

Analysis of Discrete L^2 Projection on Polynomial Spaces with Random Evaluations

G. Migliorati · F. Nobile · E. von Schwerin ·
R. Tempone

Received: 12 December 2011 / Revised: 28 June 2013 / Accepted: 20 November 2013 /
Published online: 5 March 2014
© SFoCM 2014

Abstract We analyze the problem of approximating a multivariate function by discrete least-squares projection on a polynomial space starting from *random, noise-free* observations. An area of possible application of such technique is uncertainty quantification for computational models. We prove an optimal convergence estimate, up to a logarithmic factor, in the univariate case, when the observation points are sampled in a bounded domain from a probability density function bounded away from zero and bounded from above, provided the number of samples scales quadratically with the dimension of the polynomial space. Optimality is meant in the sense that the weighted L^2 norm of the error committed by the random discrete projection is bounded with high probability from above by the best L^∞ error achievable in the given polynomial space, up to logarithmic factors. Several numerical tests are presented in both the univariate and multivariate cases, confirming our theoretical estimates. The numerical tests also clarify how the convergence rate depends on the number of sampling points, on the polynomial degree, and on the smoothness of the target function.

Keywords Approximation theory · Error analysis · Multivariate polynomial approximation · Nonparametric regression · Noise-free data · Generalized polynomial chaos · Point collocation

Communicated by Albert Cohen.

G. Migliorati (✉) · F. Nobile
CSQI-MATHICSE, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
e-mail: giovanni.migliorati@gmail.com

E. von Schwerin · R. Tempone
Applied Mathematics and Computational Sciences, and SRI Center for Uncertainty Quantification in
Computational Science and Engineering, KAUST, Thuwal 23955-6900, Saudi Arabia

Mathematical Subject Classification 41A10 · 41A25 · 65N12 · 65N15 · 65N35

1 Introduction

Given a smooth multivariate function $\phi = \phi(Y^1, \dots, Y^d)$ depending on d random variables Y^1, \dots, Y^d , we consider in this work the classical problem of approximating ϕ in a multivariate polynomial space, starting from *noise-free* observations of ϕ on random evaluations of Y^1, \dots, Y^d . The motivation for such work comes from the field of uncertainty quantification (UQ) in computational models [29,35], where uncertainty is often present in many input parameters entering the mathematical model used to describe some problem in, for example, engineering, physics, or biology. The uncertainty can be characterized in probabilistic terms by considering the input parameters as random variables. The goal of the analysis is typically to compute the statistics of the solution to the mathematical model or some output quantities of interest. The methodology proposed in this work could also be used, more generally, as a tool to approximate response functions in engineering applications for the purpose of design optimization or system control.

It is assumed here that, for each value of the input parameters, the solution or output quantity can be accessed without errors. This is of course an idealization as deterministic approximation-type errors will typically be present whenever the model involves differential or integral operators. Also, round-off errors will be present. However, these sources of error are quite different in nature from *measurement errors* appearing in an experimental setting, which are usually modeled as random and statistically independent. In the context of UQ in computational models, it is therefore reasonable to assume that the approximation errors can be kept under control by some careful a posteriori error estimation and mesh refinement (see, e.g., [1,5] and references therein).

A technique that has received considerable attention in the last few years is the so-called generalized polynomial chaos (gPC) expansion; see, e.g., [23,36]. It consists in expanding the solution in polynomials of the input random variables. The use of global polynomial spaces is sound in many situations, where the input/output (parameters-to-solution) map is smooth. This is the case, for instance, in elliptic partial differential equations with random diffusion coefficient [2,6,14,15].

Once a truncated gPC expansion has been computed by some means, it can be used later for inexpensive computations of solution statistics or as a reduced model of the input/output map for global sensitivity analysis [16,34] or optimization under uncertainty [21].

As a tool to build such a gPC approximation, we consider in this work an L^2 projection, starting from a random sample of the input parameters. Such an idea has already been proposed in the framework of UQ and been given several names: point collocation [18,26,32], nonintrusive gPC [20,29], and regression [9,10,30,31]. As a practical recipe, the size of the sample drawn from the input distribution is typically taken to be two to three times the dimension of the polynomial space.

The proposed approach falls within the classical topic of polynomial regression estimation, i.e., minimization of the empirical L^2 risk within the given polynomial

space. We insist, however, that, unlike in the traditional statistical approach, here we consider noise-free data evaluated in random points.

A relevant question is whether such a minimizer is optimal in the sense that it achieves an approximation of the (unknown) function that is equivalent to the best approximation in the polynomial space.

Many works can be found in the literature on regression estimations in the noisy case. We recall here the book [25], which provides a general framework for analysis of regression estimators with random design, as well as [7, 8], which show the optimality of noisy regression when using piecewise constant or linear functions. The aforementioned works give estimates on the expected L^2 error under the assumption that the function is bounded in L^∞ by some a priori fixed constant. Other works in the field of distribution-free regression with noise have derived convergence rates for the L^2 risk that are optimal up to logarithmic factors, e.g., [4, 27, 28].

The L^2 error in expectation is bounded by two terms: the (purely deterministic) best approximation error of the (unknown) function in the approximating space, and the statistical error due to the random sampling and the noise in the observations. The latter scales as $1/\sqrt{m}$ if m is the number of samples. In the aforementioned works, the statistical error does not vanish in a noise-free setting.

Hence, the main question that we address in this work is this: in the noise-free polynomial approximation based on random samples, does the randomness of the evaluation points introduce a statistical error $\mathcal{O}(1/\sqrt{m})$, or is it reasonable to expect better convergence rates with respect to m ?

We study the problem theoretically for a univariate function $\phi(Y)$, where Y is a bounded random variable with the probability density function bounded away from zero and bounded from above. Denoting by n the dimension of the polynomial space, we prove that the L^2 projection on a polynomial space with random evaluations leads to quasi-optimal convergence rates (up to multiplicative logarithmic factors), provided that the sample size scales as $m \propto n^2$. Here, optimality is meant in the sense that the L^2 error behaves like the best approximation error *measured in the “sup” norm* of the target function in the chosen polynomial space, up to logarithmic factors. It is reasonable to compare the L^2 error with the L^∞ best approximation error since the random discrete L^2 projection is based on pointwise evaluations. In [13] a bound in terms of the best approximation in L^2 norm is obtained by taking expectations; the proof proposed therein uses different techniques than ours. In contrast, the bound in the present work is in terms of the L^∞ norm and holds with high probability.

We also show, in the general multivariate setting, the relation between the optimality of the projection based on random points and the condition number of the corresponding design matrix, when using an orthonormal polynomial basis.

We present several numerical tests, both on univariate and multivariate functions, that clearly show that a choice $m \propto n^2$ leads to a stable regression problem and an optimal approximation, whereas $m \propto n$ leads to an ill-conditioned problem when n is large and, eventually, to a divergent approximation. Moreover, our numerical tests show some significant differences between the one-dimensional and the multidimensional cases.

Our result is based on showing an equivalence between the L^2 continuous norm and a discrete counterpart, with high probability. This has similarities with compressed

sensing [11, 19] and, in particular, the so-called restricted isometry property. In that context the relation between the number of measurements and the sparsity of the signal to recover has been widely investigated, and results are available showing that in many cases a linear relation, up to logarithmic factors, suffices for a good recovery. On the other hand, several examples in the present paper show that this is not the case in the present setting, and a quadratic proportionality $m \propto n^2$ is needed to achieve a stable and accurate approximation.

The outline of the paper is as follows: Sect. 2 introduces the approximation problem as an L^2 projection on a space of polynomials in d underlying variables; some common choices of polynomial spaces are described in Sect. 2.1. The optimality of the random L^2 projection, in terms of a best approximation constant, is shown in Sect. 2.2; the asymptotic behavior of this best approximation constant, as the number of random evaluation points goes to infinity, is analyzed in Sect. 2.3. Next, Sect. 3 restricts the study to polynomial spaces in one variable; in this case, a bound on the best approximation constant is derived and used to prove Theorem 3, which provides a rule to select the number of random points as a function of the maximal polynomial degree, which makes the discrete random L^2 projection nearly optimal (up to a logarithmic factor) with any prescribed confidence level. Section 4 gives the algebraic formulation of the random projection problem, in view of its numerical discretization. In particular, Sect. 4.1 provides an analysis of how the condition number of the design matrix depends on the sample size and dimension of the polynomial space of the random L^2 projector. Finally, Sect. 5 complements the analysis in Sects. 2–4 with numerical tests, both in the one-dimensional case and in higher dimensions.

2 Discrete L^2 Projection with Random Points

Let $Y = [Y^1, \dots, Y^d]$ be a random vector taking values in a bounded set $\Gamma \subset \mathbb{R}^d$. We assume that Γ has a tensor structure $\Gamma = \Gamma_1 \times \dots \times \Gamma_d$, with Γ_i being closed intervals in \mathbb{R} , and that the random vector Y has a joint probability density $\rho : \Gamma \rightarrow \mathbb{R}^+$. We consider the random variable $\phi(Y)$, where $\phi : \Gamma \rightarrow \mathbb{R}$ is assumed to be a smooth function (at least continuous). The goal of the analysis is to compute statistical moments of $\phi(Y)$. This will be achieved by first constructing a reduced model (approximate response function), i.e., we approximate the function $\phi(Y)$ by a suitable multivariate polynomial $\phi_\Lambda(Y)$. We then compute the statistical moments of $\phi(Y)$ using the approximate function $\phi_\Lambda(Y)$. We denote by

$$\mathbb{E}[\phi] = \int_{\Gamma} \phi(y) \rho(y) dy$$

the expected value of the random variable $\phi(Y)$ and by

$$Pr(A) = \int_A \rho(y)dy$$

the probability of the event $A \in \mathcal{B}(\Gamma)$, where $\mathcal{B}(\Gamma)$ is the Borel σ -algebra with respect to the measure $\rho(y)dy$. Throughout the paper we also make the following assumption.

Assumption 1 (*Quasi-uniform distribution*) There exist constants $0 < \rho_{\min} \leq \rho_{\max} < +\infty$ such that $\rho_{\min} \leq \rho(y) \leq \rho_{\max}$ for all $y \in \Gamma$.

We introduce the space L^2_ρ of square-integrable functions $f : \Gamma \rightarrow \mathbb{R}$, endowed with the norm

$$\|f\|_{L^2_\rho} := \left(\int_\Gamma f^2(y)\rho(y)dy \right)^{1/2}.$$

Observe that under Assumption 1, the norm $\|\cdot\|_{L^2_\rho}$ is equivalent to the standard L^2 norm (with Lebesgue measure) since

$$\sqrt{\rho_{\min}} \leq \frac{\|f\|_{L^2_\rho}}{\|f\|_{L^2}} \leq \sqrt{\rho_{\max}}, \quad \forall f \in L^2_\rho.$$

Let $\mathbf{v} = (v^1, \dots, v^d)$ be a multi-index and $\Lambda \subset \mathbb{N}_0^d$ a multi-index set featuring the following property.

Property 1 (*Downward closedness of Λ*) Consider two multi-indices $\mathbf{v}, \boldsymbol{\mu} \in \mathbb{N}_0^d$ such that $\mu^q \leq v^q$ for all $q = 1, \dots, d$. The multi-index set Λ is downward closed (or it is a lower set) if the following holds:

$$\mathbf{v} \in \Lambda \implies \boldsymbol{\mu} \in \Lambda.$$

We denote by $\mathbb{P}_\Lambda = \mathbb{P}_\Lambda(\Gamma)$ the multivariate polynomial space associated with the index set Λ

$$\mathbb{P}_\Lambda(\Gamma) := \text{span} \left\{ \prod_{q=1}^d (Y^q)^{v^q} : \mathbf{v} \in \Lambda \right\} \tag{1}$$

and by $n = \dim(\mathbb{P}_\Lambda)$ its dimension. For convenience, the set Λ can be arranged in lexicographical order:

Property 2 (Lexicographical order) Given any $\mathbf{v}, \boldsymbol{\mu} \in \Lambda$,

$$\mathbf{v} \stackrel{L}{<} \boldsymbol{\mu} \iff \exists \tilde{q} > 0 : (v^{\tilde{q}} < \mu^{\tilde{q}}) \wedge (v^q = \mu^q \quad \forall q < \tilde{q}).$$

According to this order, we can let \mathbf{v}_j denote the j th multi-index of Λ . Sometimes we refer to the elements of Λ by the generic multi-index \mathbf{v} rather than listing them by the lexicographical index.

Since the monomial basis in (1) is very ill-conditioned, in practice we use an orthonormal polynomial basis. A typical choice is to take orthogonal polynomials with respect to the measure $\rho(y)dy$. We introduce an orthonormal basis $\{\psi_j\}_{j=1}^n$ of \mathbb{P}_Λ with respect to the weighted inner product

$$(f_1, f_2)_{L^2_\rho} := \int_{\Gamma} f_1(y)f_2(y)\rho(y) dy,$$

i.e., $(\psi_i, \psi_j)_{L^2_\rho} = \delta_{ij}$. Assumption 2 ensures that the orthonormal basis is complete in L^2_ρ when $\Lambda = \mathbb{N}_0^d$, applying Theorems 3.3 and 3.5 of [22].

