Fast approximation by periodic kernel-based lattice-point interpolation with application in uncertainty quantification

Vesa Kaarnioja[1] · Yoshihito Kazashi[2] · Frances Y. Kuo[1] · Fabio Nobile[2] · Ian H. Sloan[1]

Abstract

This paper deals with the kernel-based approximation of a multivariate periodic function by interpolation at the points of an integration lattice—a setting that, as pointed out by Zeng et al. (Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, New York, 2006) and Zeng et al. (Constr. Approx. 30: 529–555, 2009), allows fast evaluation by fast Fourier transform, so avoiding the need for a linear solver. The main contribution of the paper is the application to the approximation problem for uncertainty quantification of elliptic partial differential equations, with the diffusion coefficient given by a random field that is periodic in the stochastic variables, in the model proposed recently by Kaarnioja et al. (SIAM J Numer Anal 58(2): 1068–1091, 2020). The paper gives a full error analysis, and full details of the construction of lattices needed to ensure a good (but inevitably not optimal) rate of convergence and an error bound independent of dimension. Numerical experiments support the theory.

Mathematics Subject Classification 41A15 · 41A63 · 65D07 · 65D15 · 65T40

✉ Yoshihito Kazashi
  yoshihito.kazashi@epfl.ch

  Vesa Kaarnioja
  vesa.kaarnioja@iki.fi

  Frances Y. Kuo
  f.kuo@unsw.edu.au

  Fabio Nobile
  fabio.nobile@epfl.ch

  Ian H. Sloan
  i.sloan@unsw.edu.au

[1] School of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia

[2] CSQI, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

🖄 Springer

1 Introduction

We consider a *kernel-based approximation* for a multivariate *periodic* function by *interpolation* at a quasi-Monte Carlo *lattice* point set. Kernel-based interpolation methods are by now well established (see, e.g., [26] and more discussion below). It is the unique combination of *a periodic kernel plus a lattice point set* here that will deliver us the significant advantage in computational efficiency. As already advocated by Hickernell and colleagues in [28,29], the combination of a periodic reproducing kernel with the group structure of lattice points means that the linear system for constructing the kernel interpolant involves a circulant matrix, thus can be solved very efficiently using the fast Fourier transform. So, our kernel method can be fast even if the dimensionality is high.

As also advocated in [28,29], a kernel interpolant is in many settings optimal among all approximation algorithms that use the same function values (see also known results on optimal recovery, e.g., [20,21]). We can therefore analyze the worst case approximation error of our kernel method by using, as upper bound, the worst case error of an auxiliary algorithm based on a Fourier series truncated at a hyperbolic cross index set. Using recent works [3,4], we here construct a lattice generating vector with a guaranteed good error bound for our kernel interpolant. Note, importantly, that *neither the construction of our lattice generating vector, nor the implementation of our kernel method, requires explicit knowledge or evaluation of the auxiliary hyperbolic cross index set*. In short, we know how to find a good lattice point set so that our kernel method has a small error in addition to being of low cost.

In this paper, the main contribution is to apply and analyze this periodic-kernel-plus-lattice method to uncertainty quantification of elliptic partial differential equations (PDEs), where the diffusion coefficient is given by a random field that is *periodic in the stochastic variables*, as in the model proposed recently by Kaarnioja et al. [12]. We tailor our lattice generating vector to the regularity of the PDE solution with respect to the stochastic variables. Our numerical results beat the theoretical predictions, indicating that the theory based on worst case analysis may not be sharp.

The kernel approximation developed here may have a role as a surrogate model for complicated forward problems. One popular use for surrogate models is to allow efficient sampling of the original system. If the solution of some particularly difficult PDE problem with high accuracy takes a week for a given parameter choice $\boldsymbol{y}$, then having a kernel interpolant that can be evaluated in hours or minutes could be very useful. A second possible use for the kernel interpolant is in the easy generation of derivatives, needed for example in gradient-based optimization algorithms. The surrogate might be even more useful for Bayesian inverse problems.

We now elaborate key points.

Periodic-kernel-plus-lattice method. Let $f(\boldsymbol{y}) = f(y_1, \ldots, y_s)$ be a real-valued function on the s-dimensional unit cube $[0, 1]^s$, with a somewhat smooth 1-periodic extension to $\mathbb{R}^s$. Our main interest is in problems where the dimension s is large. Following [25], we assume that f has an absolutely convergent Fourier series, and belongs to a weighted mixed Sobolev space $H := H_{s,\alpha,\boldsymbol{\gamma}}$ which is characterized by a

smoothness parameter $\alpha > 1$ and a family of positive numbers $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subset \mathbb{N}}$ called *weights*; the details are given in Sect. 2.1.

