, $\theta_{\mathbb{P}} = \widehat{\theta}$. Since $\theta_{\mathbb{P}}$ is unique and the previous arguments hold for all weakly convergent subsequences of $\{\theta_N\}$, we have $\theta_N \rightharpoonup \theta_{\mathbb{P}} = \widehat{\theta}$ as $N \to +\infty$. It remains to prove $\|\theta_N - \theta_{\mathbb{P}}\|_\Theta \to 0$.

Clearly we have the inequality

$$\liminf_N \|\theta_N\|_\Theta \ge \|\widehat{\theta}\|_\Theta \tag{8}$$

by weak lower semicontinuity of the norm $\|\cdot\|_\Theta$. On the other hand, by rearranging terms, the definition of θ_N and feasibility of $\widehat{\theta}$ yield

$$
\begin{aligned}
\frac{\alpha}{2}\|\theta_N\|_\Theta^2 &\le \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}_N}[f(\theta_N)] \\
&= \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}_N}[f(\theta_N)] + \mathbb{E}_{\mathbb{P}}[f(\theta_N)] - \mathbb{E}_{\mathbb{P}}[f(\theta_N)] \\
&\le \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\theta_N)] + d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \\
&= \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] + \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\theta_N)] + d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \\
&\le \frac{\alpha}{2}\|\widehat{\theta}\|_\Theta^2 + 2d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) + \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\theta_N)].
\end{aligned}
$$

Therefore, we again appeal to the weak lower semicontinuity of $\mathbb{E}_{\mathbb{P}}[f(\cdot)]$ on Θ to obtain

$$
\begin{aligned}
\limsup_N \frac{\alpha}{2}\|\theta_N\|_{\Theta}^2 &\leq \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 + \limsup_N \left[2d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) + \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\theta_N)]\right] \\
&= \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 + \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] - \liminf_N[f(\theta_N)]] \\
&\leq \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 + \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})] \\
&= \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2.
\end{aligned}
\tag{9}
$$

Combining (8) and (9), we have $\|\theta_N\|_{\Theta} \to \|\widehat{\theta}\|_{\Theta}$. Then since Θ is a Hilbert space and $\theta_N \rightharpoonup \widehat{\theta}$, the assertion follows. □

An alternative perspective on qualitative stability is offered by our next result. Here, we will prove convergence of the sequence of minimizers under different data assumptions on the integrands and a different form of weak convergence of measures. We note that in PDE-constrained optimization under uncertainty these assumptions are less restrictive than they may appear. In particular, we do not require $f(\theta, \cdot) : \Omega \to \mathbb{R}$ to be continuous as is needed below for the Fortet-Mourier metric. The caveat here is the requirement that $\mathbb{P}_N$ is absolutely continuous with respect to $\mathbb{P}$.

Theorem 3 *In addition to the standing assumptions, fix some $\mathbb{P} \in \mathcal{P}(\Omega)$ and suppose that for all $\theta \in \Theta_{\mathrm{ad}}$ $f(\theta, \cdot) \in L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$. Assume furthermore that the superposition operator $\Phi : \Theta \to L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ defined by*

$$
\Phi(\theta)(\omega) := f(\theta, \omega)
$$

is completely continuous. Let $\{\mathbb{P}_N\} \subset \mathcal{P}(\Omega)$ such that

1. for all $N \in \mathbb{N}$ $\mathbb{P}_N << \mathbb{P}$ ($\mathbb{P}_N$ is absolutely continuous with respect to $\mathbb{P}$) and
2. $\mathbb{P}_N \to \mathbb{P}$ with respect to the weak-star topology on $(L^{\infty}(\Omega, \mathcal{F}, \mathbb{P}))^*$.

Then $\theta_N \to \theta_{\mathbb{P}}$.

Proof As noted in Sect. 2, each $\mathbb{Q} \in \mathcal{P}(\Omega)$ is an element of $(L^{\infty}(\Omega, \mathcal{F}, \mathbb{P}))^*$ provided $\mathbb{Q} << \mathbb{P}$. The rest of the proof mirrors that of Theorem 2. Given the sequence of minimizers $\{\theta_N\}$ we immediately obtain a uniform bound on $\|\theta_N\|$ from (4) since for any $\theta \in \Theta_{\mathrm{ad}}$ $f(\theta, \cdot) \in L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{P}_N \overset{*}{\rightharpoonup} \mathbb{P}$. As before, we let $\{\theta_{N_\ell}\}_{\ell=1}^{\infty}$ denote the weakly convergent subsequence and $\widehat{\theta}$ the associated weak limit.

Turning now to the estimate derived in (6), we see that

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[f(\widehat{\theta})\right] + \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 &\leq \liminf_\ell \mathbb{E}_{\mathbb{P}}\left[f(\theta_{N_\ell})\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_{\Theta}^2 \\
&\leq \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f(\theta_{N_\ell})\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_{\Theta}^2 + \mathbb{E}_{\mathbb{P}}\left[f(\theta_{N_\ell})\right] - \mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f(\theta_{N_\ell})\right]\right] \\
&= \liminf_\ell \left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f(\theta_{N_\ell})\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_{\Theta}^2\right].
\end{aligned}
\tag{10}
$$

Here, $\Phi(\theta_{N_l}) \to \Phi(\widehat{\theta})$ strongly in $L^{\infty}(\Omega, \mathcal{F}, \mathbb{P})$ due the assumption of complete continuity. Therefore, both $\mathbb{E}_{\mathbb{P}}[f(\theta_{N_l})]$ and $\mathbb{E}_{\mathbb{P}_{N_l}}[f(\theta_{N_l})]$ converge to $\mathbb{E}_{\mathbb{P}}[f(\widehat{\theta})]$.

As in the proof of Theorem 2, we obtain optimality of $\widehat{\theta}$ by adapting the inequality (7), i.e., for every $\theta \in \Theta_{\mathrm{ad}}$ we have

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[f\left(\widehat{\theta}\right)\right] + \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 &\leq \liminf_{\ell}\left[\mathbb{E}_{\mathbb{P}_{N_\ell}}\left[f\left(\theta_{N_\ell}\right)\right] + \frac{\alpha}{2}\|\theta_{N_\ell}\|_{\Theta}^2\right] \\
&\leq \liminf_{\ell}\left[\mathbb{E}_{\mathbb{P}_{N_\ell}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2\right] \\
&= \mathbb{E}_{\mathbb{P}}[f(\theta)] + \frac{\alpha}{2}\|\theta\|_{\Theta}^2.
\end{aligned}
\tag{11}
$$

Here, the regularity of the integrand ensures that $\mathbb{E}_{\mathbb{P}_{N_\ell}}[f(\theta)]$ converges to $\mathbb{E}_{\mathbb{P}}[f(\theta)]$; from which it follows that $\widehat{\theta} = \theta_{\mathbb{P}}$. As in the proof of Theorem 2, we can again argue that the entire sequence $\{\theta_{\mathbb{P}_N}\}$ weakly converges to $\theta_{\mathbb{P}}$.