In the case where the density factorizes as $\rho(Y) = \prod_{q=1}^d \rho_q(Y^q)$, the basis can be constructed by tensorizing univariate orthogonal polynomials with respect to each weight ρ_q separately. Given q , let $\{\varphi_j^q\}_j$ be the orthogonal polynomials with respect to ρ_q . Picking the j th (according to Property 2) multi-index $\mathbf{v}_j \in \Lambda$, the corresponding j th multivariate basis function is given by

$$\psi_j(Y) := \prod_{q=1}^d \varphi_{\mathbf{v}_j^q}^q(Y^q). \tag{2}$$

Thus, using the basis function provided by (2), definition (1) of \mathbb{P}_Λ becomes

$$\mathbb{P}_\Lambda = \text{span}\{\psi_j, j = 1, \dots, n\}, \tag{3}$$

and it holds that $n = \dim(\mathbb{P}_\Lambda) = \#\Lambda$, where $\#\Lambda$ denotes the cardinality of the set Λ . Observe that in general (1) and (3) are equivalent only if the index set Λ satisfies the monotonicity Property 1.

In the sequel we will denote by $\mathbf{Y} = \{Y_1, \dots, Y_m\}$ a random sample composed of m independent random vectors Y_1, \dots, Y_m identically distributed according to the density ρ . The symbol \mathbf{Y} as a superscript or subscript will denote that a quantity depends on the random sample $\{Y_1, \dots, Y_m\}$ and, therefore, is random itself.

We now introduce the *random* discrete inner product

$$(f_1, f_2)_{\mathbf{Y}} := \frac{1}{m} \sum_{i=1}^m f_1(Y_i)f_2(Y_i) \tag{4}$$

and the corresponding norm $\|f\|_{\mathbf{Y}} := (f, f)_{\mathbf{Y}}^{1/2}$. Observe that they are actually an inner product and a norm in \mathbb{P}_Λ , respectively, if for all $v \in \mathbb{P}_\Lambda$ it holds that

$$v(Y_i) = 0, \text{ for } i = 1, \dots, m \iff v \equiv 0,$$

which, by Assumption 1, happens almost surely (a.s.) provided that $m \geq n$.

To construct the polynomial approximation $\Pi_\Lambda^{\mathbf{Y}}\phi$ of the function ϕ , we take a realization y_1, \dots, y_m of the random sample \mathbf{Y} , calculate the noise-free evaluations

of the function ϕ in the sampled points, and finally compute a discrete least-squares approximation of the values $\phi(y_i)$ in the polynomial space \mathbb{P}_Λ , i.e.,

$$\Pi_\Lambda^Y \phi := \operatorname{argmin}_{v \in \mathbb{P}_\Lambda} \sum_{i=1}^m (\phi(Y_i) - v(Y_i))^2 = \operatorname{argmin}_{v \in \mathbb{P}_\Lambda} \|\phi - v\|_Y. \tag{5}$$

The discrete least-squares problem can be written equivalently as

$$\text{find } \Pi_\Lambda^Y \phi \in \mathbb{P}_\Lambda \quad \text{s.t.} \quad (\Pi_\Lambda^Y \phi - \phi, v)_Y = 0, \quad \forall v \in \mathbb{P}_\Lambda.$$

2.1 Common Multivariate Polynomial Spaces

Some of the most common choices of polynomial spaces are the tensor product, total degree, and hyperbolic cross, which are defined by the index sets below. We index the set Λ by the subscript w denoting the maximum polynomial degree used:

$$\text{Tensor Product (TP)} : \quad \Lambda_w = \left\{ \mathbf{v} \in \mathbb{N}_0^d : \max_{q=1, \dots, d} v^q \leq w \right\},$$

$$\text{Total Degree (TD)} : \quad \Lambda_w = \left\{ \mathbf{v} \in \mathbb{N}_0^d : \sum_{q=1}^d v^q \leq w \right\},$$

$$\text{Hyperbolic Cross (HC)} : \quad \Lambda_w = \left\{ \mathbf{v} \in \mathbb{N}_0^d : \prod_{q=1}^d (v^q + 1) \leq w + 1 \right\}.$$

These spaces are isotropic, and the maximum polynomial degree w is the same in all spatial dimensions. Anisotropic versions could be considered as well [3]. The dimensions of the TP and TD spaces are

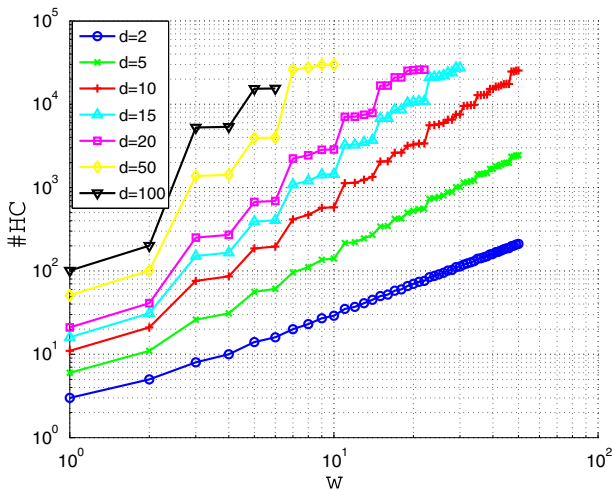


Fig. 1 Dimension of HC space, $d = 2, 5, 10, 15, 20, 50, 100$

$$\#TP(w, d) = (w + 1)^d, \tag{6}$$

$$\#TD(w, d) = \binom{d + w}{d}. \tag{7}$$

The dimension of the HC space is harder to quantify, so we report its exact dimension $\#HC(w, d)$ in Fig. 1, computed for some values of w and d . An upper bound (see [30, Appendix A.2] for the proof) is given by

$$\#HC(w, d) \leq \left\lceil (w + 1) \cdot (1 + \log(w + 1))^{d-1} \right\rceil. \tag{8}$$

This bound is accurate when $d = 2$ but becomes conservative as d increases.

2.2 L^2_ρ Optimality of Random Discrete L^2 Projection

Let us first introduce the following random variables S and Q , which depend on m, Λ , and the random sample \mathbf{Y} :

$$S(m, \Lambda) := \sup_{v \in \mathbb{P}_\Lambda \setminus \{v=0\}} \frac{\|v\|_{L^2_\rho}^2}{\|v\|_{\mathbf{Y}}^2}, \quad Q(m, \Lambda) := \sup_{v \in \mathbb{P}_\Lambda \setminus \{v=0\}} \frac{\|v\|_{\mathbf{Y}}^2}{\|v\|_{L^2_\rho}^2}. \tag{9}$$

The following result holds.

Proposition 1 *With $S(m, \Lambda)$ defined as in (9) and $m \geq n$, it holds that*

$$\|\phi - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{L^2_\rho} \leq \left(1 + \sqrt{S(m, \Lambda)}\right) \inf_{v \in \mathbb{P}_\Lambda} \|\phi - v\|_{L^\infty}. \tag{10}$$

Proof For any $v \in \mathbb{P}_\Lambda : v \neq \Pi_\Lambda^{\mathbf{Y}}\phi$ it holds that

$$\begin{aligned} \|\phi - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{L^2_\rho} &\leq \|\phi - v\|_{L^2_\rho} + \|v - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{L^2_\rho} \\ &= \|\phi - v\|_{L^2_\rho} + \left(\frac{\|v - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{L^2_\rho}^2}{\|v - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{\mathbf{Y}}^2} \right)^{\frac{1}{2}} \|v - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{\mathbf{Y}} \\ &\leq \|\phi - v\|_{L^2_\rho} + \left(\sup_{v \in \mathbb{P}_\Lambda \setminus \{v=0\}} \frac{\|v\|_{L^2_\rho}^2}{\|v\|_{\mathbf{Y}}^2} \right)^{\frac{1}{2}} \|v - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{\mathbf{Y}} \\ &= \|\phi - v\|_{L^2_\rho} + \sqrt{S(m, \Lambda)} \left(\|\phi - v\|_{\mathbf{Y}}^2 - \|\phi - \Pi_\Lambda^{\mathbf{Y}}\phi\|_{\mathbf{Y}}^2 \right)^{\frac{1}{2}} \\ &\leq \left(1 + \sqrt{S(m, \Lambda)}\right) \|\phi - v\|_{L^\infty}. \end{aligned}$$

In the fourth step we substituted definition (9) of the random variable S and used the Pythagorean theorem. The previous result is also true for $v = \Pi_{\Lambda}^{\mathbf{Y}}\phi$:

$$\|\phi - \Pi_{\Lambda}^{\mathbf{Y}}\phi\|_{L^2_{\rho}} \leq \|\phi - \Pi_{\Lambda}^{\mathbf{Y}}\phi\|_{L^{\infty}} \leq \left(1 + \sqrt{S(m, \Lambda)}\right) \|\phi - \Pi_{\Lambda}^{\mathbf{Y}}\phi\|_{L^{\infty}}$$

since S is nonnegative. The thesis follows from the arbitrariness of v . □

Clearly, the convergence property of the random discrete projection is strictly related to the properties of the quantity $S(m, \Lambda)$, which acts like a stability constant. In Sect. 3 we will characterize this quantity in the particular case of a one-dimensional problem ($d = 1$). The extension to multivariate functions is an ongoing work.

2.3 Asymptotic Limit for $S(m, \Lambda)$

In this section we show that when $m \rightarrow +\infty$, while Λ is kept fixed, the limit of $S(m, \Lambda)$ is a.s. 1. For the purpose of this analysis we also introduce the deterministic constants

$$k_{\infty}(\Lambda) := \sup_{v \in \mathbb{P}_{\Lambda} \setminus \{v \equiv 0\}} \frac{\|v\|_{L^{\infty}}^2}{\|v\|_{L^2_{\rho}}^2} < +\infty, \quad k_4(\Lambda) := \sup_{v \in \mathbb{P}_{\Lambda} \setminus \{v \equiv 0\}} \frac{\|v\|_{L^4_{\rho}}^2}{\|v\|_{L^2_{\rho}}^2} < +\infty. \quad (11)$$

Remark 1 Given any Λ , the constants k_{∞} and k_4 are always finite in any dimension d since the space \mathbb{P}_{Λ} is finite dimensional and the domain Γ is bounded. For example, in one dimension, any $v \in \mathbb{P}_w = \text{span}\{(Y)^j, j = 0, \dots, w\}$ can be expanded in a Legendre series. Introducing a suitable change of variable such that $v : \Gamma \rightarrow \mathbb{R}$ is mapped on $\widehat{v} : [-1, 1] \rightarrow \mathbb{R}$ we obtain

$$\begin{aligned} \|v\|_{L^{\infty}(\Gamma)} &= \|\widehat{v}\|_{L^{\infty}(-1,1)} \leq \frac{w+1}{\sqrt{2}} \|\widehat{v}\|_{L^2(-1,1)} \\ &\leq \frac{w+1}{\sqrt{2}} \sqrt{\frac{2}{|\Gamma|}} \|v\|_{L^2(\Gamma)} \leq \frac{w+1}{\sqrt{|\Gamma|\rho_{\min}}} \|v\|_{L^2_{\rho}}, \end{aligned}$$

so $k_{\infty} \leq (w+1)^2(|\Gamma|\rho_{\min})^{-1}$. One can tensorize the one-dimensional case to show that k_{∞} is bounded with respect to the maximum polynomial degree w also in higher dimensions. In the particular case of the uniform density $\rho = \mathcal{U}([-1, 1]^d)$, the values of k_{∞} and k_4 are reported in Lemma 1.

For any outcome $\{y_i\}_{i=1}^{\infty}$ of the random variables $\{Y_i\}_{i=1}^{\infty}$ we define the sequence $\{Q_m\}_m$ of functionals whose elements are defined as

$$Q_m(\cdot) := \frac{\|\cdot\|_{L^{\infty}}^2}{\|\cdot\|_{L^2_{\rho}}^2} : \mathbb{P}_{\Lambda} \setminus \{v \equiv 0\} \rightarrow \mathbb{R}^+ \quad (12)$$

and the polynomial set $\widehat{\mathbb{P}}_{\Lambda} := \{v \in \mathbb{P}_{\Lambda} : \|v\|_{L^2_{\rho}} = 1\} \subset L^2_{\rho}$.

Proposition 2 For any m and Λ the function Q_m in (12) is globally Lipschitz continuous over $\widehat{\mathbb{P}}_\Lambda$, and the Lipschitz constant has a deterministic bound that is independent of m .

Proof Consider the constant k_∞ defined in (11). Clearly $\|v\|_{L^\infty} \leq \sqrt{k_\infty}$ for all $v \in \widehat{\mathbb{P}}_\Lambda$, so the functions $\{Q_m\}_m$ are uniformly bounded. Taking arbitrary v_1 and v_2 in $\widehat{\mathbb{P}}_\Lambda$,

$$\begin{aligned} |Q_m(v_1) - Q_m(v_2)| &\leq \frac{1}{m} \sum_{i=1}^m |v_1(Y_i)^2 - v_2(Y_i)^2| \\ &= \frac{1}{m} \sum_{i=1}^m |(v_1(Y_i) + v_2(Y_i))(v_1(Y_i) - v_2(Y_i))| \\ &\leq \frac{1}{m} (\|v_1\|_{L^\infty} + \|v_2\|_{L^\infty}) \sum_{i=1}^m |v_1(Y_i) - v_2(Y_i)| \\ &\leq 2\sqrt{k_\infty} \|v_1 - v_2\|_{L^\infty} \leq 2k_\infty \|v_1 - v_2\|_{L^2_\rho}. \end{aligned} \tag{13}$$

□

From Proposition 2 it follows immediately that the sequence $\{Q_m\}_m$ is uniformly equicontinuous. Moreover, from the strong law of large numbers, since $\mathbb{E}[v^2] = \|v\|_{L^2_\rho}^2 = 1$ for all $v \in \widehat{\mathbb{P}}_\Lambda$, it follows that

$$\|v\|_{\mathbb{Y}}^2 = \frac{1}{m} \sum_{i=1}^m v(Y_i)^2 \xrightarrow{m \rightarrow +\infty} \mathbb{E}[v^2] = \|v\|_{L^2_\rho}^2, \quad \forall v \in \widehat{\mathbb{P}}_\Lambda$$

a.s.; hence the sequence $\{Q_m\}$ is also converging to 1 pointwise. We then have the following result.