Our ultimate application, to be analyzed in Sect. 4, concerns a class of elliptic PDEs parameterized by a very high (or possibly countably infinite) number of stochastic parameters, for which the solution, as a function of the parameters, is periodic and belongs to the weighted space H for a suitable choice of α and $\boldsymbol{\gamma}$.

The important feature of the space H is that it is a reproducing kernel Hilbert space (RKHS), with a simple reproducing kernel $K(\boldsymbol{y}, \boldsymbol{y}')$. This opens the way to the use of kernel methods to approximate functions in H from given point values. In particular, in this paper we focus on the *kernel interpolation*: given $f \in H$ and a suitable set of points $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in [0, 1]^s$, we seek for an approximation $f_n \in H$ of the form

$$f_n(\boldsymbol{y}) := \sum_{k=1}^{n} a_k \, K(\boldsymbol{t}_k, \boldsymbol{y}), \qquad \boldsymbol{y} \in [0, 1)^s, \tag{1}$$

which satisfies the interpolation condition $f_n(\boldsymbol{t}_k) = f(\boldsymbol{t}_k)$, $k = 1, \ldots, n$. We refer to f_n as the *kernel interpolant* of f.

We will interpolate the function at a set of n *lattice points* specified by a *generating vector* $\boldsymbol{z} \in \mathbb{Z}^s$. The points are then given by the formula $\boldsymbol{t}_k = (k\boldsymbol{z} \bmod n)/n$, $k = 1, \ldots, n$, with $\boldsymbol{t}_n = \boldsymbol{t}_0 = \boldsymbol{0}$. A lattice point set has an additive group structure, implying that the difference of two lattice points is another lattice point (after taking into account periodicity).

A key property of our reproducing kernel is that it depends only on the difference of the two arguments, thus $K(\boldsymbol{y}, \boldsymbol{y}') = K(\boldsymbol{y} - \boldsymbol{y}', \boldsymbol{0})$, and $K(\cdot, \boldsymbol{0})$ is a periodic function with an easily computable expression when α is an even integer. Combining this with the group structure of lattice points means that the matrix $[K(\boldsymbol{t}_k - \boldsymbol{t}_{k'}, \boldsymbol{0})]_{k,k'=1,\ldots,n}$ contains only n distinct values and indeed is a circulant matrix. Therefore the linear system arising from collocating (1) at the points $\boldsymbol{t}_{k'}$, $k' = 1, \ldots, n$, can be solved using the fast Fourier transform with a cost of $\mathcal{O}(n \log(n))$.

Once we have the coefficients a_k, we can use (1) to evaluate the interpolant f_n at L arbitrary points $\boldsymbol{y}_\ell$, $\ell = 1, \ldots, L$, with a cost of $\mathcal{O}(Ln)$. Remarkably, with almost the same cost we can evaluate f_n at all the Ln points of the *union of shifted lattices* $\boldsymbol{y}_\ell + \boldsymbol{t}_{k'}$, $\ell = 1, \ldots, L$, $k' = 1, \ldots, n$. Indeed, since $K(\boldsymbol{t}_k, \boldsymbol{y}_\ell + \boldsymbol{t}_{k'}) = K(\boldsymbol{t}_k - \boldsymbol{t}_{k'}, \boldsymbol{y}_\ell)$ and the matrix $[K(\boldsymbol{t}_k - \boldsymbol{t}_{k'}, \boldsymbol{y}_\ell)]_{k,k'=1,\ldots,n}$ is circulant, we have

$$f_n(\boldsymbol{y}_\ell + \boldsymbol{t}_{k'}) = \sum_{k=1}^{n} a_k \, K(\boldsymbol{t}_k - \boldsymbol{t}_{k'}, \boldsymbol{y}_\ell),$$

which can be evaluated for each $\boldsymbol{y}_\ell$ for all $\boldsymbol{t}_{k'}$ together by fast Fourier transform with a cost of $\mathcal{O}(n \log(n))$, leading to the total cost of $\mathcal{O}(Ln \log(n))$. Comprehensive cost analysis taking into account also the evaluations of f and K is given in Sect. 5.