In order to prove norm convergence, we note that

$$
\frac{\alpha}{2}\|\theta_N\|_{\Theta}^2 \leq \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}_N}[f(\theta_N)].
$$

Then by the complete continuity and regularity assumptions, we have

$$
\limsup_{N} \frac{\alpha}{2}\|\theta_N\|_{\Theta}^2 \leq \limsup_{N} \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2 + \mathbb{E}_{\mathbb{P}_N}[f(\widehat{\theta})] - \mathbb{E}_{\mathbb{P}_N}[f(\theta_N)] = \limsup_{N} \frac{\alpha}{2}\|\widehat{\theta}\|_{\Theta}^2.
$$

This completes the proof.   □

Next, we return to the setting using probability metrics to obtain some important implications of the Theorem 2 under further regularity assumptions on the integrands. In our setting, we recall that the space of all (Borel) probability measures with finite p-th moments is defined by

$$
\mathcal{P}_p(\Omega) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \,\middle|\, \int_{\Omega} d(\omega_0, \omega)^p \mathrm{d}\mathbb{P}(\omega) < +\infty \right\},
$$

for some arbitrary $\omega_0 \in \Omega$. We recall that a sequence $\{\mathbb{P}_N\} \subset \mathcal{P}_p(\Omega)$ converges weakly (narrowly) provided for all $\varphi \in C_b^0(\Omega)$

$$
\mathbb{E}_{\mathbb{P}_N}[\varphi] \to \mathbb{E}_{\mathbb{P}}[\varphi] \text{ and } \mathbb{E}_{\mathbb{P}_N}\left[d(\omega_0, \cdot)^p\right] \to \mathbb{E}_{\mathbb{P}}\left[d(\omega_0, \cdot)^p\right]
$$

as $N \to \infty$. This type of weak convergence shares an intimate link with a certain class of ζ-distances known as Fortet-Mourier metrics. To start, for $p \in [1, \infty)$, we define the sets $\mathcal{F}_p(\Omega)$ of locally Lipschitz functions with a certain p-related growth condition by

$$\mathcal{F}_p(\Omega) := \{f : \Omega \to \mathbb{R} \mid |f(\omega_1) - f(\omega_2)|$$
$$\leq \max\left\{1, d(\omega_1, \omega_0)^{p-1}, d(\omega_2, \omega_0)^{p-1}\right\} d(\omega_1, \omega_2) \quad \forall \omega_1, \omega_2 \in \Omega\right\}.$$

We then define the Fortet-Mourier metric of order p for two measures $\mathbb{P}, \mathbb{Q} \in \mathcal{P}_p(\Omega)$ by

$$\zeta_p(\mathbb{P}, \mathbb{Q}) = d_{\mathcal{F}_p(\Omega)}(\mathbb{P}, \mathbb{Q}).$$

In particular, ζ_p is equivalent to the so-called Kantorovich-Rubinstein functional with cost function given by

$$c(\omega_1, \omega_2) = \max\left\{1, d(\omega_1, \omega_0)^{p-1}, d(\omega_2, \omega_0)^{p-1}\right\} d(\omega_1, \omega_2).$$

(see [19, Theorem 5.3.3] along with the discussion on page 93 in [19]). Furthermore, it follows from [19, Theorem 6.2.1] that $\{\mathbb{P}_N\} \subset \mathcal{P}_p(\Omega)$ converges weakly (narrowly) to $\mathbb{P} \in \mathcal{P}_p(\Omega)$ if and only if $\zeta_p(\mathbb{P}_N, \mathbb{P}) \to 0$ as $N \to +\infty$. We may therefore connect Theorem 2 directly to the weak convergence of probability measures.

Proposition 1 *In the setting of Theorem 2, suppose there exists a $p \in [1, \infty)$ and some $L > 0$ such that*

$$\mathfrak{F} = \mathcal{F}_p(\Omega) \text{ and } \left\{\frac{1}{L} f(\theta, \cdot) : \Omega \to \mathbb{R} \mid \theta \in \Theta_{\mathrm{ad}}\right\} \subset \mathfrak{F}.$$

Then the solution mapping $\mathcal{P}_p(\Omega) \ni \mathbb{Q} \mapsto \theta_{\mathbb{Q}} \in \Theta_{\mathrm{ad}}$ is continuous with respect to weak (narrow) convergence of probability measures on $\mathcal{P}_p(\Omega)$.

Proof After rescaling the integrands by $1/L > 0$, this is a direct consequence of Theorem 2 in light of the preceding arguments. □

Finally, we note that an alternative means of obtaining the sequential convergence result in Proposition 1 would be to appeal to the link between the weak topology on $\mathcal{P}_p(\Omega)$ and the topologies generated by the well-known Wasserstein distance W_p of order p. Let γ_i ($i = 1, 2$) be the projection onto the first or second term of $\Omega \times \Omega$, respectively, and for $\pi \in \mathcal{P}(\Omega \times \Omega)$ denote the marginals by $\pi^i := \pi_{\#}\gamma_i := \pi \circ \gamma_i^{-1}$. Then the Wasserstein distance of order p is given by

$$W_p^p(\mathbb{P}, \mathbb{Q}) = \inf\left\{\int_{\Omega \times \Omega} d(\omega_1, \omega_2)^p \mathrm{d}\pi(\omega_1, \omega_2) \,\Big|\, \pi \in \mathcal{P}(\Omega \times \Omega), \pi^1 = \mathbb{P} \text{ and } \pi^2 = \mathbb{Q}\right\}.$$

For this distance we have the estimate:

$$\zeta_p(\mathbb{P}, \mathbb{Q}) \leq \left(1 + \int_{\Omega} d(\omega_0, \omega)^p \, \mathrm{d}\mathbb{P}(\omega) + \int_{\Omega} d(\omega_0, \omega)^p \, \mathrm{d}\mathbb{Q}(\omega)\right)^{\frac{p-1}{p}} W_p(\mathbb{P}, \mathbb{Q}). \tag{12}$$

Therefore, if we start with a sequence of Borel probability measures $\{\mathbb{P}_N\}$ and $\mathbb{P} \in \mathcal{P}(\Omega)$ such that $W_p(\mathbb{P}_N, \mathbb{P}) \to 0$, then we obtain the same statement as in Proposition 1. However, as shown in [20] the convergence in the Wasserstein metric is potentially strictly slower than in the Fortet-Mourier metric.

5 Quantitative stability

As mentioned above, quantitative stability provides us with Lipschitz or Hölder-type estimates of the optimal values and solutions. This is first done using the "weakest" possible ζ-distance $d_\mathfrak{F}$ in which $\mathfrak{F}$ is directly related to the integrands without additional regularity assumptions on the dependence on ω. Further estimates related to Fortet-Mourier and Wasserstein metrics then follow as corollaries under Lipschitz conditions on the integrands.

Theorem 4 *Under the standing assumptions, let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(\Omega)$ and let $\mathfrak{F}$ be any set of Borel measurable functions that contains $g_\theta(\cdot) := f(\theta, \cdot)$, where $\theta = \theta_\mathbb{P}$ and $\theta_\mathbb{Q}$. Then we have the estimates:*

$$|\nu(\mathbb{Q}) - \nu(\mathbb{P})| \leq d_\mathfrak{F}(\mathbb{Q}, \mathbb{P}) \tag{13}$$

$$\|\theta_\mathbb{Q} - \theta_\mathbb{P}\| \leq 2\sqrt{\frac{2}{\alpha} d_\mathfrak{F}(\mathbb{Q}, \mathbb{P})}. \tag{14}$$

Proof For the Lipschitz estimate (13), we observe that

$$
\begin{aligned}
|\nu(\mathbb{Q}) - \nu(\mathbb{P})| &= \max\left\{\nu(\mathbb{Q}) - \nu(\mathbb{P}), \nu(\mathbb{P}) - \nu(\mathbb{Q})\right\} \\
&= \max\left\{\mathbb{E}_\mathbb{Q}\left[f(\theta_\mathbb{Q})\right] + \frac{\alpha}{2}\|\theta_\mathbb{Q}\|^2 - \mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{P})\right] - \frac{\alpha}{2}\|\theta_\mathbb{P}\|^2, \right. \\
&\qquad\quad \left. \mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{P})\right] + \frac{\alpha}{2}\|\theta_\mathbb{P}\|^2 - \mathbb{E}_\mathbb{Q}\left[f(\theta_\mathbb{Q})\right] - \frac{\alpha}{2}\|\theta_\mathbb{Q}\|^2\right\} \\
&\leq \max\left\{\mathbb{E}_\mathbb{Q}[f(\theta_\mathbb{P})] - \mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{P})\right], \mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{Q})\right] - \mathbb{E}_\mathbb{Q}\left[f(\theta_\mathbb{Q})\right]\right\} \\
&\leq \max\{|\mathbb{E}_\mathbb{Q}\left[f(\theta_\mathbb{P})\right] - \mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{P})\right]|, \\
&\qquad\quad |\mathbb{E}_\mathbb{P}\left[f(\theta_\mathbb{Q})\right] - \mathbb{E}_\mathbb{Q}[f(\theta_\mathbb{Q})]|\} \\
&\leq d_\mathfrak{F}(\mathbb{Q}, \mathbb{P}).
\end{aligned}
$$