Proposition 3 It holds that

$$Q_m(v) \xrightarrow{m \rightarrow +\infty} 1, \quad \text{uniformly in } \widehat{\mathbb{P}}_\Lambda, \quad \text{a.s.} \tag{14}$$

Proof For any outcome, the sequence $\{Q_m\}_{m=1}^\infty$ is (uniformly) equicontinuous on $\widehat{\mathbb{P}}_\Lambda$. The sequence also converges a.s. on $\widehat{\mathbb{P}}_\Lambda$; therefore, it a.s. converges uniformly in $\widehat{\mathbb{P}}_\Lambda$ (e.g., [33, Theorem 7.25]). □

Theorem 1 Let $S(m, \Lambda)$ and $Q(m, \Lambda)$ be the random variables defined in (9). Then, for any downward closed set Λ and $d \geq 1$, it holds that

$$\lim_{m \rightarrow +\infty} S(m, \Lambda) = \lim_{m \rightarrow +\infty} Q(m, \Lambda) = 1, \quad \text{a.s.}$$

Proof Since $Q_m(v) = Q_m(v/\|v\|_{L^2_\rho})$ for any $v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}$, we have

$$Q_m(v) \xrightarrow{m \rightarrow +\infty} 1, \quad \text{a.s. uniformly in } \mathbb{P}_\Lambda \setminus \{v \equiv 0\},$$

from which we also deduce that

$$\left(Q_m(v)\right)^{-1} \xrightarrow{m \rightarrow +\infty} 1, \quad \text{a.s. uniformly in } \mathbb{P}_\Lambda \setminus \{v \equiv 0\}.$$

This implies by uniform equicontinuity

$$\lim_{m \rightarrow +\infty} Q(m, \Lambda) = \lim_{m \rightarrow +\infty} \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} Q_m(v) = \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} \lim_{m \rightarrow +\infty} Q_m(v) = 1, \quad \text{a.s.,}$$

and the same holds for $S(m, \Lambda)$. □

Theorem 2 Consider the random variables Q and S defined in (9). In any dimension and given any downward closed set Λ , for all $\varepsilon > 0$ there exists an a.s. finite random variable \overline{M}_ε such that

$$1 - \eta(m)(L + \varepsilon) \leq Q(m, \Lambda) \leq 1 + \eta(m)(L + \varepsilon), \quad \forall m > \overline{M}_\varepsilon \quad \text{a.s.,} \quad (15)$$

$$\frac{1}{1 + \eta(m)(L + \varepsilon)} \leq S(m, \Lambda) \leq \frac{1}{1 - \eta(m)(L + \varepsilon)}, \quad \forall m > \overline{M}_\varepsilon, \quad \text{a.s.,} \quad (16)$$

with

$$\eta(m) = \sqrt{\frac{\log \log m}{m}}$$

and

$$L = \sqrt{2} \left((k_4(\Lambda))^2 - 1 \right)^{\frac{1}{2}}.$$

Proof Let Q_m be defined as in (12) and $S_m(v) = Q_m^{-1}(v)$ given any $v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}$. It holds that

$$\begin{aligned} \text{Var}(Q_m(v)) &= \frac{1}{m^2 \|v\|_{L^2_\rho}^4} \text{Var}\left(\sum_{i=1}^m (v(Y_i))^2\right) = \frac{1}{m} \frac{\|v\|_{L^4_\rho}^4 - \|v\|_{L^2_\rho}^4}{\|v\|_{L^2_\rho}^4} \\ &\leq \frac{1}{m} \left((k_4(\Lambda))^2 - 1 \right), \quad \forall v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}, \end{aligned} \quad (17)$$

where $k_4(\Lambda)$ denotes the constant of the inverse inequality between L^4_ρ and L^2_ρ introduced in (11). The value of $k_4(\Lambda)$ is reported in Lemma 1 for some polynomial spaces and uniform measure in $[-1, 1]^d$. Denote by \mathbb{P}_0 the subspace of \mathbb{P}_Λ , which contains only the constant functions. When $v \in \mathbb{P}_0 \setminus \{v \equiv 0\}$, it holds that $Q_m(v) = S_m(v) = 1$ for any m . Moreover, notice that the supremums over $\mathbb{P}_\Lambda \setminus \{v \equiv 0\}$ of the random variables Q_m and S_m are by definition the random variables Q and S :

$$Q(m, \Lambda) = \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} Q_m(v) \quad \text{and} \quad S(m, \Lambda) = \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} S_m(v).$$

Given any $v \in \mathbb{P}_\Lambda \setminus \mathbb{P}_0$, the summands, $\frac{(v(Y_i))^2}{\|v\|_{L^2_\rho}^4} - 1$, of the sequence $\{m(Q_m(v) - 1)\}_m$ are independent and identically distributed (i.i.d.) with zero mean and a finite variance that can be bounded as in (17). Therefore, using the law of the iterated logarithm (see, e.g., [24]), it holds that

$$\limsup_{m \rightarrow +\infty} \frac{Q_m(v) - 1}{\sqrt{2\text{Var}(Q_m(v)) \log \log m}} = 1, \quad \text{a.s.},$$

and, using (17),

$$-L \leq \liminf_{m \rightarrow +\infty} \frac{Q_m(v) - 1}{\eta(m)} \leq \limsup_{m \rightarrow +\infty} \frac{Q_m(v) - 1}{\eta(m)} \leq L.$$

Consequently, given any $v \in \mathbb{P}_\Lambda \setminus \mathbb{P}_0$, for all $\varepsilon > 0$ there exists $\overline{M}_\varepsilon(v)$ such that it holds

$$1 - \eta(m)(L + \varepsilon) \leq Q_m(v) \leq 1 + \eta(m)(L + \varepsilon), \quad \forall m > \overline{M}_\varepsilon(v) \quad \text{a.s.} \quad (18)$$

$$\frac{1}{1 + \eta(m)(L + \varepsilon)} \leq S_m(v) \leq \frac{1}{1 - \eta(m)(L + \varepsilon)}, \quad \forall m > \overline{M}_\varepsilon(v) \quad \text{a.s.} \quad (19)$$

Relations (18) and (19) trivially hold also for any $v \in \mathbb{P}_0 \setminus \{v \equiv 0\}$. From Proposition 2, $S_m(v)$ and $Q_m(v)$ are uniformly equicontinuous on $\widehat{\mathbb{P}}_\Lambda$; therefore, there exists an \overline{M}_ε (independent of v) for which (18) and (19) hold for all $m > \overline{M}_\varepsilon$ and for any $v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}$ a.s. Finally, since

$$\begin{aligned} Q(m, \Lambda) &= \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} Q_m(v) = \sup_{v \in \widehat{\mathbb{P}}_\Lambda} Q_m(v) \quad \text{and} \\ S(m, \Lambda) &= \sup_{v \in \mathbb{P}_\Lambda \setminus \{v \equiv 0\}} S_m(v) = \sup_{v \in \widehat{\mathbb{P}}_\Lambda} S_m(v), \end{aligned}$$

this implies the thesis. □

Lemma 1 *In the case of the uniform density $\rho = \mathcal{U}([-1, 1]^d)$, the constants $k_\infty(\Lambda)$ and $k_4(\Lambda)$ of the inverse inequalities between $L^\infty - L^2_\rho$ and $L^4_\rho - L^2_\rho$ introduced in (11) satisfy*

$$k_\infty(w) \leq (w + 1)^2, \quad k_4(w) \leq w + 1 \quad (20)$$

in the one-dimensional case $d = 1$ and

$$\begin{aligned} k_\infty(\Lambda) &\leq \begin{cases} (\#\Lambda)^2, & \text{with the TP and TD spaces,} \\ 2^d(w + 1)\#\Lambda, & \text{with the HC space,} \end{cases} \\ k_4(\Lambda) &\leq \begin{cases} \#\Lambda, & \text{with the TP and TD spaces,} \\ (2^d(w + 1)\#\Lambda)^{\frac{1}{2}}, & \text{with the HC space,} \end{cases} \end{aligned} \quad (21)$$

in the multidimensional case.

See [30, Appendix B] for the proof of this lemma.

3 Error Analysis in One Dimension

We restrict the analysis to the case where $d = 1$, with Γ a closed interval in \mathbb{R} , and consider the polynomial space $\mathbb{P}_\Lambda = \text{span}\{y^j, j = 0, \dots, w\}$. For convenience, we rename the polynomial space \mathbb{P}_w and the random discrete projector Π_w^Y . The main result of this section is Theorem 3. Its proof is postponed until the end of the section because we will need several lemmas and intermediate results.

Theorem 3 *For any $\alpha \in (0, 1)$, when the density ρ satisfies Assumption 1 and under the condition*

$$\frac{2m\rho_{min}}{3 \log((m + 1)/\alpha)} \geq \frac{8\sqrt{3} w^2}{|\Gamma|} \tag{22}$$

it holds that

$$Pr \left(\|\phi - \Pi_w^Y \phi\|_{L^2_\rho} \leq \left(1 + \sqrt{\frac{3\rho_{max}}{\rho_{min}} \log \frac{m + 1}{\alpha}} \right) \inf_{v \in \mathbb{P}_w} \|\phi - v\|_{L^\infty} \right) \geq 1 - \alpha. \tag{23}$$

The previous theorem states that, with a confidence level of $1 - \alpha$, the discrete projection with random points is (nearly) optimal up to a logarithmic factor in m , provided m is large enough with respect to the polynomial degree w , and satisfies condition (22). Theorem 3 holds for any density that satisfies Assumption 1. In [30, Chapter 3] it is shown that this assumption is strictly required, i.e., the thesis of the theorem does not hold when the density is not bounded away from zero. In such cases, a number of points larger than that prescribed by condition (22) is required to achieve an optimal convergence. Moreover, in [30, Chapter 3] it is proven that the same holds for the analysis proposed in [15] by estimating the growth of the Jacobi polynomials that are associated with the beta density family.

We now proceed to derive some probabilistic results on the distribution of order statistics for the random sample points Y_i .

3.1 Useful Results on Order Statistics of Random Sample Y

To study the order statistics of the random points Y_i , it is more convenient to transform them into uniform random variables in $[0, 1]$. Let $F(y) = \int_{-\infty}^y \rho(z) dz$ be the cumulative distribution of the random variable Y . Then, $U = F^{-1}(Y)$ is a uniform random variable in $[0, 1]$, and we introduce the sample $U_i = F^{-1}(Y_i)$ for all $i = 1, \dots, m$, so that $U_1, \dots, U_m \stackrel{iid}{\sim} \mathcal{U}(0, 1)$. We know that the order statistics $U_{(1)} < U_{(2)} < \dots < U_{(m)}$ are

$$U_{(i)} \sim \text{beta}(i, m + 1 - i), \quad i = 1, \dots, m,$$

where we recall that a $\text{beta}(i, m + 1 - i)$ random variable has distribution

$$\rho(y) = \frac{m!}{(i-1)!(m-i)!} y^{i-1} (1-y)^{m-i}, \quad y \in [0, 1].$$

Let us define $\Xi_0 = U_{(1)}$, $\Xi_i = U_{(i+1)} - U_{(i)}$ for $i = 1, \dots, m-1$ and $\Xi_m = 1 - U_{(m)}$. It can be shown that $\Xi_i \sim \text{beta}(1, m)$ for all $i = 0, \dots, m$; see [17, p. 14], where a more general result on the distribution of $U_{(i_2)} - U_{(i_1)}$ is proven, namely

$$U_{(i_2)} - U_{(i_1)} \sim \text{beta}(i_2 - i_1, m - i_2 + i_1 + 1), \quad 1 \leq i_1 < i_2 \leq m.$$

In particular, the distribution is independent of the values i_1 and i_2 and depends only on their difference $i_2 - i_1$. The following result gives a bound on the probability distribution of $\max_{i=0, \dots, m} \Xi_i$.