Brief survey on kernel methods in high dimensions. Griebel and Rieger [10] considered a (non-interpolatory) kernel approximation based on a regularized reconstruction technique from machine learning for a class of parameterized elliptic PDEs similar to

the one considered in this work, yet with non-periodic dependence on the parameters. They used an anisotropic kernel, behaving differently in different variables, to address the high dimensionality of the problem. However, their error estimate was in terms of the mesh norm or fill distance of the point set, which is the Euclidean radius of the largest Euclidean ball that contains no points in its interior. Since the fill distance behaves at best like $n^{-1/s}$, where n is the number of sampling points, their estimates inevitably suffer the *curse of dimensionality*.

Kempf et al. [15] considered the same PDE problem and anisotropic kernel as [10]. However, they considered a penalized least-squares approach for kernel approximation and an isotropic sparse grid as point set, which allowed them to obtain error estimates with a mitigated (but still present) curse of dimensionality.

As noted above, lattice points have already been used in a kernel interpolation method. Zeng et al. [28] seem to be the first to work in this direction, however the question of dependence on dimension was not considered in their analysis. Zeng et al. [29] established dimension independent error estimates in weighted spaces in the case of product weights (i.e., weights that have the form $\gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_j$). We note, however, that the assumption of product weights is rather limiting. For instance, for integration problems involving parameterized PDEs, the best convergence rates known up to now are obtained by considering weighted space for the parameter-to-solution map with (S)POD weights [9,12,17], whereas weighted spaces with product weights lead to the best known rates only for special models [7,11,14]. In this paper we extend these results to the case of kernel approximation (as opposed to integration) of the parameter-to-solution map, and we are able to show dimension-independent convergence rates using (S)POD weights in the general case. To the best of our knowledge, this is the first paper to use non-product weights for approximation in parameterized PDE problems.

PDEs with periodic dependence on random variables. Our motivating application is a class of parameterized elliptic PDEs with periodic dependence on the parameters, for which we will establish dimension independent error estimate for the kernel interpolant, by deriving suitable choices of smoothness parameter and weights for the problem at hand. To the best of our knowledge, this is the first paper presenting dimension independent kernel approximation methods using lattice points for this class of problems.

We consider uncertainty quantification for an elliptic PDE (see details in Sect. 4) on a physical domain $D \subset \mathbb{R}^d$, $d = 1, 2$ or 3, in a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, with an input random field of the form

$$a(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \sum_{j \geq 1} \Theta_j(\omega) \, \psi_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in D, \ \omega \in \Omega,$$

where a_0 and ψ_j are uniformly bounded in D, and $\Theta_j(\omega)$ are i.i.d. random variables following a prescribed distribution. In the popular affine model, Θ_j are i.i.d. random variables uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. In the periodic model [12], Θ_j are i.i.d. random variables distributed according to the arcsine distribution and can be parameterized as

$$\Theta_j = \frac{1}{\sqrt{6}} \sin(2\pi y_j), \qquad j \geq 1,$$

with y_j uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. The mean of the random field is a_0, and the scaling $1/\sqrt{6}$ is chosen here so that the covariance of the random field is also exactly the same as in the affine case. Higher moments are of course somewhat different, but as argued in [12], there seems to be no clear reason for preferring one over the other.

Due to periodicity, it is equivalent to work with y_j uniformly distributed in the interval $[0, 1]$ instead of $[-\frac{1}{2}, \frac{1}{2}]$, thus from now on we consider the parameter space

$$\boldsymbol{y} \in U := [0, 1]^{\mathbb{N}}.$$

In the earlier paper [12] the aim was to develop and analyze a method for computing the expected value of a given quantity of interest, expressed as a linear functional of the PDE solution, hence facing a high dimensional integration problem. Here, in contrast, the aim is to develop and analyze a fast method for approximating the solution $u(\boldsymbol{x}, \boldsymbol{y})$, or some quantity of interest $Q(\boldsymbol{y})$ derived from $u(\boldsymbol{x}, \boldsymbol{y})$, as an explicit function of $\boldsymbol{y}$. To that end we will develop a kernel-based approximation, using the kernel of a reproducing kernel Hilbert space of periodic functions, and interpolation at a lattice point set.