For the Hölder estimate on the solution mapping (14), we start by letting $\delta := d_\mathfrak{F}(\mathbb{Q}, \mathbb{P})$ and observing that

$$
\begin{aligned}
2\delta &\geq \delta + |\nu(\mathbb{Q}) - \nu(\mathbb{P})| \\
&\geq \delta + \nu(\mathbb{Q}) - \nu(\mathbb{P}) \\
&= \delta + \mathbb{E}_\mathbb{Q}[f(\theta_\mathbb{Q})] + \frac{\alpha}{2}\|\theta_\mathbb{Q}\|^2 - \mathbb{E}_\mathbb{P}[f(\theta_\mathbb{P})] - \frac{\alpha}{2}\|\theta_\mathbb{P}\|^2 \\
&= \delta \pm \left(\mathbb{E}_\mathbb{P}[f(\theta_\mathbb{Q})] + \frac{\alpha}{2}\|\theta_\mathbb{Q}\|^2\right) + \mathbb{E}_\mathbb{Q}[f(\theta_\mathbb{Q})] + \frac{\alpha}{2}\|\theta_\mathbb{Q}\|^2 - \mathbb{E}_\mathbb{P}[f(\theta_\mathbb{P})] - \frac{\alpha}{2}\|\theta_\mathbb{P}\|^2 \\
&= \delta - \left(\mathbb{E}_\mathbb{P}[f(\theta_\mathbb{Q})] - \mathbb{E}_\mathbb{Q}[f(\theta_\mathbb{Q})]\right)
\end{aligned}
$$

$$+ \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{Q}})] + \frac{\alpha}{2}\|\theta_{\mathbb{Q}}\|^2 - \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}})] - \frac{\alpha}{2}\|\theta_{\mathbb{P}}\|^2$$

$$\geq \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{Q}})] + \frac{\alpha}{2}\|\theta_{\mathbb{Q}}\|^2 - \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}})] - \frac{\alpha}{2}\|\theta_{\mathbb{P}}\|^2. \tag{15}$$

Using the quadratic term $\frac{\alpha}{2}\|\cdot\|^2$, the convexity of the integrand $f(\cdot, \omega)$, the convexity of Θ_{ad}, and optimality of $\theta_{\mathbb{P}}$, we have

$$\mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}})] + \frac{\alpha}{2}\|\theta_{\mathbb{P}}\|^2 \leq \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}}/2 + \theta_{\mathbb{Q}}/2)] + \frac{\alpha}{2}\|\theta_{\mathbb{P}}/2 + \theta_{\mathbb{Q}}/2\|^2$$

$$\leq \frac{1}{2}(\mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}})] + \frac{\alpha}{2}\|\theta_{\mathbb{P}}\|^2)$$

$$+ \frac{1}{2}(\mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{Q}})] + \frac{\alpha}{2}\|\theta_{\mathbb{Q}}\|^2) - \frac{\alpha}{8}\|\theta_{\mathbb{P}} - \theta_{\mathbb{Q}}\|^2.$$

It follows that

$$\mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{Q}})] + \frac{\alpha}{2}\|\theta_{\mathbb{Q}}\|^2 - \mathbb{E}_{\mathbb{P}}[f(\theta_{\mathbb{P}})] - \frac{\alpha}{2}\|\theta_{\mathbb{P}}\|^2 \geq \frac{\alpha}{8}\|\theta_{\mathbb{Q}} - \theta_{\mathbb{P}}\|^2. \tag{16}$$

Combining (16) with (15) above yields

$$\|\theta_{\mathbb{Q}} - \theta_{\mathbb{P}}\| \leq \sqrt{\frac{8\delta}{\alpha}} = 2\sqrt{\frac{2d_{\mathfrak{F}}(\mathbb{Q}, \mathbb{P})}{\alpha}}$$

as was to be shown. $\qquad\square$

We may now return to the results at the end of Sect. 4 in order to derive quantitative stability results using the familiar Fortet-Mourier and Wasserstein distances.

Corollary 1 *In the setting of Theorem 4, suppose there exists a $p \in [1, \infty)$ and some $L > 0$ such that*

$$\mathfrak{F} = \mathcal{F}_p(\Omega) \text{ and } \left\{\frac{1}{L}f(\theta, \cdot) : \Omega \to \mathbb{R} \mid \theta \in \Theta_{\mathrm{ad}}\right\} \subset \mathfrak{F}.$$

Then the following estimates hold for the associated Fortet-Mourier metric:

$$|\nu(\mathbb{Q}) - \nu(\mathbb{P})| \leq L\zeta_p(\mathbb{Q}, \mathbb{P}), \quad \|\theta_{\mathbb{Q}} - \theta_{\mathbb{P}}\| \leq 2L\sqrt{\frac{2}{\alpha}\zeta_p(\mathbb{Q}, \mathbb{P})}.$$

Consequently, the following estimates hold for the Wasserstein metric W_p:

$$|\nu(\mathbb{Q}) - \nu(\mathbb{P})| \leq Lc(\omega_0, \mathbb{P}, \mathbb{Q})W_p(\mathbb{Q}, \mathbb{P}),$$

$$\|\theta_{\mathbb{Q}} - \theta_{\mathbb{P}}\| \leq 2L\sqrt{\frac{2}{\alpha}c(\omega_0, \mathbb{P}, \mathbb{Q})W_p(\mathbb{Q}, \mathbb{P})}.$$

Here, we set $c(\omega_0, \mathbb{P}, \mathbb{Q}) = \left(1 + \int_{\Omega} d(\omega_0, \omega)^p \mathrm{d}\mathbb{P}(\omega) + \int_{\Omega} d(\omega_0, \omega)^p \mathrm{d}\mathbb{Q}(\omega)\right)^{\frac{p-1}{p}}$.

6 An application to PDE-constrained optimization under uncertainty

We conclude the theoretical discussion with an example from PDE-constrained optimization under uncertainty to demonstrate the applicability of our results. For the purpose of discussion, we start with an arbitrary Borel probability measure $\mathbb{P}$ in order to introduce the problem. The notation mirrors in part that of [17]. For readability, we indicate the associated quantities in the general discussions above.

Our goal is twofold. Under reasonable data assumptions, we will define a class of integrands $\mathfrak{F}$, which allow us to use the m.i. metric in our stability results to prove convergence of optimal values and optimal solutions under weak convergence of a sequence of measures $\{\mathbb{P}_N\}$. Afterwards, assuming the underlying function spaces are replaced by finite-dimensional subspaces defined by a standard finite-element discretization, we derive an a priori-type error bound and argue that the fully discrete problems converge to the original continuous problems.

We will consider a class of optimization problems in which we seek to minimize the objective function

$$
\begin{aligned}
\mathcal{J}(u, z) &:= \frac{1}{2} \int_\Omega \int_D |u(x, \omega) - \widetilde{u}_d(x)|^2 \, \mathrm{d}x \mathrm{d}\mathbb{P}(\omega) + \frac{\alpha}{2} \int_D |z(x)|^2 \, \mathrm{d}x \\
&= \frac{1}{2} \mathbb{E}_\mathbb{P} \left[\|u - \widetilde{u}_d\|_{L^2(D)}^2 \right] + \frac{\alpha}{2} \|z\|_{L^2(D)}^2
\end{aligned}
\tag{17}
$$

subject to the condition that $z \in Z_{\mathrm{ad}} \subset L^2(D)$, a closed bounded convex set, and for $z \in Z_{\mathrm{ad}}$, u solves a random partial differential equation (PDE), which we define below. The function $\widetilde{u}_d \in L^2(D)$ can be thought of as a desired state or general target function.