Lemma 2 For any $\alpha \in (0, 1)$ and $m \in \mathbb{N}$

$$\Pr \left(\max_{i=0, \dots, m} \Xi_i > \frac{\log((m+1)/\alpha)}{m} \right) \leq \alpha.$$

Proof Trivially, if $0 < \alpha \leq \frac{m+1}{\exp(m)}$, then $\frac{\log((m+1)/\alpha)}{m} \geq 1$ and

$$\Pr \left(\max_{i=0, \dots, m} \Xi_i > 1 \right) = 0 < \alpha.$$

Consider now $\frac{m+1}{\exp(m)} < \alpha < 1$. Rewriting the random event

$$\left\{ \max_{i=0, \dots, m} \Xi_i > \delta \right\} = \bigcup_{i=0}^m \left\{ \Xi_i > \delta \right\},$$

we have, for $0 < \delta < 1$,

$$\begin{aligned} \Pr \left(\max_{i=0, \dots, m} \Xi_i > \delta \right) &= \Pr \left(\bigcup_{i=0}^m \left\{ \Xi_i > \delta \right\} \right) \leq \sum_{i=0}^m \Pr \left(\Xi_i > \delta \right) \\ &= (m+1)(1-\delta)^m. \end{aligned} \tag{24}$$

The last step in (24) exploits $\Xi_i \sim \text{beta}(1, m)$, so that $\Pr(\Xi_i > \delta) = (1-\delta)^m$ for each $i = 0, \dots, m$. Therefore,

$$\begin{aligned} &\Pr \left(\max_{k=0, \dots, m} \Xi_k > \frac{\log((m+1)/\alpha)}{m} \right) \\ &\leq (m+1) \left(1 - \frac{\log((m+1)/\alpha)}{m} \right)^m \end{aligned}$$

$$\begin{aligned}
 &= (m + 1) \left(\left(1 - \frac{\log((m + 1)/\alpha)}{m} \right)^{\frac{m}{\log((m+1)/\alpha)}} \right)^{\log((m+1)/\alpha)} \\
 &\leq (m + 1) e^{-\log((m+1)/\alpha)} = \alpha.
 \end{aligned}$$

□

Let now $\Gamma = [\Gamma_{\min}, \Gamma_{\max}]$, $Y_0 = \Gamma_{\min}$, and $Y_{m+1} = \Gamma_{\max}$, and denote by $Y_{(i)}$ the ordered sample Y_0, \dots, Y_{m+1} . The following lemma is an immediate consequence of Lemma 2.

Lemma 3 For any $\alpha \in (0, 1)$ and $m \in \mathbb{N}$

$$Pr \left(\max_{i=1, \dots, m+1} |Y_{(i)} - Y_{(i-1)}| > \frac{\log((m + 1)/\alpha)}{m\rho_{\min}} \right) \leq \alpha.$$

Proof The result follows by observing that

$$Y_{(i+1)} - Y_{(i)} \leq \frac{1}{\rho_{\min}} \Xi_i, \quad \forall i = 0, \dots, m.$$

□

3.2 Relation Between $\|\cdot\|_{L^2_\rho}$ and $\|\cdot\|_{\mathbf{Y}}$ on \mathbb{P}_w

We are interested in finding an inequality between the continuous norm $\|\cdot\|_{L^2_\rho}$ and the discrete norm $\|\cdot\|_{\mathbf{Y}}$, i.e., in finding a suitable random variable B such that

$$\|v\|_{L^2_\rho}^2 \leq B \|v\|_{\mathbf{Y}}^2, \quad \forall v \in \mathbb{P}_w.$$

This will be the random variable appearing in Proposition 1.

Now we introduce the notion of covering of the interval Γ , i.e., to each ordered random point $Y_{(i)}$ we associate an interval I_i satisfying the requirement that

$$\bigcup_{i=1}^m I_i = \Gamma. \tag{25}$$

In other words, the family of intervals $\{I_i\}_i$ is a (finite) covering of the domain Γ . Since the points $\{Y_i\}_i$ are random variables, the intervals $\{I_i\}_i$ generate a random covering of Γ . In one dimension, it is easy to build a finite covering of mutually disjoint intervals

$$I_1 = [Y_{(1)} - \Delta_0, Y_{(1)} + \Delta_1], \quad I_i = (Y_{(i)} - \Delta_{i-1}, Y_{(i)} + \Delta_i], \quad i = 2, \dots, m, \tag{26}$$

satisfying (25) by taking

$$\Delta_i = \begin{cases} |Y_{(0)} - Y_{(1)}|, & i = 0, \\ \frac{|Y_{(i+1)} - Y_{(i)}|}{2}, & i = 1, \dots, m - 1, \\ |Y_{(m+1)} - Y_{(m)}|, & i = m. \end{cases} \tag{27}$$

In general, the sets I_i are not centered in the corresponding random point $Y_{(i)}$. It is useful to split each interval I_i into its left and right parts, $I_i^- = I_i \cap \{y \in \Gamma : y \leq Y_{(i)}\}$, $I_i^+ = I_i \cap \{y \in \Gamma : y > Y_{(i)}\}$, with measures $|I_i^-| = \Delta_{i-1}$ and $|I_i^+| = \Delta_i$, respectively. We also define the random variable ΔY as

$$\Delta Y = \max_{i=1, \dots, m} |I_i|. \tag{28}$$

The following result is a consequence of Lemma 3.

Corollary 1 *For any $\alpha \in (0, 1)$ and $m \in \mathbb{N}$, $m \geq 2$ it holds that*

$$Pr \left(\Delta Y > \frac{3 \log((m + 1)/\alpha)}{2m\rho_{\min}} \right) \leq \alpha.$$

Proof It is enough to notice that

$$\begin{aligned} |I_1| &= (Y_{(1)} - Y_{(0)}) + \frac{Y_{(2)} - Y_{(1)}}{2} \leq \frac{3}{2} \max_{i=1, \dots, m+1} (Y_{(i)} - Y_{(i-1)}), \\ |I_i| &= \frac{Y_{(i)} - Y_{(i-1)}}{2} + \frac{Y_{(i+1)} - Y_{(i)}}{2} \\ &\leq \max_{j=1, \dots, m+1} (Y_{(j)} - Y_{(j-1)}), \quad \forall i = 2, \dots, m - 1, \\ |I_m| &= \frac{Y_{(m)} - Y_{(m-1)}}{2} + (Y_{(m+1)} - Y_{(m)}) \leq \frac{3}{2} \max_{i=1, \dots, m+1} (Y_{(i)} - Y_{(i-1)}), \end{aligned}$$

so that

$$Pr (\Delta Y > \zeta) \leq Pr \left(\frac{3}{2} \max_{i=1, \dots, m-1} (Y_{(i)} - Y_{(i-1)}) > \zeta \right), \quad \forall \zeta > 0.$$

□

In what follows we will need a Markov inequality for the polynomial space $\widehat{\mathbb{P}}_w$, (e.g., [12]),

$$\left\| \frac{\partial v}{\partial y} \right\|_{L^2(\Gamma)} \leq c_I w^2 \|v\|_{L^2(\Gamma)}, \quad \forall v \in \mathbb{P}_w, \tag{29}$$

where the constant $c_I = 2\sqrt{3}/|\Gamma|$ depends on the length $|\Gamma|$ of the interval.

We now define two events:

$$\widehat{\Sigma}_w = \left\{ \Delta Y \leq \frac{1}{4c_I w^2} \right\} \quad \text{and} \quad \Sigma_m = \left\{ \Delta Y \leq \frac{3 \log((m + 1)/\alpha)}{2m\rho_{\min}} \right\}. \tag{30}$$

Observe that, taking m large enough, the probability of $\widehat{\Sigma}_w$ can be made arbitrarily close to 1. The following lemma makes the point more precisely.

Lemma 4 For any $\alpha \in (0, 1)$, under the condition

$$\frac{2m\rho_{\min}}{3 \log((m + 1)/\alpha)} \geq 4 c_I w^2, \tag{31}$$

the following inclusion holds:

$$\Sigma_m \subset \widehat{\Sigma}_w$$

and

$$Pr(\widehat{\Sigma}_w) \geq Pr(\Sigma_m) \geq 1 - \alpha.$$

Proof Clearly, under (31),

$$\Delta Y \leq \frac{3 \log((m + 1)/\alpha)}{2m\rho_{\min}} \implies \Delta Y \leq \frac{1}{4c_I w^2},$$

and the inclusion $\Sigma_m \subset \widehat{\Sigma}_w$ is proven. The bound from below on the corresponding probabilities is an immediate consequence of Lemma 1. \square

Now we are ready to formulate the main result of this subsection.

Theorem 4 Define on $\widehat{\Sigma}_w$ the random variable

$$B = \frac{m \Delta Y \rho_{\max}}{1 - 2 \Delta Y c_I w^2}. \tag{32}$$

Then, in $\widehat{\Sigma}_w$ it holds that

$$\|v\|_{L^2_\rho}^2 \leq B \|v\|_{\mathbb{Y}}^2, \quad \forall v \in \mathbb{P}_w. \tag{33}$$

Moreover, under condition (31), in $\Sigma_m \subset \widehat{\Sigma}_w$ it holds that

$$B \leq \frac{3\rho_{\max}}{\rho_{\min}} \log((m + 1)/\alpha).$$

Proof For convenience the proof uses the standard $L^2(\Gamma)$ norm instead of the weighted norm L^2_ρ . Remember that in this case, $\|\cdot\|_{L^2_\rho}^2 \leq \rho_{\max} \|\cdot\|_{L^2(\Gamma)}^2$. To lighten the notation, we also introduce on each interval I_i^\pm the short notation $\|\cdot\|_{I_i^-} := \|\cdot\|_{L^2(I_i^-)}$ and $\|\cdot\|_{I_i^+} := \|\cdot\|_{L^2(I_i^+)}$.

Take $v \in \mathbb{P}_w$ arbitrarily. The following standard inequalities are used in the sequel:

$$\begin{aligned} & \left| v(y)^2 - v(Y_{(i)})^2 \right| \\ &= \left| \int_{Y_{(i)}}^y (v^2)'(\xi) d\xi \right| = \left| \int_{Y_{(i)}}^y 2v(\xi)v(\xi)' d\xi \right| \leq \begin{cases} 2\|v\|_{I_i^+} \|v'\|_{I_i^+}, & \forall y \in I_i^+, \\ 2\|v\|_{I_i^-} \|v'\|_{I_i^-}, & \forall y \in I_i^-. \end{cases} \end{aligned} \tag{34}$$

Integrating (34) over I_j^- yields

$$\int_{I_i^-} v(y)^2 dy - v(Y_{(i)})^2 |I_i^-| = \int_{I_i^-} (v(y)^2 - v(Y_{(i)})^2) dy \leq \int_{I_i^-} |v(y)^2 - v(Y_{(i)})^2| dy \leq 2 |I_i^-| \|v\|_{I_i^-} \|v'\|_{I_i^-}, \tag{35}$$

and analogously over I_j^+

$$\int_{I_i^+} v(y)^2 dy - v(Y_{(i)})^2 |I_i^+| \leq 2 |I_i^+| \|v\|_{I_i^+} \|v'\|_{I_i^+}. \tag{36}$$

Summing (35) and (36) we obtain

$$\int_{I_i} v(y)^2 dy - |I_i| v(Y_{(i)})^2 \leq 2 (|I_i^-| \|v\|_{I_i^-} \|v'\|_{I_i^-} + |I_i^+| \|v\|_{I_i^+} \|v'\|_{I_i^+}) \leq 2 |I_i| \|v\|_{I_i} \|v'\|_{I_i},$$

which implies

$$\int_{I_i} v(y)^2 dy \leq |I_i| (v(Y_{(i)})^2 + 2 \|v\|_{I_i} \|v'\|_{I_i}). \tag{37}$$

Summing (37) over all the intervals and using the definitions of ΔY and $\|\cdot\|_Y$, followed by the Cauchy–Schwarz inequality and (29) we have

$$\begin{aligned} \|v\|_{L^2(\Gamma)}^2 &= \sum_{i=1}^m \int_{I_i} v(y)^2 dy \\ &\leq \sum_{i=1}^m |I_i| v(Y_{(i)})^2 + 2 \sum_{i=1}^m |I_i| \|v\|_{I_i} \|v'\|_{I_i} \\ &\leq \Delta Y \sum_{i=1}^m v(Y_{(i)})^2 + 2 \Delta Y \sum_{i=1}^m \|v\|_{I_i} \|v'\|_{I_i} \\ &= m \Delta Y \|v\|_Y^2 + 2 \Delta Y \sum_{i=1}^m \|v\|_{I_i} \|v'\|_{I_i} \\ &\leq m \Delta Y \|v\|_Y^2 + 2 \Delta Y \|v\|_{L^2(\Gamma)} \|v'\|_{L^2(\Gamma)} \\ &\leq m \Delta Y \|v\|_Y^2 + 2 \Delta Y c_I w^2 \|v\|_{L^2(\Gamma)}. \end{aligned} \tag{38}$$

Rearranging the terms in (38) we obtain

$$(1 - 2\Delta Y c_I w^2) \|v\|_{L^2(\Gamma)}^2 \leq m \Delta Y \|v\|_{\mathbf{Y}}^2. \tag{39}$$

The coefficient on the left-hand side is nonzero on $\widehat{\Sigma}_w$, so we have proven (32)–(33). If we now restrict (39) to Σ_m under condition (31), then from Lemma 4 we have

$$\Delta Y \leq \frac{1}{4c_I w^2} \quad \text{and} \quad \Delta Y \leq \frac{3 \log((m + 1)/\alpha)}{2m\rho_{\min}},$$

so that

$$B = \frac{\rho_{\max} m \Delta Y}{1 - 2\Delta Y c_I w^2} \leq \frac{3\rho_{\max}}{\rho_{\min}} \log((m + 1)/\alpha),$$

and this concludes the proof. □

Remark 2 Theorem 4 states that on Σ_m , which is an event of probability at least $1 - \alpha$, the discrete and continuous norms are equivalent up to a logarithmic factor if condition (31) is fulfilled, which, roughly speaking, corresponds to $m \propto w^2$.