Structure of the paper. In Sect. 2 we define the function space setting and the kernel interpolant, and establish its principal properties, while giving a simple proof of a known optimality result, namely that in the sense of worst case error the kernel interpolant is an optimal L_p approximation among all approximations that use the same information about the target function $f \in H$. Then in Sect. 3 we establish upper and lower bounds on the error. For the upper bound we use the optimality result together with the error analysis for a trigonometric polynomial method established by two of the current authors together with Cools and Nuyens [3,4]. For the lower bound we provide another proof of a recent result by Byrenheid et al. [1], namely that a method that draws information only from function values at lattice points inevitably has a rate of convergence that is at best only half of the best possible rate, thereby obtaining matching upper and lower bounds up to logarithmic factors. In Sect. 4, we apply the error analysis developed in Sect. 3 to a parameterized PDE problem, thereby obtaining rigorous upper error bounds that are independent of dimension and have explicit rates of convergence. Section 5 is concerned with the cost analysis of our proposed method. In Sect. 6 we give the results of some numerical experiments.

2 The kernel interpolant

2.1 The function space setting

Let $f(\boldsymbol{y}) = f(y_1, \ldots, y_s)$ be a real-valued function on $[0, 1]^s$ with a somewhat smooth 1-periodic extension to $\mathbb{R}^s$ with respect to each variable y_j. Our main interest

is in problems where the dimension s is large. Following [25], we assume that f has absolutely convergent Fourier series (and so is continuous),

$$f(\boldsymbol{y}) = \sum_{\boldsymbol{h} \in \mathbb{Z}^s} \widehat{f}(\boldsymbol{h}) \, e^{2\pi i \boldsymbol{h} \cdot \boldsymbol{y}}, \quad \text{with} \quad \widehat{f}(\boldsymbol{h}) := \int_{[0,1]^s} f(\boldsymbol{y}) \, e^{-2\beta i \boldsymbol{h} \cdot \boldsymbol{y}} \, d\boldsymbol{y};$$

and moreover belongs to a weighted mixed Sobolev space $H := H_{s,\alpha,\boldsymbol{\gamma}}$, a Hilbert space with inner product and norm

$$\langle f, g \rangle_H := \langle f, g \rangle_{s,\alpha,\boldsymbol{\gamma}} := \sum_{\boldsymbol{h} \in \mathbb{Z}^s} r(\boldsymbol{h}) \, \widehat{f}(\boldsymbol{h}) \, \overline{\widehat{g}(\boldsymbol{h})},$$

$$\|f\|_H := \|f\|_{s,\alpha,\boldsymbol{\gamma}} := \left(\sum_{\boldsymbol{h} \in \mathbb{Z}^s} r(\boldsymbol{h}) \, |\widehat{f}(\boldsymbol{h})|^2 \right)^{1/2},$$

where

$$r(\boldsymbol{h}) := r_{s,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}) := \frac{1}{\gamma_{\mathrm{supp}(\boldsymbol{h})}} \prod_{j \in \mathrm{supp}(\boldsymbol{h})} |h_j|^{\alpha},$$

with $\mathrm{supp}(\boldsymbol{h}) := \{ j \in \{1 : s\} : h_j \neq 0 \}$ and $\{1 : s\} := \{1, 2, \ldots, s\}$, and with the $\boldsymbol{h} = \boldsymbol{0}$ term in the sum to be interpreted as $\gamma_{\emptyset}^{-1} |\widehat{f}(\boldsymbol{0})|^2$. The weighted space $H_{s,\alpha,\boldsymbol{\gamma}}$ is characterized by the smoothness parameter $\alpha > 1$ and a family of positive numbers $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subset \mathbb{N}}$ called *weights*, where a positive weight $\gamma_{\mathfrak{u}}$ is associated with each subset $\mathfrak{u} \subseteq \{1 : s\}$. We fix the scaling of the weights by setting $\gamma_{\emptyset} := 1$, so that the norm of a constant function in H matches its L_2 norm.

It can easily be verified that if α is an even integer then the norm can be rewritten as the norm in an "unanchored" weighted Sobolev space of dominating mixed smoothness of order $\alpha/2$,

$$\|f\|_H = \sqrt{ \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{(2\pi)^{\alpha|\mathfrak{u}|} \gamma_{\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \left| \int_{[0,1]^{s-|\mathfrak{u}|}} \left(\prod_{j \in \mathfrak{u}} \frac{\partial^{\alpha/2}}{\partial y_j^{\alpha/2}} \right) f(\boldsymbol{y}) \, d\boldsymbol{y}_{-\mathfrak{u}} \right|^2 d\boldsymbol{y}_{\mathfrak{u}} },$$

(2)

where $\boldsymbol{y}_{\mathfrak{u}}$ denotes the components of $\boldsymbol{y}$ with indices that belong to the subset $\mathfrak{u}$, and $\boldsymbol{y}_{-\mathfrak{u}}$ denotes the components that do not belong to $\mathfrak{u}$, and $|\mathfrak{u}|$ denotes the cardinality of $\mathfrak{u}$.