To be precise, let $D \subset \mathbb{R}^n$ be an open, bounded Lipschitz domain, $V = H_0^1(D)$ the classical Sobolev space with inner product $(\cdot, \cdot)_V$, and $V^\star = H^{-1}(D)$ its dual with norm $\|\cdot\|_\star$ und dual pairing $\langle \cdot, \cdot \rangle$. In addition, let $H = L^2(D)$ with inner product $(\cdot, \cdot)_H$. Furthermore, let Ω be a metric space with metric ρ and Borel σ-field $\mathcal{F}$ and let $\mathbb{P}$ be a Borel probability measure.

Within this framework, we consider the bilinear form $a(\cdot, \cdot; \omega) : V \times V \to \mathbb{R}$ defined by

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

where $\omega \in \Omega$. The associated random PDE can be defined pointwise as:

$$
a(u, v; \omega) = \int_D (z(x) + g(x, \omega))v(x)dx \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega,
$$

for all test functions $v \in C_0^\infty(D)$, z varying in a constraint set $Z_{\mathrm{ad}} \subset H$, and $g, b_{ij} : D \times \Omega \to \mathbb{R}$, which are assumed to be at least measurable in Ω and square (Lebesgue) integrable in D.

In order to use our stability results in this context, we will need further data assumptions on the bilinear form. For each $\omega \in \Omega$, we let $A(\omega) : V \to V^\star$ be the mapping

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

The existence of $A(\omega)$ as a bounded linear operator is due to the Riesz representation theorem and the Lax-Milgram lemma based on the following assumptions, see e.g., [1, Chap. 6., 6.1, 6.2]. First, we impose the condition that there exist $L > \gamma > 0$ such that

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

for all $x \in D$ and $\mathbb{P}$-a.e. $\omega \in \Omega$. This implies that each b_{ij} is essentially bounded in $D \times \Omega$ with respect to the associated product measure. Moreover, the mapping $A(\omega) : V \to V^\star$ is uniformly positive definite (with constant γ) and uniformly bounded (with constant L) with respect to $\mathbb{P}$-a.e. $\omega \in \Omega$, i.e.,

$$\gamma \|u\|_V^2 \leq \langle A(\omega)u, u \rangle \leq L\|u\|_V^2 \quad (\forall u \in V).$$

In addition, the inverse mapping $A(\omega)^{-1} : V^\star \to V$ is again uniformly positive definite (with constant $\frac{1}{L}$) and uniformly bounded (with constant $\frac{1}{\gamma}$) with respect to $\mathbb{P}$-a.e. $\omega \in \Omega$.

Under these data assumptions, we may now define a class of integrands for the m.i. metric for which $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) \to 0$ for any sequence of Borel probability measures $\{\mathbb{P}_N\}$ that converges weakly to $\mathbb{P}$. To this end, we define the functions $f : Z_{\text{ad}} \times \Omega \to \mathbb{R}$ by

$$f(z, \omega) = \frac{1}{2}\|A(\omega)^{-1}(z + g(\omega)) - \widetilde{u}_d\|_H^2 = \frac{1}{2}\|A(\omega)^{-1}z - \left(\widetilde{u}_d - A(\omega)^{-1}g(\omega)\right)\|_H^2$$
$$= \frac{1}{2}\int_D ([A(\omega)^{-1}z](x) - (\widetilde{u}_d(x) - [A(\omega)^{-1}g(\cdot, \omega)](x)))^2 dx.$$

Our aim is to derive conditions implying that the class $\mathfrak{F} = \{f(z, \cdot) : z \in Z_{\text{ad}}\}$ is uniformly bounded and equicontinuous, and consequently a $\mathbb{P}$-uniformity class, cf. [23].

Lemma 1 *In addition to the standing assumptions, suppose that $\widetilde{u}_d \in H$ and $g \in L^2(\Omega, \mathcal{F}, \mathbb{P}; V^\star)$. Then for some $C > 0$ we have*

$$|f(z, \omega)| \leq C(1 + \|g(\omega)\|_\star^2) \quad (\mathbb{P}\text{- a.e. } \omega \in \Omega, \, z \in Z_{\text{ad}}).$$

Proof For $z \in Z_{\text{ad}}$ and $\omega \in \Omega$ we obtain

$$|f(z, \omega)| \leq \left(\|A(\omega)^{-1}z\|_H^2 + \|\widetilde{u}_d - A(\omega)^{-1}g(\omega)\|_H^2\right)$$
$$\leq \left(\|A(\omega)^{-1}z\|_H^2 + 2\|\widetilde{u}_d\|_H^2 + 2\|A(\omega)^{-1}g(\omega)\|_H^2\right)$$

$$\leq \left(\|A(\omega)^{-1}z\|_V^2 + 2\|A(\omega)^{-1}g(\omega)\|_V^2 + 2\|\widetilde{u}_d\|_H^2 \right)$$

$$\leq \left(\frac{c}{\gamma}(\|z\|_\star^2 + 2\|g(\omega)\|_\star^2) + 2\|\widetilde{u}_d\|_H^2 \right),$$

where we used the Poincaré-Friedrichs inequality twice (with some constant c) and the uniform boundedness of $\|A(\omega)^{-1}\|$ by $\frac{1}{\gamma}$. Since z varies in the bounded set Z_{ad}, there is a positive constant C such that the assertion holds. □

Lemma 1 provides us with a uniform bound on all functions in $\mathfrak{F}$. The proof of the following Lemma makes use of a result in [12]. Since this book is not readily available in English, we provide it and a short proof in the "Appendix".

Lemma 2 *In addition to the assumptions of Lemma 1, suppose there is a constant $C > 0$ such that*

$$\sqrt{\sum_{i,j=1}^n |b_{ij}(x,\omega) - b_{ij}(x,\omega')|^2} \leq C\rho(\omega,\omega') \quad (\forall \omega, \omega' \in \Omega).$$

Then for any $g \in V^\star$ and $\omega, \omega' \in \Omega$ we have

$$\|A(\omega)^{-1}g - A(\omega')^{-1}g\|_V \leq \frac{t}{1 - \kappa(t)}C\|g\|_\star \rho(\omega,\omega') \quad \forall t \in \left(0, \frac{2\gamma}{L^2} \right),$$

where $\kappa(t) = \sqrt{1 - 2\gamma t + L^2 t^2}$.

Proof First we study the dependence of $A(\omega)u$ on ω. Let $u, v \in V$ and $\omega, \omega' \in \Omega$.

$$|\langle (A(\omega) - A(\omega'))u, v \rangle| = \left| \int_D \sum_{i,j=1}^n (b_{ij}(x,\omega) - b_{ij}(x,\omega')) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} dx \right|$$

$$= \left| \left(B(\cdot; \omega, \omega') \nabla u(\cdot), \nabla v(\cdot) \right) \right|$$

$$\leq \int_D \|B(x; \omega, \omega')\| |\nabla u(x)| |\nabla v(x)| dx$$

$$\leq \sup_{x \in D} \|B(x; \omega, \omega')\| \|u\|_V \|v\|_V,$$

where $B(x; \omega, \omega')$ denotes the $n \times n$-matrix

$$B(x; \omega, \omega') = (b_{ij}(x,\omega) - b_{ij}(x,\omega'))_{i,j=1,\dots,n}$$

with Frobenius norm $\|B(x; \omega, \omega')\|$, ∇u the gradient of u in the sense of Sobolev and $|\nabla u|$ its Euclidean norm. Hence, we obtain

$$\|(A(\omega) - A(\omega'))u\|_\star \leq \|u\|_V \sup_{x \in D} \sqrt{\sum_{i,j=1}^n |b_{ij}(x,\omega) - b_{ij}(x,\omega')|^2} \leq C\|u\|_V \rho(\omega,\omega').$$