Remark 3 In the arguments used in Lemmas 1 and 4 and Theorem 3, the covering (26) can be slightly improved by taking an m -dependent partition

$$\Delta_i^+ = \frac{(m - i) |Y_{(i+1)} - Y_{(i)}|}{m}, \quad \Delta_i^- = \frac{i |Y_{(i+1)} - Y_{(i)}|}{m}, \quad i = 0, \dots, m,$$

and building the covering as

$$I_1 = \left[Y_{(1)} - \Delta_0^+, Y_{(1)} + \Delta_1^- \right], \quad I_i = \left(Y_{(i)} - \Delta_{i-1}^+, Y_{(i)} + \Delta_i^- \right), \quad i = 2, \dots, m. \tag{40}$$

Using this covering in the proof of Lemma 1, it holds that

$$|I_i| \leq \frac{m + 1}{m} \max_{j=1, \dots, m+1} (Y_{(j)} - Y_{(j-1)}), \quad \forall i = 1, \dots, m.$$

This relation is sharper since $(m + 1)/m \leq 3/2$ when $m \geq 2$. The use of the sharper covering (40) allows us to obtain a sharper condition

$$\frac{m^2 \rho_{\min}}{(m + 1) \log((m + 1)/\alpha)} \geq 4 c_I w^2, \quad m \geq 2,$$

which slightly improves condition (22), and a sharper bound

$$B \leq \frac{2(m + 1)\rho_{\max}}{m\rho_{\min}} \log((m + 1)/\alpha), \quad m \geq 2,$$

but with more complicated expressions.

Remark 4 The proof of Theorem 4 relies on the uniform covering properties of the sampling points $\{Y_1, \dots, Y_m\}$. The result then applies to any sequence, not necessarily random, of points that provides a covering with maximal spacing $\Delta Y \propto m^{-1}$. In particular, an analogous result to Theorem 3 holds for a deterministic sequence of equally spaced points, provided that a relation $m \propto w^2$ is fulfilled to retain the stability of the projection and avoid the blow-up of the random variable B . The use of random uniform points instead of deterministic equally spaced points is clearly of interest in higher dimensions since in this case the approach based on a uniform deterministic sampling is of limited interest because the number of points required to preserve uniformity must scale exponentially with respect to the dimension.

3.3 Proof of Theorem 3

The proof of this theorem is merely a collection of results from Proposition 1, Theorem 4, and Lemma 1.

Proof of Theorem 3 We consider the event Σ_m defined in (30). From Lemma 1 we know that under condition (22) the probability of this event is at least $1 - \alpha$. From Theorem 4 it holds that

$$\|v\|_{L^2_\rho}^2 \leq \frac{3\rho_{\max}}{2\rho_{\min}} \log\left(\frac{m+1}{\alpha}\right) \|v\|_{\mathbf{Y}}^2, \quad \text{in } \Sigma_m,$$

uniformly with respect to $v \in \mathbb{P}_w$. We then apply Proposition 1 to any outcome in Σ_m to conclude that

$$\|\phi - \Pi_w^{\mathbf{Y}}\phi\|_{L^2_\rho} \leq \left(1 + \sqrt{\frac{3\rho_{\max}}{2\rho_{\min}} \log\left(\frac{m+1}{\alpha}\right)}\right) \inf_{v \in \mathbb{P}_w} \|\phi - v\|_{L^\infty}, \quad \forall \phi \in L^\infty(\Gamma).$$

This concludes the proof. □

4 Algebraic Formulation

The value of n depends on the particular polynomial space (TP, TD, HC, ...), the maximal polynomial degree used, w , and the dimension of the physical space, d . The number of points m must satisfy the constraint

$$m \geq n$$

to have an overdetermined problem (more data than unknowns). We showed in Sect. 3 that for univariate functions, m should scale as $m \propto w^2$ to have a stable discrete projection. As a general rule, to choose m for multivariate functions and arbitrary polynomial spaces, we consider the formula

$$m = c n^\gamma, \tag{41}$$

where c is a positive constant and $\gamma \geq 1$ a real number. We restrict our numerical tests in Sect. 5 to the cases $\gamma = 1$ and $\gamma = 2$.

Given the polynomial space, we denote by $D \in \mathbb{R}^{m \times n}$ the design matrix. Its element $[D]_{ij}$ contains the j th L^2_ρ -orthonormal basis function ψ_j evaluated in the i th random point Y_i :

$$[D]_{ij} = \psi_j(Y_i).$$

The discrete random projection $\Pi_{\Lambda}^{\mathbf{Y}}\phi$ can be expressed in terms of the orthonormal basis $\{\psi_j\}_j$ as

$$\Pi_{\Lambda}^{\mathbf{Y}}\phi(y) = \sum_{j=1}^n x_j \psi_j(y), \quad y \in \Gamma. \tag{42}$$

Then the algebraic problem to determine the unknown vector of coefficients x can be formulated as

$$x = \operatorname{argmin}_{z \in \mathbb{R}^n} \|Dz - b\|_2, \tag{43}$$

where $b \in \mathbb{R}^{m \times 1}$ contains the evaluations of the target function $[b]_i = \phi(Y_i)$. Although we omit the subscript \mathbf{Y} to simplify the notation, D is a random matrix and x, b are random vectors since they depend on the random sample \mathbf{Y} . The normal equations allow us to rewrite the rectangular system embedded in (43) as a square system:

$$D^T D x = D^T b. \tag{44}$$

In practice we solve problem (43) by a QR factorization of the matrix D . Still, the formulation (44) will be useful to measure the ill-conditioning of the problem through the condition number of the matrix $D^T D$.

To evaluate the approximation error, we have considered a cross-validation approach: a random set $y_1^{cv}, \dots, y_{100}^{cv}$ of 100 cross-validating points is chosen initially, and the corresponding design matrix D_{cv} is computed. The evaluations of ϕ in these points are stored in b_{cv} . Then the cross-validated error in ∞ -norm is defined as

$$\|\phi - \Pi_{\Lambda}^{\mathbf{Y}}\phi\|_{cv} := \|D_{cv}x - b_{cv}\|_{\infty}. \tag{45}$$

Note that $\|\cdot\|_{cv}$ is not a norm on the function space of ϕ ; we abuse the norm notation in the figures with cross-validation errors below to emphasize the dependence on ϕ . To estimate the variability of (45) due to the random sampling of the m collocation points, we have repeated the calculation of the coefficient vector x over r independent sets of outcomes $\{y_i^p, i = 1, \dots, m\}$, with $p = 1, \dots, r$. Accordingly, we denote by $x_{(p)}$ the outcome of x obtained with the p th set. Subsequently we compute the sample average error by

$$\bar{E}_{cv} = \frac{\sum_{p=1}^r \|D_{cv}x_{(p)} - b_{cv}\|_{\infty}}{r} \tag{46}$$

and the sample standard deviation by

$$s_E = \sqrt{\frac{1}{r-1} \sum_{p=1}^r \left(\|D_{cv}x^{(p)} - b_{cv}\|_\infty - \bar{E}_{cv} \right)^2}. \tag{47}$$

We also aim to analyze the condition number of the random matrix $D^T D$,

$$K(D^T D) := \frac{\sigma_{\max}(D^T D)}{\sigma_{\min}(D^T D)}, \tag{48}$$

where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ are the maximum and minimum singular values, respectively. Again, denoting by $D_{(p)}$ the design matrix built with the p th set $\{y_i^p, i = 1, \dots, m\}$ of outcomes, we estimate the mean condition number \bar{K} over the r repetitions by

$$\bar{K} = \frac{\sum_{p=1}^r K(D_{(p)}^T D_{(p)})}{r} \tag{49}$$

and the standard deviation by

$$s_K = \sqrt{\frac{1}{r-1} \sum_{p=1}^r \left(K(D_{(p)}^T D_{(p)}) - \bar{K} \right)^2}. \tag{50}$$

4.1 Condition Number of $D^T D$

In the rest of this section we show how the condition number of problem (44) relates to some quantities previously introduced. All the contents of this section hold for a generic polynomial space, in any dimension. Accordingly, we refer to the polynomial space as \mathbb{P}_Λ .

Proposition 4 *The spectral condition number (2-norm) of the matrix $D^T D$, as defined in (48), is equal to*

$$K(D^T D) = Q(m, \Lambda) S(m, \Lambda), \tag{51}$$

where $Q(m, \Lambda)$ and $S(m, \Lambda)$ are defined in (9).

Proof The symmetric random matrix $D^T D$ is a.s. positive definite under Assumption 1 on ρ . Let $v(y) = \sum_{j=1}^n x_j \psi_j(y)$ be an arbitrary polynomial in \mathbb{P}_Λ . Then the i th element of the vector Dx is

$$[Dx]_i = \sum_{j=1}^n x_j \psi_j(Y_i) = v(Y_i). \tag{52}$$

By definition (4) of the random discrete inner product

$$\|v\|_{\mathbf{Y}}^2 = \frac{1}{m} \sum_{i=1}^m (v(Y_i))^2 \tag{53}$$

and by (42) and the L^2_ρ -ortonormality of the basis $\{\psi_j\}_j$, we have

$$\|v\|_{L^2_\rho}^2 = \left(\sum_{j=1}^n x_j \psi_j, \sum_{j=1}^n x_j \psi_j \right)_{L^2_\rho} = \sum_{j=1}^n (x_j)^2 = x^T x. \tag{54}$$

Using in sequence (54), (52), and (53) yields

$$\frac{x^T D^T D x}{x^T x} = \frac{\sum_{i=1}^m (v(Y_i))^2}{\|v\|_{L^2_\rho}^2} = \frac{m \|v\|_{\mathbf{Y}}^2}{\|v\|_{L^2_\rho}^2}. \tag{55}$$

Therefore, the largest and smallest eigenvalues of $D^T D$ correspond to $mQ(m, \Lambda)$ and $mS(m, \Lambda)^{-1}$, respectively. The conclusion follows from definition (48) of the condition number. □

Remark 5 In the one-dimensional case we can easily establish an equivalence of norms between $\|\cdot\|_{\mathbf{Y}}$ and $\|\cdot\|_{L^2_\rho}$, collecting the results of Theorem 4 and Remark 1. That is, under condition (22), in the event Σ_m , which occurs with probability at least $1 - \alpha$, we have

$$\frac{\sqrt{|\Gamma| \rho_{\min}}}{1 + w} \|v\|_{\mathbf{Y}}^2 \leq \|v\|_{L^2_\rho}^2 \leq \frac{3\rho_{\max}}{\rho_{\min}} \log\left(\frac{m + 1}{\alpha}\right) \|v\|_{\mathbf{Y}}^2, \quad v \in \mathbb{P}_w,$$

from which we obtain the bound on the condition number

$$K(D^T D) \leq \frac{3\rho_{\max}}{\sqrt{|\Gamma|}(\rho_{\min})^{3/2}} (w + 1) \log((m + 1)/\alpha), \quad \text{in } \Sigma_m. \tag{56}$$

However, we have observed numerically that the bound (56) is very pessimistic because under condition (22) the condition number seems to be uniformly bounded with respect to m and w .

A direct consequence of Theorem 1 is that

$$K(D^T D) \xrightarrow{m \rightarrow +\infty} 1, \quad \text{a.s.}$$

This is confirmed numerically. Figure 2 shows the numerical results obtained for a one-dimensional problem with an overkilling rule $m = 100n^4$ to simulate the asymptotic case $m \rightarrow +\infty$.

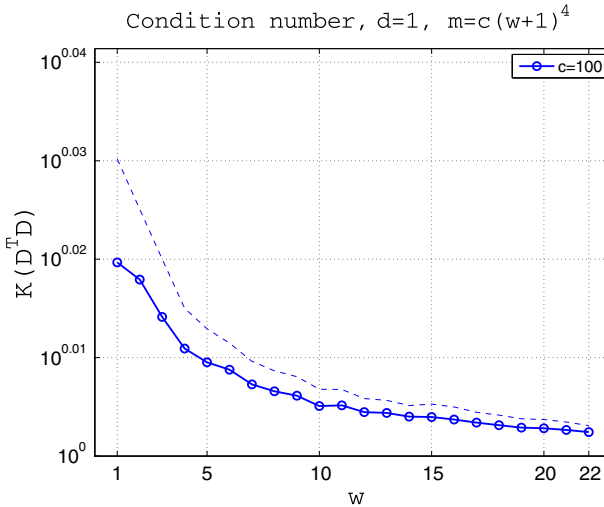


Fig. 2 Condition number (48), $d = 1, m = 100n^4$. *Continuous marked line*: mean condition number (49) over $r = 200$ repetitions; *dashed line*: mean (49) plus one standard deviation (50). The scale on the y-axis ranges from $10^0 = 1$ to $10^{0.04} = 1.0965$

5 Numerical Results

We present an illustrative selection of results from an extensive set of numerical tests. The aim is to seek the correct relation between the number of sampling points, m , and the dimension of the polynomial space to obtain a stable and accurate approximation. The following issues are addressed:

- How the condition number (48) depends on w, d, c, γ and the choice of the polynomial space;
- Analogously, how the cross-validation error (45) behaves.

In the convergence plots presented in the remainder of this section, we show the average error (46) and condition number (49), as well as their average plus one standard deviation.

5.1 One-Dimensional Case

We first investigate the dependence of the condition number (48) on the rule (41) used to select the number $m(w)$ of sampling points. Observe that in the one-dimensional case, $n = w + 1$.