The important feature of the space H is that it is an RKHS, with an explicitly known and analytically simple reproducing kernel, namely

$$K(\boldsymbol{y}, \boldsymbol{y}') := K_{s,\alpha,\boldsymbol{\gamma}}(\boldsymbol{y}, \boldsymbol{y}') := \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}} \prod_{j \in \mathfrak{u}} \eta_\alpha(y_j, y_j'),$$

where

$$\eta_\alpha(y, y') := \eta_\alpha(y - y') := \sum_{h \neq 0} \frac{e^{2\pi i h (y - y')}}{|h|^\alpha} = \sum_{h \neq 0} \frac{\cos 2\pi h (y - y')}{|h|^\alpha}.$$

Note that the *reproducing property*

$$\langle f, K(\cdot, y) \rangle_H = f(y) \quad \text{for all } f \in H \text{ and all } y \in [0, 1]^s, \tag{3}$$

is easily verified.

Of special interest are even integer values of α, because, when α is even, η_α can be expressed in the especially simple closed form

$$\eta_\alpha(y, y') = \frac{(2\pi)^\alpha}{(-1)^{\alpha/2+1}\alpha!} B_\alpha(\{y - y'\}), \quad y, y' \in [0, 1],$$

where the braces indicate that $y - y'$ is to be replaced by its fractional part in $[0, 1)$, and $B_\alpha(y)$ is the Bernoulli polynomial of degree α. For example, for $\alpha = 2$ and $\alpha = 4$ we have

$$B_2(y) = y^2 - y + \frac{1}{6} \quad \text{and} \quad B_4(y) = y^4 - 2y^3 + y^2 - \frac{1}{30}.$$

2.2 The kernel interpolant

We are interested in approximating a given function $f \in H$ by an approximation of the form

$$A_n^*(f) := f_n(y) := f_{s,\alpha,\gamma,n,z}(y) := \sum_{k=1}^{n} a_k K\left(\left\{\frac{kz}{n}\right\}, y\right), \quad y \in [0, 1)^s, \tag{4}$$

where $z \in \{1, \ldots, n-1\}^s$, and the braces around the vector of length s indicate that each component of the vector is to be replaced by its fractional part. The points

$$t_k := \left\{\frac{kz}{n}\right\} \quad \text{for} \quad k = 1, \ldots, n \tag{5}$$

are the points of a lattice cubature rule of rank 1, see [24]. In what follows, we omit these braces because functions we consider are, unless otherwise stated, periodic.

In particular, we define $f_n \in H$ to be the function of the form (4) that *interpolates* f at the lattice points,

$$f_n(t_k) = f(t_k) \quad \text{for all} \quad k = 1, \ldots, n, \tag{6}$$

and refer to f_n as the *kernel interpolant* of f.

The coefficients a_k in (4) are given by the linear system based on (6)

$$\sum_{k=1}^{n} \mathcal{K}_{k,k'} \, a_k \; = \; f\left(\boldsymbol{t}_{k'}\right) \qquad \text{for all} \qquad k' = 1, \dots, n, \tag{7}$$

where $\mathcal{K}_{k',k} = \mathcal{K}_{k,k'} := K(\boldsymbol{t}_k, \boldsymbol{t}_{k'}), k, k' = 1, \dots, n$. Note that the matrix elements can be expressed, using periodicity, as

$$\mathcal{K}_{k,k'} \; = \; K\left(\frac{(k - k')z}{n}, \boldsymbol{0}\right),$$

where $\boldsymbol{0}$ is the s-vector of all zeroes. It follows that the $n \times n$ matrix $\mathcal{K}$ is a *circulant matrix*, which contains only n distinct elements, and can be diagonalised in a time of order $n \log n$ by fast Fourier transform. This is a major motivation for using lattice points.

2.3 The kernel interpolant is the minimal norm interpolant

The following property is a well known result for interpolation in a reproducing kernel Hilbert space; for completeness we give a proof.