Next we consider the mapping $K_t(\omega)u = u - tJ^{-1}(A(\omega)u - g)$ for some $t \in (0, \frac{2\gamma}{L^2})$, $\omega \in \Omega$, $g \in V^\star$ and any $u \in V$. Then it follows from Proposition 2 that

$$\|K_t(\omega)u - K_t(\omega)u'\|_V \leq \kappa(t)\|u - u'\|_V$$

for any $u, u' \in V$ and $\kappa(t) = \sqrt{1 - 2t\gamma + t^2L^2} < 1$. Furthermore, the unique fixed point of $K_t(\omega)$ belongs to the ball around zero with radius

$$r = \frac{\|K_t 0 - 0\|_V}{1 - \kappa(t)} = \frac{t}{1 - \kappa(t)}\|J^{-1}g\|_V = \frac{t}{1 - \kappa(t)}\|g\|_\star.$$

For any $u \in V$ and $\omega, \omega' \in \Omega$ we have

$$\|K_t(\omega)u - K_t(\omega')u\|_V = t\|J^{-1}(A(\omega) - A(\omega'))u\|_V = t\|(A(\omega) - A(\omega'))u\|_\star$$

and apply Proposition 3 from the "Appendix" with $P = \Omega$, $X = \mathbb{B}(0, r) = \{u \in V : \|u\|_V \leq r\}$, $F(p, u) = K_t(\omega)u$ and $F(p', u) = K_t(\omega')u$. We obtain

$$\|\bar{x}(g, \omega) - \bar{x}(g, \omega')\| \leq \frac{t}{1 - \kappa(t)}Cr\rho(\omega, \omega')$$

where $\bar{x}(g, \omega) = A(\omega)^{-1}g$ and $r = \frac{t}{1-\kappa(t)}\|g\|_\star$. This completes the proof. $\qquad\square$

We now have enough results to prove that $\mathfrak{F}$ constitutes a $\mathbb{P}$-uniformity class, which is a direct consequence of the following theorem.

Theorem 5 *In addition to the hypotheses of Lemma 2, assume $g \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}; H)$ and there exists $\bar{C} > 0$ such that*

$$|g(x, \omega) - g(x, \omega')| \leq \bar{C}\rho(\omega, \omega') \quad (\forall \omega, \omega' \in \Omega, a.\, e.\, in\, D).$$

Then $\mathfrak{F}$ is uniformly bounded and equi-Lipschitz continuous with respect to ρ on Ω.

Proof For any $z \in Z_{\mathrm{ad}}$ and $\omega' \in \Omega$ let $F(z, \omega) = A(\omega)^{-1}(z + g(\omega)) - \tilde{u}_d$. Our assumptions imply that $F(\cdot, \cdot)$ is $\mathbb{P}$-a.s. uniformly bounded in H by some constant $\hat{C}$ (see the proof of Lemma 1). Furthermore, we obtain for any $z \in Z_{\mathrm{ad}}$ and $\omega, \omega' \in \Omega$:

$$
\begin{aligned}
|f(z, \omega) - f(z, \omega')| &= \frac{1}{2}(\|F(z, \omega)\|_H + \|F(z, \omega')\|_H)\|F(z, \omega)\|_H - \|F(z, \omega')\|_H| \\
&\leq \hat{C}\|F(z, \omega) - F(z, \omega')\|_H \\
&\leq \hat{C}(\|(A(\omega)^{-1} - A(\omega')^{-1})z\|_V + \|A(\omega)^{-1}g(\omega) - A(\omega')^{-1}g(\omega')\|_V),
\end{aligned}
$$

where we used the uniform boundedness and the Poincaré-Friedrichs' inequality. For the first term on the right-hand side we argue as in Lemma 2. The second term is estimated by

$$\|A(\omega)^{-1}g(\omega) - A(\omega')^{-1}g(\omega')\|_V \leq \|(A(\omega)^{-1} - A(\omega')^{-1})g(\omega)\|_V$$
$$+ \|A(\omega')^{-1}(g(\omega) - g(\omega'))\|_V.$$

Now, we use again Lemma 2 for the first term and both Lemma 1 and the assumption on g for the second. Combining these observations, we obtain the assertion. □

Theorem 5 establishes the $\mathbb{P}$-uniformity of $\mathfrak{F}$ under relatively mild assumptions. In particular, we do not require the terms b_{ij} and g to be smooth in any way with respect to ω. Of course in many interesting applications, see e.g. [4,5], one can demonstrate much higher regularity of $F(z, \omega) = A(\omega)^{-1}(z + g(\omega)) - \tilde{u}_d$ in ω for each $z \in Z_{ad}$ if some smoothness of b_{ij} and g is in fact available. And though the presence of $\|\cdot\|_H^2$ in $f(z, \omega)$ rules out 1-Lipschitz continuity, as required by estimates using the Wasserstein distances, the quantitative estimates using the Fortet-Mourier metric, e.g., ζ_2, which are incidentally strictly sharper than the Wasserstein estimates, are still applicable provided the local growth conditions for functions in $\mathcal{F}_p(\Omega)$ are fulfilled. On the other hand, as a general point of critique, the minimal information metric along with Theorems 4 and 5 preclude the need to enlarge the set of integrands $\mathfrak{F}$ in order make use of the rougher estimates given by the Fortet-Mourier estimates.

This brings us to our second goal of this section. In order to solve optimization problems of the type

$$\min \ \mathbb{E}_{\mathbb{P}}[f(z)] + \frac{\alpha}{2}\|z\|_H^2 \text{ over } z \in Z_{ad} \tag{18}$$

with $f \in \mathfrak{F}$ numerically, not only $\mathbb{P}$ but the decision variables z and the underlying partial differential equation must be approximated. In order words, using a finite-sample-based approximation $\mathbb{P}_N$ of $\mathbb{P}$ and a finite-element discretization for the deterministic quantities in H and V, we would typically consider the finite-dimensional problems of the type:

$$\min \ \frac{1}{2}\sum_{i=1}^{N} \pi_i \left[\|(A_i^h)^{-1}(z^h) - u_{d,i}^h\|_H^2 \right] + \frac{\alpha}{2}\|z^h\|_H^2 \text{ over } z^h \in Z_{ad}^h. \tag{19}$$

Here, A_i^h is defined from $A(\omega)$ by replacing ω with a realization ω_i and V by a finite-dimensional subspace V_h derived by a standard finite-element approximation. The term $u_{d,i}^h = (A_i^h)^{-1}(g_i^h - \tilde{u}_{d,h})$ and Z_{ad}^h is an approximate of Z_{ad} using a finite-element approximation of H.

For piecewise constant approximations of the control z, the original ideas date back to Falk [10], see the recent chapter [2] for a quick reference. For another more recent, comprehensive treatment see [26] (for a posteriori error estimates) as well as the monograph [15, Chap. 3] and the many references therein. In many of these works, as well as in our concrete example in the next section, the set Z_{ad} has the concrete form:

$$Z_{ad} := \left\{ z \in L^2(D) \,\middle|\, \underline{a}(x) \leq z(x) \leq \overline{a}(x) \text{ a.e. } x \in D \right\},$$

where $\underline{a}, \overline{a}$ are sufficiently regular; often $L^\infty(D)$. Therefore, for the sake of argument, we may assume that if D, B, g, $\underline{a}$, $\overline{a}$, and $\widetilde{u}_d$ are sufficiently regular, then there exists a (random variable) $C_N \geq 0$ along with a real $q \in (0, 3/2]$ such that

$$\|z_{\mathbb{P}_N}^h - z_{\mathbb{P}_N}\|_H \leq C_N h^q \quad \omega \in \Omega. \tag{20}$$

Here, the dependence on N results from the fact that the typical estimates, see e.g., [2, Thm. 10], depend on constants related to the coefficients of the PDE, the right-hand side g, and u_d, which is stochastic. Therefore, in the estimate (20), C_N is related to a realization of a random sample of length N. Using (20), we can apply Theorem 4 and the triangle inequality to obtain the estimate

$$\|z_{\mathbb{P}} - z_{\mathbb{P}_N}^h\|_H \leq 2\sqrt{2}\alpha^{-1/2}d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)^{1/2} + C_N h^q. \tag{21}$$

It should also be noted that for piecewise constant approximations of Z_{ad}^h, we can only expected $q = 1$. The case for $q = 3/2$ requires a significant amount of regularity, and $q \in (1, 3/2)$ depends on both the regularity as well as the type of discretization. In light of Theorem 5, (21) guarantees the convergence of the fully discrete solutions $z_{\mathbb{P}_N}^h$ to the original infinite dimensional solution, provided $h \downarrow 0$, and $\mathbb{P}_N \to \mathbb{P}$ weakly.