As seen in Fig. 3, the condition number behaves differently depending on the rule chosen. In Fig. 3 (left), we report the results obtained with the linear rule $m = cn$, corresponding to $\gamma = 1$ in (41). We tested several values for c ranging from 2 to 20. All cases show an exponential growth of the condition number with respect to w , with rates decreasing with increasing c , as one would expect. Using $r = 10,000$ repetitions the observed average condition number still shows a large variability. This is due to

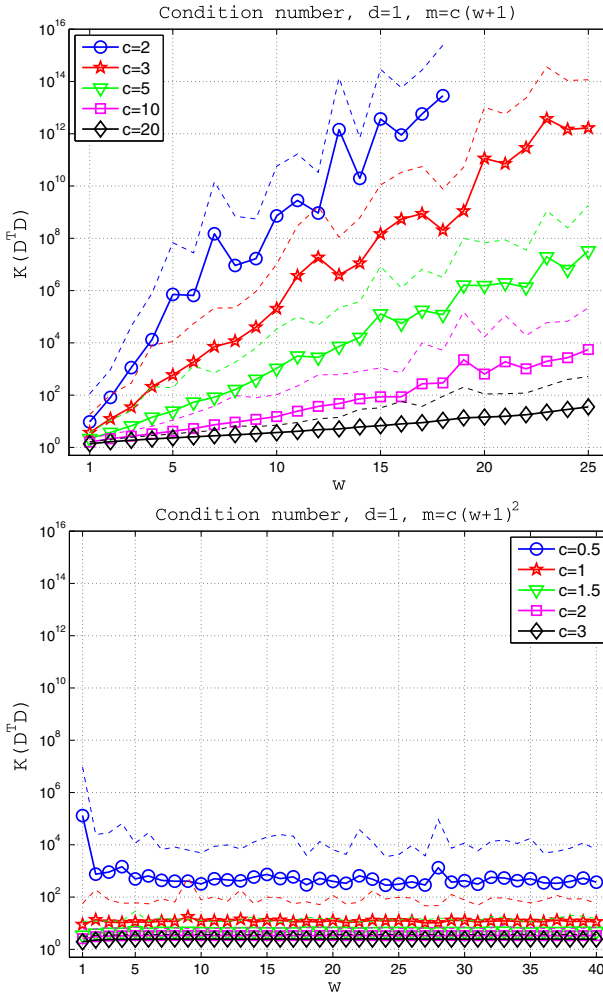


Fig. 3 Condition number (48), $d = 1$. Continuous marked lines: mean condition number (49) over $r = 10,000$ repetitions; dashed lines: mean (49) plus one standard deviation (50). Top: $m = cn$; bottom: $m = cn^2$

the large standard deviations of the condition number, as indicated in the figure by the dashed line.

Note that the range of w goes up to 25, so in this range the choice of the largest c yields a linear rule that uses more sample points than some of the quadratic rules (shown on the right-hand side of Fig. 3).

In contrast to the exponential growth observed when using the linear rule, the results using the quadratic rule exhibit a condition number that is approximately constant for w ranging from 1 to 40. Fluctuations become smaller when c increases. This behavior is consistent with the theoretical results in Sect. 3.

We now proceed to illustrate the convergence of the error for a few functions of varying regularity in Figs. 4–7. We focus on three target functions: an exponential

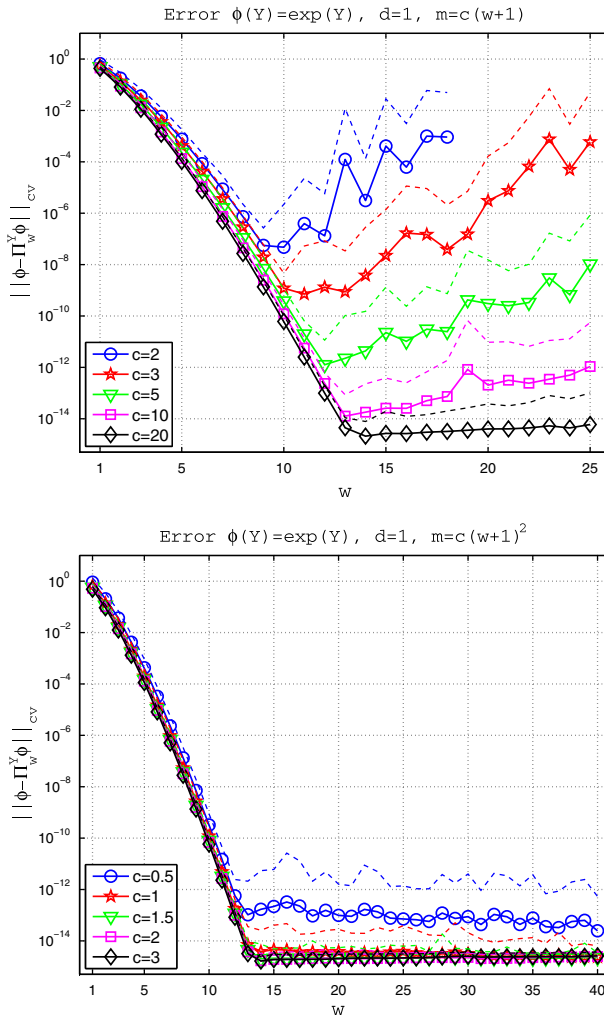


Fig. 4 Error (45) for function (57). *Continuous marked lines*: mean error (46) over $r = 10,000$ repetitions; *dashed lines*: mean (46) plus one standard deviation (47). *Top*: $m = cn$; *bottom*: $m = cn^2$

entire function,

$$\phi(Y) = \exp(Y), \quad Y \sim \mathcal{U}([-1, 1]); \tag{57}$$

a meromorphic function,

$$\phi(Y) = \frac{1}{1 + \beta Y}, \quad Y \sim \mathcal{U}([-1, 1]), \tag{58}$$

that is, a function that is analytic in $[-1, 1]$ provided that $\beta \in (-1, 1)$; and a function with lower regularity,

$$\phi(Y) = |Y|^3, \quad Y \sim \mathcal{U}([-1, 1]). \tag{59}$$

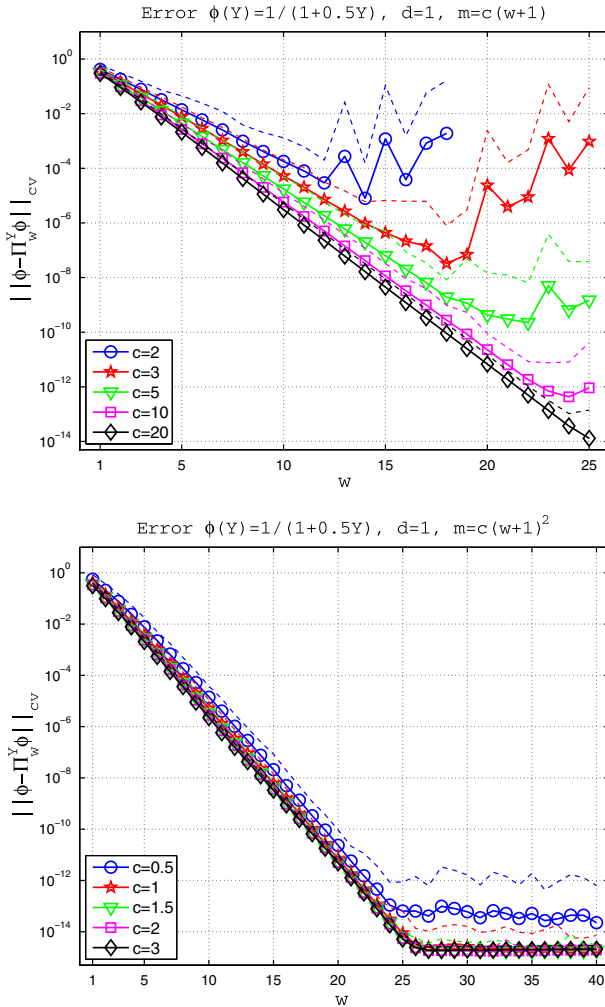


Fig. 5 Error (45) for function (58) with $\beta = 0.5$. Continuous marked lines: mean error (46) over $r = 10,000$ repetitions; dashed lines: mean (46) plus one standard deviation (47). Top: $m = c n$; bottom: $m = c n^2$

Figure 4 shows the error computed as in (45), in approximating the exponential function (57) with different choices of c and γ in rule (41). The quadratic rule (on the right) displays the same superexponential, optimal, convergence with respect to w independently of the constant c . The convergence is up to machine precision.

In contrast, the linear rule (left) displays a deterioration of the convergence using small values of c . The deterioration is due to the ill-conditioning of the matrix $D^T D$ when w increases. As noted earlier, the largest value of c yields at least as many sample points as the quadratic rule with the smallest value of c in the shown range, and the errors behave accordingly. Again, the fluctuations in the average error decrease with increasing c .

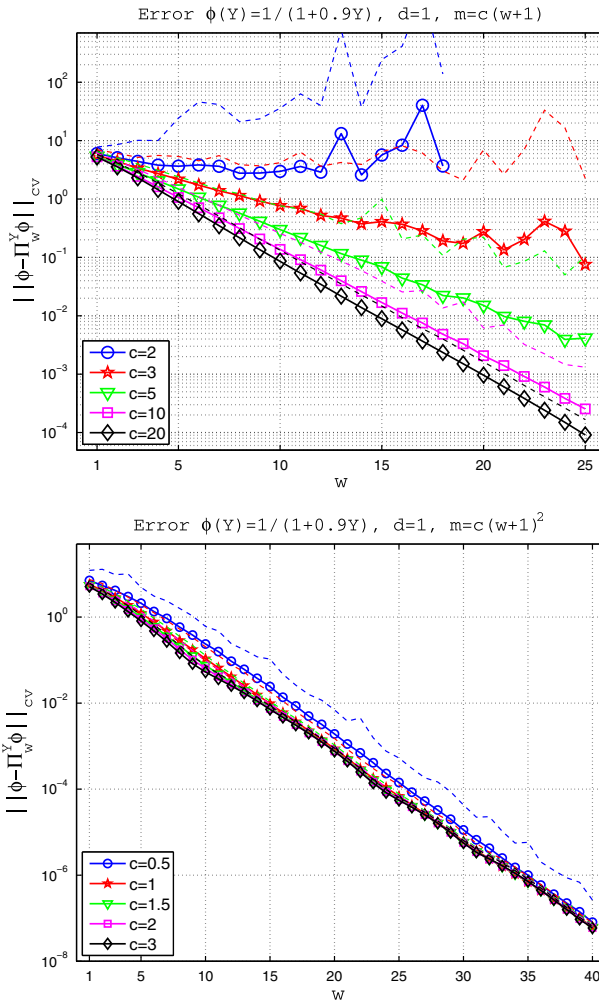


Fig. 6 Error (45) for function (58) with $\beta = 0.9$. *Continuous marked lines*: mean error (46) over $r = 10,000$ repetitions; *dashed lines*: mean (46) plus one standard deviation (47). *Top*: $m = c n$; *bottom*: $m = c n^2$

The use of the meromorphic function (58) with $\beta = 0.5$ (Fig. 5) and $\beta = 0.9$ (Fig. 6) yields analogous error graphs, but with a slower convergence rate.

Unlike function (58), which is analytic in $[-1, 1]$, function (59) is only $C^2([-1, 1])$, but not $C^3([-1, 1])$. This decreased regularity manifests in the slower decay of the approximation error in Fig. 7. Note that the dependence of the error on the polynomial degree w is displayed in log–log scale, so that the error no longer decreases exponentially with respect to w .

When taking the number of sample points according to the quadratic rule (Fig. 7, right), the error decreases as w^{-3} , and in this range of w the error shows no tendency to blow up for any of the studied values of c .

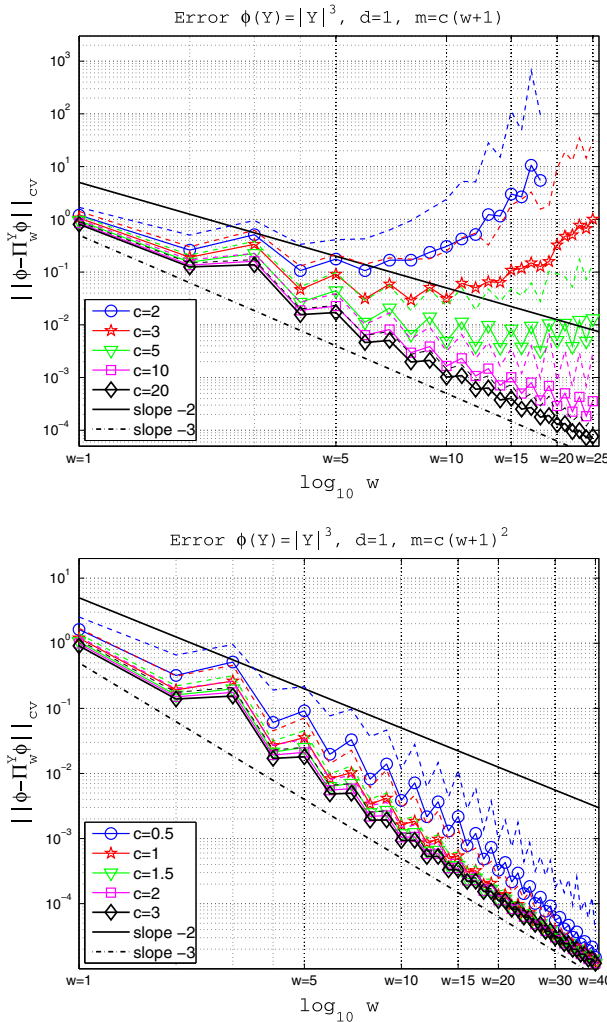


Fig. 7 Error (45) for function (59). *Continuous marked lines*: mean error (46) over $r = 10,000$ repetitions; *dashed lines*: mean (46) plus one standard deviation (47). *Top*: $m = c n$; *bottom*: $m = c n^2$

On the other hand, using the linear rule (Fig. 7, left) yields a deterioration: the critical w , above which the error starts to grow, increases with increasing c .

Note, in particular, that sooner or later the error starts to blow up for all shown constants. This is a clear indication that the linear rule does not lead to a stable and convergent approximation.

From a practical point of view, we are mainly interested in the error as a function of the computational work, not the polynomial degree itself. Figure 8 shows how the error depends on the total number of sampling points when we consider function (58) with $\beta = 0.5$. Note that Fig. 8 shows the same errors as Fig. 5 but with m instead of w on the abscissas.

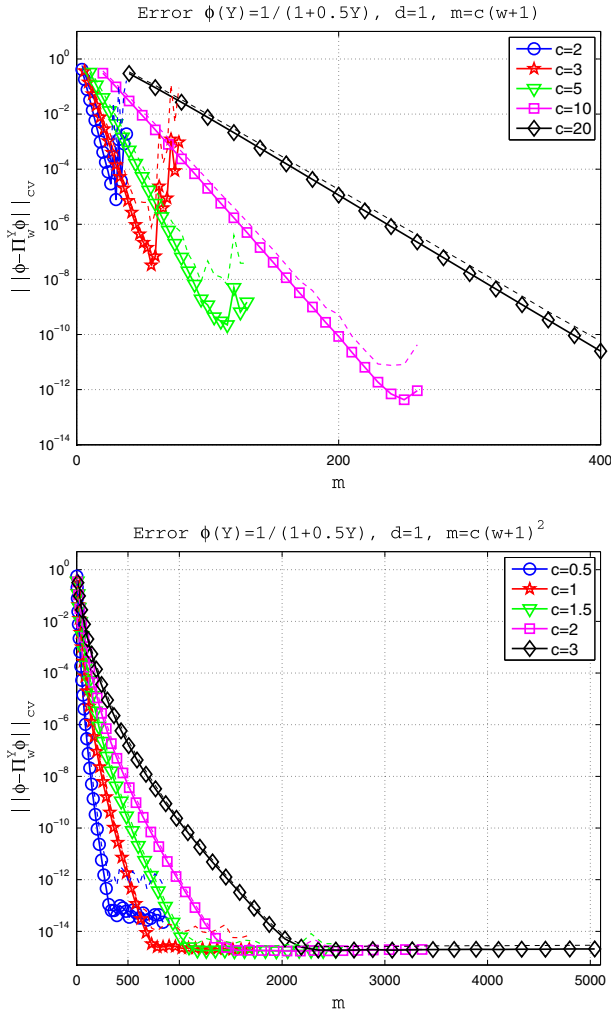


Fig. 8 Dependence of error (45) on number of sample points m . The function is (58) with $\beta = 0.5$. *Continuous marked lines*: mean error (46) over $r = 10,000$ repetitions; *dashed lines*: mean (46) plus one standard deviation (47). *Top*: $m = cn$; *bottom*: $m = cn^2$

In Fig. 8, left we show the linear case: the error decays exponentially with increasing m in an initial phase until the error starts to deteriorate. The convergence is faster for small values of c , but the deterioration also happens earlier, which prevents higher accuracies.

In Fig. 8, right we show the quadratic case. In contrast to the linear case, the convergence becomes subexponential with respect to m . On the other hand, all choices of $c \geq 1$ avoid the deterioration of the errors that we see using the linear rule, and the approximation remains stable and convergent. Figure 9 compares the convergence of the error obtained with the linear and quadratic rules, with respect to m . We remark

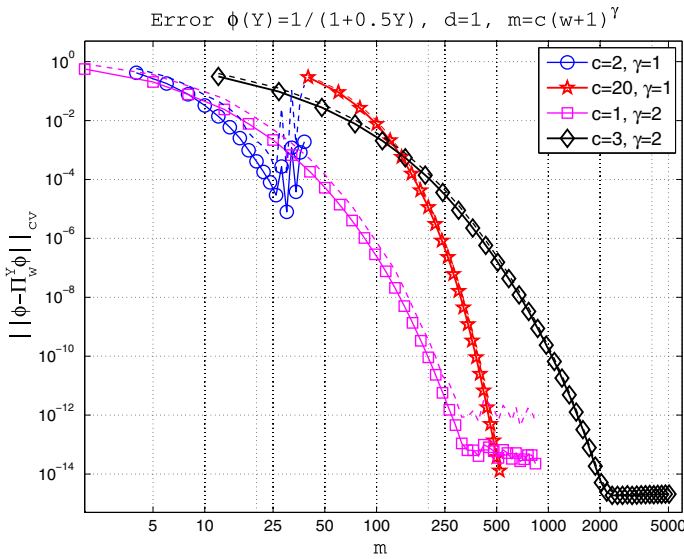


Fig. 9 Dependence of error (45) on number of sample points m . Selected data from Fig. 8 corresponding to rules $m = 2n$, $m = 20n$, $m = 1n^2$, and $m = 3n^2$

that, even though the error deteriorates for high w when using the linear rule with a small c , we can still obtain an accuracy that suffices in many applications.

5.2 Multidimensional case

We now proceed to the multidimensional case, where we have greater freedom to choose the space \mathbb{P}_Λ . We will restrict our examples to isotropic versions of the TP, TD, and HC spaces mentioned earlier. In this section we choose $r = 100$ repetitions to estimate the variability of the error and condition number. In the linear case, the values assumed by c are 1.1, 1.25, 2, 5, 20. In the multidimensional case, a constant c slightly larger than 1 is enough to have a good approximation. This is in contrast to the one-dimensional case, where the linear rule with a constant $c = 2$ features fast growth of the condition number and a large variability of the polynomial approximation.

5.2.1 Condition Number

In Fig. 10, the condition number is compared among the TP, TD, and HC spaces in the case where $d = 2$. We see again an exponential growth of the condition number when m is linearly proportional to n , and the larger the dimension of the space (6)–(8) with respect to w , the more ill-conditioned the problem. As in the one-dimensional case, choosing the number of sample points m like $m \propto n^2$ yields condition numbers that are approximately constant in the studied range of w . Compared to the one-dimensional results of Fig. 3, the two-dimensional results exhibit smaller values and a lower variability. Therefore, the choice of the space does not

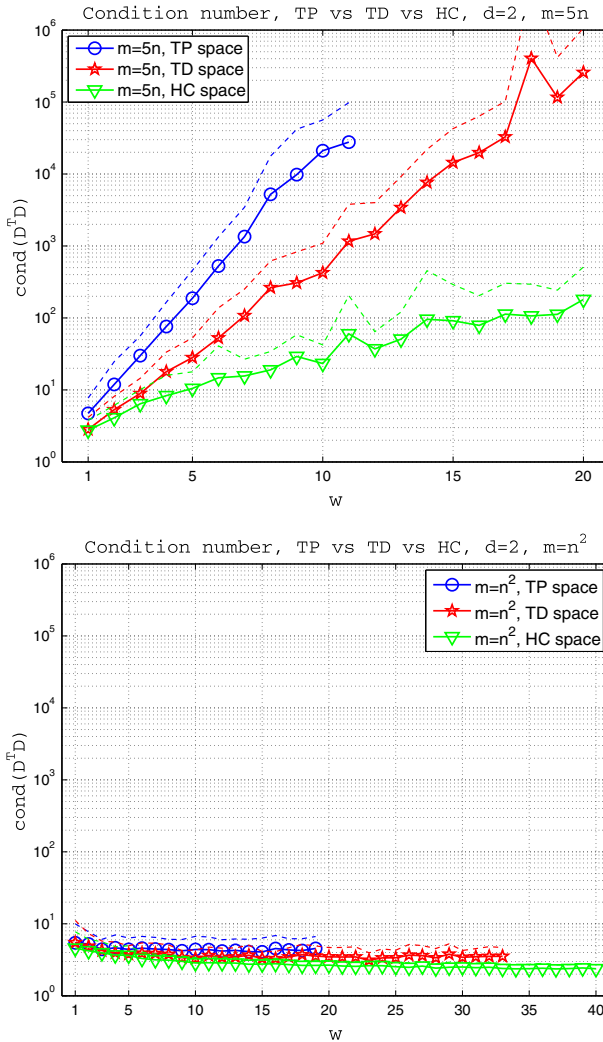


Fig. 10 Condition number (48), TP versus TD versus HC spaces, $d = 2$. *Continuous, marked lines:* mean condition number (49) over $r = 100$ repetitions; *dashed lines:* mean (49) plus one standard deviation (50). *Top:* $m = cn$; *bottom:* $m = cn^2$

seem to play a major role in the behavior of the condition number (48) under the condition $m \propto n^2$.

The situation is similar in higher dimensions; we report, for example, in Fig. 11 the one-dimensional case $d = 1$ compared with the TD space in dimensions $d = 2$ and $d = 4$. We observe in Fig. 11, left, that $m \propto n$ features a lower variability of the condition number with the linear rule, which, however, is still exponentially growing. On the other hand, Fig. 11, right, confirms that the condition $m \propto n^2$ ensures a condition number bounded independently of w , and the higher the dimension d , the more stable the problem.

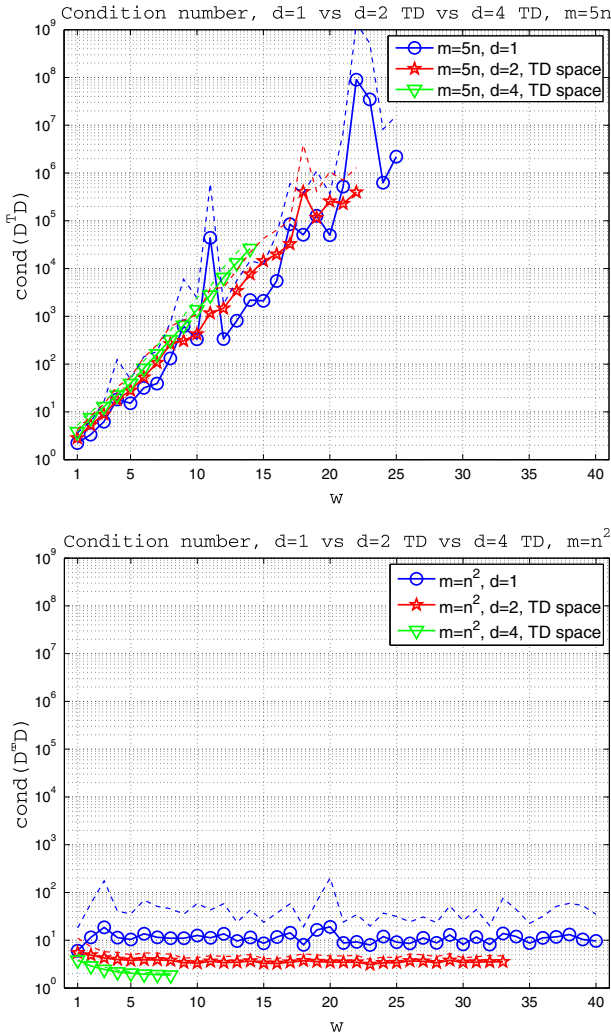


Fig. 11 Condition number (48), $d = 1$ versus $d = 2$ TD space versus $d = 4$ TD space. *Continuous marked lines*: mean condition number (49) over $r = 100$ repetitions; *dashed lines*: mean (49) plus one standard deviation (50). *Top*: $m = cn$; *bottom*: $m = cn^2$

5.2.2 Approximation Error

Let us consider the error in approximating the target function

$$\phi(Y) = \frac{1}{1 + \frac{\beta}{d} \sum_{q=1}^d Y^q}, \quad Y \sim \mathcal{U}([-1, 1]^d), \tag{60}$$

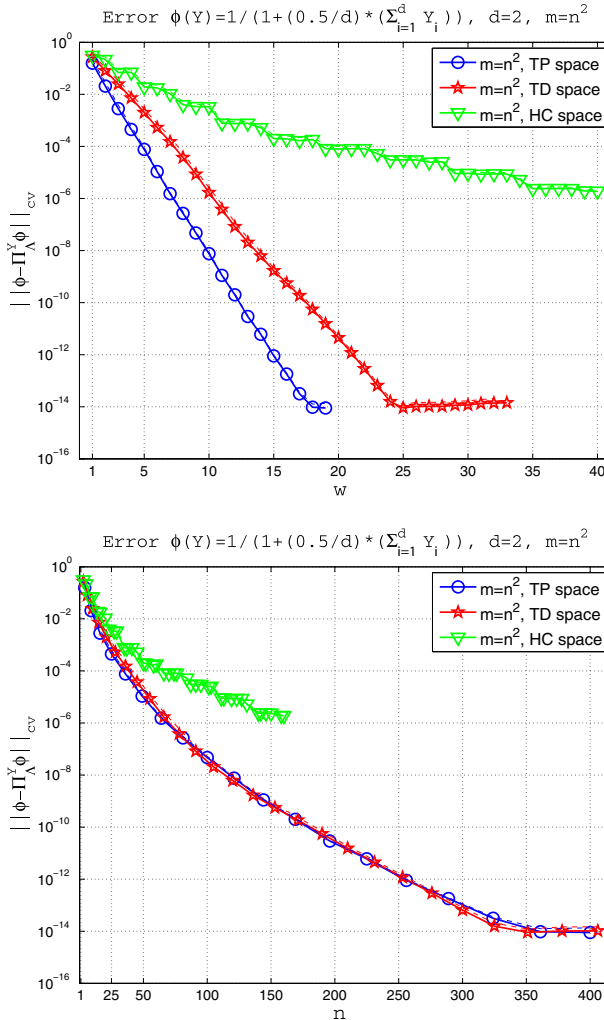


Fig. 12 Error (45) with function (60), $d = 2$, TP versus TD versus HC, $m = cn^2$. Continuous marked lines: mean error (46) over $r = 100$ repetitions; dashed lines: mean (46) plus one standard deviation (47). Top: error versus w ; bottom: error versus n

which is a multidimensional generalization of (58) and inherits its regularity. This function is paradigmatic since it resembles the solution of an elliptic differential equation with the diffusion coefficient parametrized by a Karhunen–Loeve expansion of random variables. We take $\beta = 0.5$ and start by considering the quadratic rule. Figure 12 shows the optimal convergence rates obtained with TP, TD, and HC when approximating function (60) with $d = 2$, $m = cn^2$. The same convergence error plots are reported versus w in Fig. 12, left, and versus n in Fig. 12, right. This shows that the TP and TD spaces with the same dimension introduce similar errors when approximating function (60), while the convergence of the HC space is slower also when looking at the effective dimension of the space n .