7 Monte Carlo approximation and numerical illustration for PDE-constrained optimization

In this final section, we provide additional discussions on the behavior of $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)$ for Monte Carlo approximations $\mathbb{P}_N$ of $\mathbb{P}$ in order to give the reader a better impression of the potential rate of convergence; in particular, for the setting in the previous section.

7.1 Monte Carlo approximation

For the sake of argument, we assume that the integrands f can be written

$$f(z, \omega) = f(z, \boldsymbol{\xi}(\omega)) + \frac{\alpha}{2}\|z\|_H^2,$$

where $\boldsymbol{\xi} : \Omega \to \varXi \subset \mathbb{R}^d$ is a random vector. This is often the case for PDE-models and is used in the example in Sect. 7.2 below.

Let $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \ldots, \boldsymbol{\xi}^N, \ldots$ be independent identically distributed $\varXi$-valued random vectors on some probability space $(\Omega, \mathcal{F}, P)$ having the common probability law $\mathbb{P}$, i.e., $\mathbb{P} = P \circ (\boldsymbol{\xi}^1)^{-1}$. In this context, we define the empirical measures

$$\mathbb{P}_N(\cdot) = \frac{1}{N}\sum_{i=1}^N \delta_{\boldsymbol{\xi}^i(\cdot)} \quad (n \in \mathbb{N}) \tag{22}$$

and the empirical or Monte Carlo approximation of the stochastic program (18) with sample size N, i.e.,

$$\min\left\{\frac{1}{N}\sum_{i=1}^{N} f(z, \boldsymbol{\xi}^i(\cdot)) : z \in Z_{\text{ad}}\right\}. \tag{23}$$

The optimal value $v(\mathbb{P}_N(\cdot))$ of (23) is a real random variable and the solution $z_{\mathbb{P}_N(\cdot)}$ an H-valued random element. It is well known that the sequence $\{\mathbb{P}_N(\cdot)\}$ of empirical measures converges weakly to $\mathbb{P}$ P-almost surely (see, e.g., [8, Theorem 11.4.1]). Theorem 5 implies that the class $\mathfrak{F}$ is a P-uniformity class and, hence, the sequence $d_{\mathfrak{F}}(\mathbb{P}_N(\cdot), \mathbb{P})$ converges to zero P-almost surely. According to Theorem 4, the sequences $\{v(\mathbb{P}_N(\cdot))\}$ and $\{z_{\mathbb{P}_N(\cdot)}\}$ of empirical optimal values and solutions converge P-almost surely to their true optimal values and solutions $v(\mathbb{P})$ and $z_{\mathbb{P}}$, respectively.

In order to obtain rates of convergence for the sequences of empirical optimal values and solutions, we consider their mean or mean square distance and conclude from Theorem 4 and Corollary 1 the estimates

$$\mathbb{E}\left[|v(\mathbb{P}_N) - v(\mathbb{P})|\right] \le \mathbb{E}\left[d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P})\right] \le L\,\mathbb{E}\left[\zeta_1(\mathbb{P}_N, \mathbb{P})\right]$$

$$\left(\mathbb{E}\left[\|z(\mathbb{P}_N) - z(\mathbb{P})\|_H^2\right]\right)^{\frac{1}{2}} \le \left(\mathbb{E}\left[d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P})\right]\right)^{\frac{1}{2}} \le \left(\hat{L}\,\mathbb{E}\left[\zeta_1(\mathbb{P}_N, \mathbb{P})\right]\right)^{\frac{1}{2}},$$

where L is the constant appearing in Corollary 1 and $\hat{L}$ is given by $\hat{L} = 2L\sqrt{\frac{2}{\alpha}}$. Unfortunately, convergence rates of the mean convergence of $d_{\mathfrak{F}}(\mathbb{P}_N(\cdot), \mathbb{P})$ are not known, but for $\mathfrak{F}$ containing uniformly bounded and equi-Lipschitz continuous functions, the rate coincides essentially with that of $\mathbb{E}[\zeta_1(\mathbb{P}_N, \mathbb{P})]$. For the latter it follows from [6, Theorem 1] that the estimate

$$\mathbb{E}\left[\zeta_1(\mathbb{P}_N, \mathbb{P})\right] \le \kappa_s M_s(\mathbb{P}) N^{-\frac{1}{d}} \tag{24}$$

is valid if $d \ge 3$, $s > \frac{d}{d-1}$, κ_s is a constant only depending on s and the sth absolute moment

$$M_s(\mathbb{P}) = \left(\int_{\mathbb{R}^d} \|\xi\|^s d\xi\right)^{\frac{1}{s}}$$

is finite. It is argued in [11] that the rate (24) is sharp if $\mathbb{P}$ is the uniform distribution on the unit cube $[-1, 1]^d$. For the numerical experiments in the next subsection, we observe better rates than guaranteed by the probability metrics. This is a limitation of the probability metrics themselves, not their application to the problem at hand.

7.2 Numerical illustration

The previous discussion provides useful upper bounds for the case of the empirical measure $\mathbb{P}_N$. However, these bounds are based on estimates of the Wasserstein metrics,

not the minimal information metric. Therefore, in order to obtain a better impression of the possible rate of convergence associated with $d_{\mathfrak{F}}(\mathbb{P}, \mathbb{P}_N)$ in practice, we provide here a concrete numerical example.

We propose a model problem, which is based on [16, Ex. 6.1, Ex. 6.2]. Let $\alpha = 10^{-3}$, $D = (0, 1)$, $\widetilde{u}(x) = \sin(50.0 * x/\pi)$, and consider the optimal control problem

$$\text{minimize}_{z \in L^2(D)} \frac{1}{2} \mathbb{E}_{\mathbb{P}} \left[\|u - \widetilde{u}_d\|_H^2 \right] + \frac{\alpha}{2} \|z\|_H^2 \text{ over } z \in L^2(D) \quad (25)$$

where $z \in Z_{\text{ad}}$ with

$$Z_{\text{ad}} := \left\{ w \in L^2(D) \,|\, -0.75 \leq w(x) \leq 0.75 \text{ a.e. } x \in D \right\}$$

and $u = u(z) \in L^\infty(\Omega, \mathcal{F}, \mathbb{P}; H^1(D))$ solves the weak form of

$$-\nu(\omega)\partial_{xx} u(\omega, x) = g(\omega, x) + z(x) \qquad (\omega, x) \in \Omega \times D, \quad (26a)$$
$$u(\omega, 0) = d_0(\omega), \quad u(\omega, 1) = d_1(\omega) \qquad\qquad \omega \in \Omega. \quad (26b)$$

Furthermore, we suppose that

$$\nu(\omega) := 10^{\xi_1(\omega) - 2}, \qquad g(\omega, x) := \frac{\xi_2(\omega)}{100}$$
$$d_0(\omega) := 1 + \frac{\xi_3(\omega)}{1000} \qquad d_1(\omega) := \frac{\xi_4(\omega)}{1000},$$

with random variables $\xi_i : \Omega \to \mathbb{R}, i = 1, 2, 3, 4$, such that the supports $\xi_i, i = 1, 2, 3$, are $[-1, 1]$ and the support of ξ_4 is $[1, 3]$. For the sake of illustration, we assume that each of these random variables is uniformly distributed and after the usual change of variables, we consider instead

$$-\nu(\xi)\partial_{xx} u(\xi, x) = g(\xi, x) + z(x) \qquad (\xi, x) \in \Xi \times D, \quad (27a)$$
$$u(\xi, 0) = d_0(\xi), \quad u(\xi, 1) = d_1(\xi) \qquad\qquad \xi \in \Xi. \quad (27b)$$

with $\Xi = [-1, 1] \times [-1, 1] \times [-1, 1] \times [1, 3]$, endowed with the associated uniform density. We define $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_4) \in \Xi$.