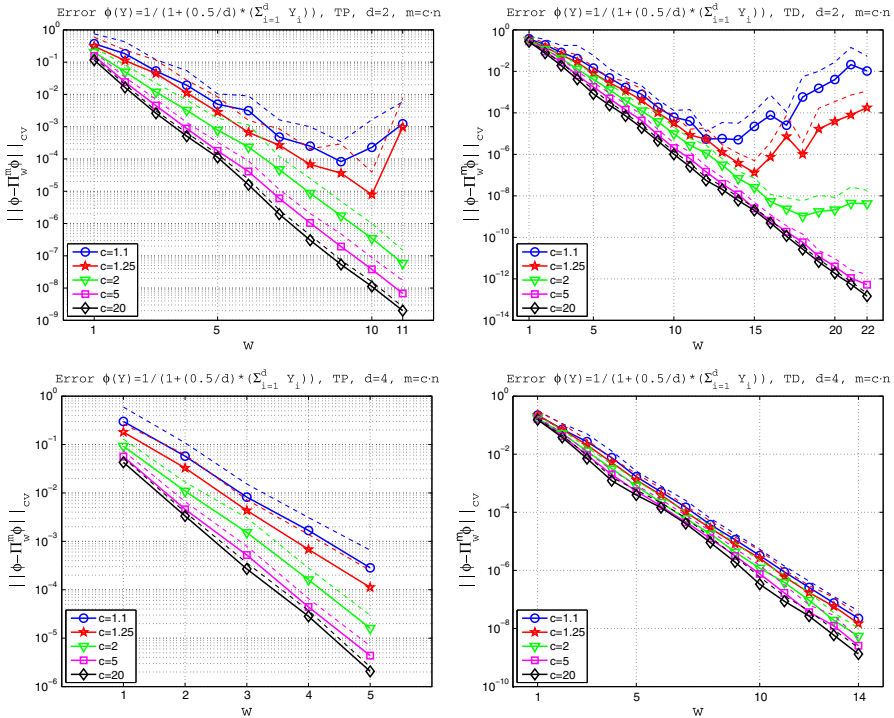


Fig. 13 Error (45) with function (60), $m = cn$. Continuous marked lines: mean error (46) over $r = 100$ repetitions; dashed lines: mean (46) plus one standard deviation (47). Top left $d = 2$, TP space; top right $d = 2$, TD space; bottom left $d = 4$, TP space; bottom right $d = 4$, TD space

Figure 13 shows the error in approximating function (60) in TP or TD spaces for $d = 2$ or $d = 4$ with $m = cn$. Compared to the one-dimensional case $d = 1$ in Fig. 5, we observe a lower variability in the error due to the reduced variability in the corresponding condition number. Despite the reduced variability, in the two-dimensional case $d = 2$, we observe also in this case that the linear rule eventually leads to divergence when w increases if c is chosen too small.

Now we give an example of a function that is hard to approximate in the TD spaces. When $d = 2$, we consider the target function

$$\phi(Y) = \left| \sum_{q=1}^2 Y_q^2 - 0.5 \right|^3, \quad Y \sim \mathcal{U}([-1, 1]^d), \tag{61}$$

which features a discontinuity in its derivatives over a circle with radius equal to $\sqrt{0.5}$ and centered in the origin. Note that (61) is a class C^2 continuous function.

Choosing the quadratic proportionality $m = n^2$ leads to the expected theoretical convergence rates for both TP and TD spaces (Fig. 14).

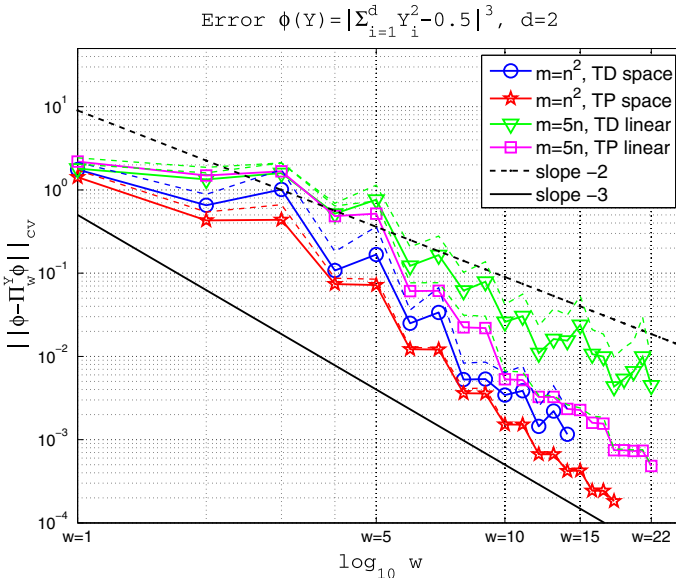


Fig. 14 Error (45) with function (61), $d = 2, m = cn^2$ with TD space versus $m = cn^2$ with TP space versus $m = 5n$ with TD space versus $m = 5n$ with TP space. *Continuous marked lines*: mean error (46) over $r = 100$ repetitions; *dashed lines*: mean (46) plus one standard deviation (47)

When choosing the linear proportionality $m = 5n$ (Fig. 14), the convergence rate is slower than the theoretically predicted one, but again the TP space outperforms the TD space.

6 Conclusions

In this work we have analyzed the problem of approximating a multivariate function by discrete L^2 projection on a polynomial space starting from random, noise-free observations.

In the one-dimensional case with sampling points drawn from a bounded domain and a probability density function bounded away from zero and bounded from above, we showed that the discrete L^2 projection leads to optimal convergence rates, equivalent to the best approximation error in L^∞ , up to a logarithmic factor, *provided the sample size m scales quadratically with the dimension of the polynomial space n* . We also showed how this result reflects on the condition number of the design matrix.

The numerical tests we performed confirm the theoretical results. In our one-dimensional tests, we clearly see that the condition $m \propto n^2$ ensures a condition number of the design matrix bounded independently of the polynomial degree and an optimal convergence rate. On the other hand, the relation $m \propto n$ leads to a condition number growing exponentially fast with the polynomial degree and a convergece plot that features initially a suboptimal rate up to a critical polynomial degree beyond which divergence is observed. In addition, the sensitivity to the proportionality constant was examined.

In high dimensions, we observed numerically that in many cases a choice $m \propto n$ does lead to an optimal convergence rate within all reasonable tolerances (up to machine precision). Whether this is an indication that in high dimensions the relation $m \propto n$ is enough to have a stable and optimal approximation or just that the blow-up of the error occurs at tolerances smaller than machine precision is still an open question and a topic of current research.

In one dimension, similar results could be obtained using a deterministic sample of equispaced points, still ensuring the condition $m \propto n^2$ to guarantee the stability of the projection. In this respect, random sampling is much more attractive in high-dimensional problems.

In this work we considered only functions with values in \mathbb{R} . In the field of UQ, one is often interested in functions with values in some Banach space, representing the solution of a (possibly nonlinear) differential or integral problem. Future research directions will include the extension of these results to Banach-valued functions.

Acknowledgments The authors would like to recognize the support of the PECOS Center at ICES, University of Texas at Austin (Project Number 024550, Center for Predictive Computational Science). Support from the VR project “Effektiva numeriska metoder for stokastiska differentialekvationer med tillampningar” and King Abdullah University of Science and Technology (KAUST) through the AEA projects “Predictability and Uncertainty Quantification for Models of Porous Media” and “Tracking Uncertainties in Computational Modeling of Reactive Systems” is also acknowledged. R. Tempone is a member of the KAUST SRI Center for Uncertainty Quantification. The first and second authors were supported by the Italian Grant FIRB-IDEAS (Project RBID08223Z) “Advanced numerical techniques for uncertainty quantification in engineering and life science problems.” We are indebted to A. Cohen and R. DeVore for giving us valuable feedback on the convergence proof. We would also like to thank the anonymous referees for their useful comments that helped us to improve considerably the manuscript, and in particular for the suggestion in the proof of Theorem 2.

References

1. M. Ainsworth and J.T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.
2. I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Rev.*, 52(2):317–355, 2010.
3. J. Bäck, F. Nobile, L. Tamellini, and R. Tempone. Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In J.S. Hesthaven and E.M. Ronquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76, pages 43–62. Springer, Berlin, 2011.
4. A.M. Bagirov, C. Clausen, and M. Kohler. Estimation of a regression function by maxima of minima of linear functions. *IEEE Trans. Inform. Theory*, 55(2):833–845, 2009.
5. W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.
6. J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Mod. Methods, Appl. Sci. (M3AS)*, 22(9):1250023–1, 1250023–33, 2012.
7. P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Universal algorithms for learning theory. II. Piecewise polynomial functions. *Constr. Approx.*, 26(2):127–152, 2007.
8. P. Binev, A. Cohen, W. Dahmen, R. DeVore, and Vladimir Temlyakov. Universal algorithms for learning theory. I. Piecewise constant functions. *J. Mach. Learn. Res.*, 6:1297–1321, 2005.
9. G. Blatman and B. Sudret. Sparse Polynomial Chaos expansions and adaptive stochastic finite elements using regression approach. *C.R.Mechanique*, 336:518–523, 2008.

10. G. Blatman and B. Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, 2011.
11. E.J. Candès, J.K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
12. C. Canuto and A. Quarteroni. Approximation results for orthogonal polynomials in Sobolev spaces. *Math. Comp.*, 38(157):67–86, 1982.
13. A. Cohen, M. Davenport, and D. Leviatan. On the stability and accuracy of least squares approximations. *Found. Comput. Math.*, 13:819–834, 2013.
14. A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best N -term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.
15. A. Cohen, R. DeVore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.
16. T. Crestaux, O. Le Maître, and J.M. Martinez. Polynomial Chaos Expansion for Sensitivity Analysis. *Reliability Engineering and System Safety*, 94(7):1161–1172, 2009.
17. H. A. David and H. N. Nagaraja. *Order statistics*. Wiley-Interscience, Hoboken, Third edition, 2003.
18. B. Debusschere, H. Najm, P. Pebray, O. Knio, R. Ghanem, and O. Le Maître. Numerical Challenges in the Use of Polynomial Chaos Representations for Stochastic Processes. *SIAM J. Sci. Comput.*, 26(2):698–719, 2005.
19. D.L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
20. M.S. Eldred. Recent Advances in Non-Intrusive Polynomial Chaos and Stochastic Collocation Methods for Uncertainty Analysis and Design. In *Proceedings of the AIAA 50th conference on Structural Dynamics, and Materials 2009, Palm Springs, California, American Institute of Aeronautics and Astronautics, Reston, VA*, 2009.
21. M.S. Eldred. Design under uncertainty employing stochastic expansion methods. *International Journal for Uncertainty Quantification*, 1:119–146, 2011.
22. O.G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM Math. Model. Numer. Anal.*, 46:317–339, 2012.
23. G.R. Ghanem and P.D. Spanos. *Stochastic Finite Elements: a spectral approach*. Dover, Mineola, 2003.
24. A. Gut. *Probability: A Graduate Course*, volume 75 of *Springer texts in statistics*. Springer, Berlin, 2005.
25. L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer, Berlin, first edition, 2002.
26. S. Hosder, R. W. Walters, and M. Balch. Point-Collocation Nonintrusive Polynomial Chaos Method for Stochastic Computational Fluid Dynamics. *AIAA Journal*, 48:2721–2730, 2010.
27. M. Kohler. Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *J. Statist. Plann. Inference*, 89(1–2):1–23, 2000.
28. M. Kohler. Nonlinear orthogonal series estimates for random design regression. *J. Statist. Plann. Inference*, 115(2):491–520, 2003.
29. O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, New York, 2010.
30. G. Migliorati. *Polynomial approximation by means of the random discrete L^2 projection and application to inverse problems for PDEs with stochastic data*. PhD thesis, Dipartimento di Matematica “Francesco Brioschi”, Politecnico di Milano and Centre de Mathématiques Appliquées, École Polytechnique, 2013.
31. G. Migliorati, F. Nobile, E. von Schwerin, and R. Tempone. Approximation of Quantities of Interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces. *SIAM J. Sci. Comput.*, 35(3):A1440–A1460, 2013.
32. F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.
33. W. Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976.
34. B. Sudret. Global sensitivity analysis using polynomial chaos expansion. *Reliability Engineering and System Safety*, 93:964–979, 2008.
35. D. Xiu. Fast numerical methods for stochastic computations: a review. *Commun. Comput. Phys.*, 5(2–4):242–272, 2009.
36. D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644 (electronic), 2002.