Finally, we note that the linearity of the differential equation allows us to use the superposition principle to separate the unique, z-dependent solution $u(z)$ into the sum of random fields as $u(z) = S(z) + \widehat{u}$, where $S(z)$ maps z from H into the solution space and $\widehat{u}$ is a fixed random field. With the aim of showing that the necessary continuity properties used in the previous section are fulfilled, let $\widehat{u} : D \times \Xi \to \mathbb{R}$ denote a function such that $\widehat{u}(\cdot, \xi)$ belongs to $H^1(D)$ and $\widehat{u}$ satisfies the inhomogeneous random boundary condition in (27b). If we then recast (27) as the random elliptic PDE with random boundary conditions

$$A(\xi)u = z + g(\xi), \quad u(x) = b(x, \xi) \quad (x \in \partial D, \xi \in \Xi) \quad (28)$$
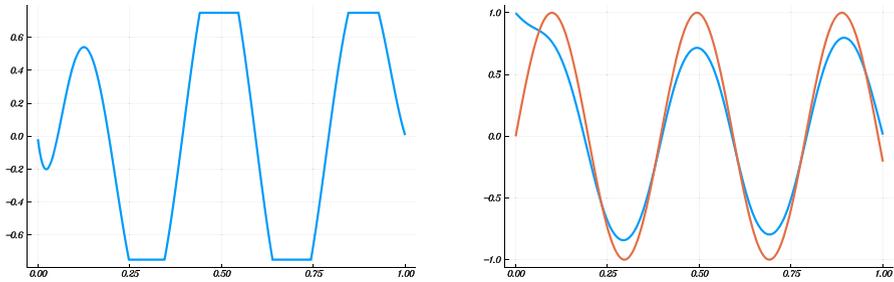
Fig. 1 (l) Optimal solution $z_{\mathbb{P}}$ of (25) for $M = 100$, $h = 1/(2^{10} - 1)$. (r) Empirical mean value of $u(z_{\mathbb{P}})$ out of sample using 1000 iid random realizations of $(\xi_1, \ldots, \xi_4)$

and determine $u \in H_0^1(D)$ such that

$$A(\xi)u = z + g(\xi) + A(\xi)\widehat{u}(\xi) \quad (\xi \in \varXi).$$

Then $u + \widehat{u}$ solves the random boundary value problem (28) of the random elliptic PDE. Based on the model assumptions, we see that both A and the altered right-hand side $g + A\widehat{u}$ satisfy the necessary conditions for our theory.

In order to illustrate the sensitivities for this model problem, we generate a sample on $\varXi$ of size M and replace $\mathbb{P}$ by the corresponding empirical measure $\mathbb{P}_M$. To avoid confusion, we leave off the M subscript in the following discussion. The underlying spaces for the control, state, and adjoint variables are discretized using standard piecewise linear finite elements on a uniform mesh with parameter $h = 1/(2^8 - 1)$.

The resulting finite-dimensional deterministic optimization problems are solved by a semismooth Newton method as proposed in [14,24,25], which in the current context is mesh-independent. The forward and adjoint equations as well as the linear equations for the Hessian-vector products are solved directly and the reduced system for the semismooth Newton step is calculated using conjugate gradients as some of the associated operators are only implicitly given.

The algorithm terminates once the discrete ℓ^2-norm of the residual of the optimality system reaches a tolerance of 1e-8. On average, the semismooth Newton algorithm required between three to four iterations with approximately 13 to 15 inner iterations for the CG solver. The solution $z_{\mathbb{P}}$ of this problem is treated as the "true" solution. See Fig. 1 for the solution $z_{\mathbb{P}}$ with $M = 100$ and its behavior out of sample on the mean value of the state u. As expected, the control behaves well on average.

We now investigate $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P})$, where $\mathbb{P}_N$ is an empirical probability measure based on a random sample of size N ($1 \leq N \leq M$) of the M realizations $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^M$. Since a direct calculation of $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P})$ for even simple examples can be quite challenging, we calculate instead $v(\mathbb{P}_N)$ and $z_{\mathbb{P}_N}$ and compute the associated errors

$$|v(\mathbb{P}_N) - v(\mathbb{P})| \text{ and } \|z_{\mathbb{P}_N} - z_{\mathbb{P}}\|_H.$$

The experiment is repeated for each $N = 1, \ldots, M$ 100 times. The plots of these errors can be see in Fig. 2.
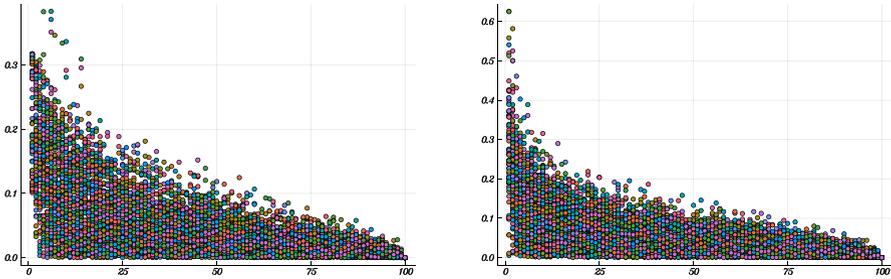
Fig. 2 Scatter plots of the computed errors (l) $|\nu(\mathbb{P}_N) - \nu(\mathbb{P})|$ (r) $\|z_{\mathbb{P}_N} - z_{\mathbb{P}}\|_H$ for $M = 100$, $N \in \{1, \ldots, 100\}$ repeated 100 times

The choice of setting $M = 100$ was made for computationally expediency. Though the semismooth Newton method implemented for this example works exceptionally well, it still requires hundreds if not thousands of PDE solves for each N. Nevertheless, due to the relatively small choice of M we cannot use all the data points in Fig. 2 in order to estimate a rate of convergence. As a compromise, we only use the data points for $N = 1, \ldots, 25$. This provides us with the estimates

$$\|z_{\mathbb{P}_N} - z_{\mathbb{P}}\|_H = O(N^{-0.4995}) \ \text{ and } \ |\nu(\mathbb{P}_N) - \nu(\mathbb{P})| = O\left(N^{-0.5591}\right)$$

The estimates were generated by solving a linear least-squares regression problem with ridge (ℓ^2) regularization term. Whereas the convergence of the optimal values would indicate a rate of roughly $1/2$, the rate of convergence for the optimal solutions would be a marked improvement over the theoretical estimates based on probability metrics. One explanation for this would be that the integrands for this specific model problem have much better smoothness properties than assumed. We leave a rigorous analysis of this question for future research.

8 Conclusion

We have shown that a number known results for stability of stochastic programs with finite-dimensional decision spaces can be carried over to infinite dimensions, provided certain convexity conditions are satisfied. As perhaps expected the best possible growth rates for the convergence of solutions using probability metrics are of Hölder-type. Our analysis gives rise to a number of possible future directions and open questions. For example, we consider a setting that is primarily related to risk-neutral problems, whereas problems using typically non-smooth risk measures in the objective are of significant interest for robust engineering design. In addition, the assumptions on the objective and linear operator Σ are essential for the qualitative stability analysis as the infinite-dimensional setting often requires us to make use of the weak topology. Without such regularity properties, it is unclear how to proceed in general. Finally, in order to develop adaptive numerical optimization methods based on estimates of the type (21), we need to more closely investigate the convergence of $d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P})$ for the

class of functions $\mathfrak{F}$ used in Sect. 6. and specific approximations $\mathbb{P}_N$. Section 7 provides some deeper insight into this question, however the picture is far from complete.

A Results from fixed point theory

The following is a result can be found in the monograph [12].
    We provide a translation here along with a short proof for the reader's convenience

Proposition 2 (Lemma 3.1 in *[12]*) *Let V be a Hilbert space with inner product $(\cdot, \cdot)$, dual $V^\star$ and dual pairing $\langle \cdot, \cdot \rangle$, $b \in V^\star$, $A : V \to V^\star$ a strongly monotone (with constant $\gamma > 0$) and Lipschitz continuous (with modulus $L > 0$) operator, and $J : V \to V^\star$ the duality mapping, i.e., $\langle Ju, v \rangle = (u, v), \forall u, v \in V$.*
    Then the mapping $K_t : V \to V$ given by

$$K_t x = x - t J^{-1}(Ax - b)$$

is a contraction with constant $0 < \kappa(t) < 1$, where

$$\kappa(t) = \sqrt{1 - 2\gamma t + L^2 t^2}$$

and $t \in (0, \frac{2\gamma}{L^2})$. Moreover, the unique fixed point of K_t is the unique solution of $Ax = b$ and belongs to the ball around zero with radius $r = (1 - \kappa(t))^{-1} \|K_t 0 - 0\| = t(1 - \kappa(t))^{-1} \|A0 - b\|_\star$.

Remark 2 Note that $\min \kappa(t) = L^{-1} \sqrt{L^2 - \gamma^2}$ and, hence, $\kappa(t)$ is typically close to 1.

Proof Let $x, x' \in H$. Then

$$\|K_t x - K_t x'\|^2 = \|x - x'\|^2 - 2t \left(J^{-1}\left(Ax - Ax'\right), x - x' \right) + t^2 \|J^{-1}(Ax - Ax')\|^2$$

$$= \|x - x'\|^2 - 2t \langle Ax - Ax', x - x' \rangle + t^2 \|Ax - Ax'\|_\star^2$$

$$\leq \left(1 - 2t\gamma + t^2 L^2\right) \|x - x'\|^2 = \kappa^2(t)\|x - x'\|^2.$$

Clearly, $0 < \kappa(t) < 1$ iff $t \in (0, \frac{2\gamma}{L^2})$. Furthermore, the unique solution $\bar{x}(b)$ of $Ax = b$ satisfies

$$\|\bar{x}(b)\| = \|K_t\bar{x}(b)\| \leq \|K_t\bar{x}(b) - K_t0\| + \|K_t0 - 0\| \leq \kappa(t)\|\bar{x}(b)\| + \|K_t0 - 0\|,$$

from which immediately obtain the estimate

$$\|\bar{x}(b)\| \leq \frac{\|K_t0 - 0\|}{1 - \kappa(t)} = \frac{t}{1 - \kappa(t)}\|A0 - b\|_\star.$$

This finishes the proof. $\qquad\square$

The following result can be found, e.g., in [7].

Proposition 3 (Theorem 1A.4 in *[7]*) *Let P be a metric space with metric ρ and X a complete metric space with metric d. Let $F : P \times X \rightarrow X$ and assume that there exist $\alpha \in (0, 1)$ and $\lambda > 0$ such that*

$$d\left(F(p, x), F(p, x')\right) \leq \alpha d(x, x') \quad \left(\forall x, x' \in X, p \in P\right)$$
$$d\left(F(p, x), F(p', x)\right) \leq \lambda\rho(p, p') \quad \left(\forall p, p' \in P, x \in X\right).$$

Then, for each $p \in P$, there exists a unique fixed point $x(p)$ of $F(p, \cdot)$ in X and we have the estimate

$$d(x(p), x(p')) \leq \frac{\lambda}{1 - \alpha}\rho(p, p') \quad \left(\forall p, p' \in P\right).$$

References

1. Alt, H.W.: Linear functional analysis. Universitext: an application-oriented introduction. Springer, London Ltd, London. Translated from the German edition by Robert Nürnberg (2016). https://doi.org/10.1007/978-1-4471-7280-2
2. Antil, H., Leykekhman, D.: A brief introduction to PDE-constrained optimization. Frontiers in PDE-constrained optimization. IMA Vol. Math. Appl., vol. 163, pp. 3–40. Springer, New York (2018)
3. Attouch, H., Buttazzo, G., Michaille, G.: Variational analysis in Sobolev and BV spaces. MPS/SIAM Series on Optimization, vol. 6. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2006)
4. Bachmayr, M., Cohen, A., Migliorati, G.: Sparse polynomial approximation of parametric elliptic PDES. PART 1: Affine coefficients. ESAIM: M2AN **51**(1), 321–339 (2017). https://doi.org/10.1051/m2an/2016045
5. Cohen, A., Devore, R., Schwab, C.: Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. Anal. Appl. (Singapore) **9**(1), 11–47 (2011). https://doi.org/10.1142/S0219530511001728
6. Dereich, S., Scheutzow, M., Schottstedt, R.: Constructive quantization: approximation by empirical measures. Ann. Inst. Henri Poincaré Probab. Stat. **49**(4), 1183–1203 (2013). https://doi.org/10.1214/12-AIHP489

7. Dontchev, A.L., Rockafellar, R.T.: Implicit functions and solution mappings: a view from variational analysis. Springer Series in Operations Research and Financial Engineering, 2nd edn. Springer, New York (2014)

8. Dudley, R.M.: Real analysis and probability. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA (1989)

9. Dunford, N., Schwartz, J.T.: Linear operators. Part I. Wiley Classics Library. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. Wiley, New York (1988)

10. Falk, R.S.: Approximation of a class of optimal control problems with order of convergence estimates. J. Math. Anal. Appl. **44**, 28–47 (1973). https://doi.org/10.1016/0022-247X(73)90022-X

11. Fournier, N., Guillin, A.: On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Related Fields **162**(3–4), 707–738 (2015). https://doi.org/10.1007/s00440-014-0583-7

12. Gajewski, H., Gröger, K., Zacharias, K.: Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen. Mathematische Lehrbücher und Monographien. II, Abteilung, Mathematische Monographien, Band 38. Akademie-Verlag, Berlin (1974)

13. Hille, E., Phillips, R.S.: Functional analysis and semi-groups. vol. 31, Rev. ed. American Mathematical Society Colloquium Publications, American Mathematical Society, Providence (1957)

14. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**(3), 865–888 (2002)

15. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints. Mathematical Modelling: Theory and Applications, vol. 23. Springer, New York (2009)

16. Kouri, D.P., Surowiec, T.M.: Risk-averse PDE-constrained optimization using the conditional value-at-risk. SIAM J. Optim. **26**(1), 365–396 (2016). https://doi.org/10.1137/140954556

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

18. Müller, A.: Integral probability metrics and their generating classes of functions. Adv. Appl. Probab. **29**(2), 429–443 (1997)

19. Rachev, S.T.: Probability metrics and the stability of stochastic models. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley, Chichester (1991)

20. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. Math. Oper. Res. **27**(4), 792–818 (2002). https://doi.org/10.1287/moor.27.4.792.304

21. Ramsay, J.O., Silverman, B.W.: Functional data analysis. Springer Series in Statistics, 2nd edn. Springer, New York (2005)

22. Römisch, W.: Stability of stochastic programming problems. In: Stochastic programming, *Handbooks Oper. Res. Management Sci.*, vol. 10, pp. 483–554. Elsevier, Amsterdam (2003). https://doi.org/10.1016/S0927-0507(03)10008-4

23. Topsøe, F.: On the connection between P-continuity and P-uniformity in weak convergence. Probab. Theory Appl. **12**, 281–290 (1967)

24. Ulbrich, M.: Semismooth Newton methods for operator equations in function spaces. SIAM J. Optim. **13**(3), 805–841 (2002)

25. Ulbrich, M.: Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces. *MOS-SIAM Series on Optimization*, vol. 11. SIAM, MOS, Philadelphia, PA (2011). https://doi.org/10.1137/1.9781611970692

26. Vexler, B., Wollner, W.: Adaptive finite elements for elliptic optimization problems with control constraints. SIAM J. Control Optim. **47**(1), 509–534 (2008). https://doi.org/10.1137/070683416

27. Zolotarev, V.M.: Probability metrics. Theory Probab. Theory Appl. **28**(2), 278–302 (1983)