Theorem 2.1 *The kernel interpolant f_n defined by (4), (5) and (6) is the minimal norm interpolant in H.*

Proof Denoting the linear span of the kernels with one leg at $\boldsymbol{t}_k, k = 1, \dots, n$ by

$$P_n \; := \; \mathrm{span}\{K(\boldsymbol{t}_k, \cdot) : k = 1, \dots, n\},$$

we observe the well known fact (see, e.g., [5,8]), that f_n is the orthogonal projection of f on P_n with respect to the inner product $\langle \cdot, \cdot \rangle_H$, since from the reproducing property (3) and the interpolation property (6) we have

$$\langle f - f_n, K(\boldsymbol{t}_k, \cdot) \rangle_H \; = \; f(\boldsymbol{t}_k) - f_n(\boldsymbol{t}_k) \; = \; 0 \qquad \text{for all} \qquad k = 1, \dots, n.$$

In turn, there follows the Pythagoras theorem,

$$\|f\|_H^2 \; = \; \|f - f_n\|_H^2 + \|f_n\|_H^2, \tag{8}$$

and the *minimal norm* property of f_n,

$$f_n \; = \; \mathrm{argmin}\big\{\|g\|_H \; : \; g \in H \text{ and } g(\boldsymbol{t}_k) = f(\boldsymbol{t}_k) \text{ for all } k = 1, \dots, n\big\},$$

since if g is any other interpolant of f at the lattice points then

$$\langle g - f_n, f_n \rangle_H \; = \; \sum_{k=1}^{n} a_k \, \langle g - f_n, K(\boldsymbol{t}_k, \cdot) \rangle_H \; = \; 0,$$

and hence $\|g\|_H^2 = \|g - f_n\|_H^2 + \|f_n\|_H^2$, from which the uniqueness of the minimal norm interpolant also follows. □

2.4 The kernel interpolant is optimal for given function values

In this subsection we show that the kernel interpolant f_n defined by (4), (5) and (6) is optimal among all approximations that use only the same function values of f, in the sense of giving the least possible worst case error measured in any given norm $\|\cdot\|_W$ such that $H \subset W$ for functions in H. This is a special case of a general result for optimal recovery problems in Hilbert spaces (see for example [20, Example 1.1] and [21, Section 3]), but for completeness we give a short proof here. Our proof follows the exposition of [26, Proof of Theorem 13.5], but suitably adapted to our setting.

Let $A_n : H \to H$ be an algorithm (linear or non-linear) that uses as information about the argument only its values at the points (5), i.e., it is a mapping of the form $A_n(f) = \mathcal{I}_n(f(t_1), \ldots, f(t_n))$ for a mapping $\mathcal{I}_n : \mathbb{R}^n \to H$. The worst case W-error for this algorithm is defined by

$$e^{\mathrm{wor}}(A_n; W) := \sup_{f \in H, \|f\|_H \leq 1} \|f - A_n(f)\|_W.$$

Theorem 2.2 *Let $A_n : H \to H$ be an algorithm* (*linear or non-linear*) *such that $A_n(f)$ uses as information about f only its values $f(t_1), \ldots, f(t_n)$ at the points* (5). *For $f \in H$, let $A_n^*(f) := f_n$ be the kernel interpolant defined by* (4), (5) *and* (6). *Then, for any normed space $W \supset H$ we have*

$$e^{\mathrm{wor}}(A_n^*; W) \leq e^{\mathrm{wor}}(A_n; W).$$

Proof Define $\mathcal{C} := \{g \in H : \|g\|_H \leq 1 \text{ and } g(t_k) = 0 \text{ for all } k = 1, \ldots, n\}$. For any $g \in \mathcal{C}$ we have

$$
\begin{aligned}
\|g\|_W &\leq \frac{1}{2}(\|g - A_n(0)\|_W + \|g + A_n(0)\|_W) \\
&\leq \max\left(\|g - A_n(0)\|_W, \|g + A_n(0)\|_W\right) \\
&= \max\left(\|g - A_n(g)\|_W, \|(-g) - A_n(-g)\|_W\right) \leq e^{\mathrm{wor}}(A_n; W), \quad (9)
\end{aligned}
$$

where in the penultimate step we used $g(t_k) = 0$ for all $k = 1, \ldots, n$, from which it follows that $A_n(0) = A_n(g) = A_n(-g)$. For any $f \in H$ such that $\|f\|_H \leq 1$, since f_n is interpolatory, the Pythagoras theorem (8) implies $\|f - f_n\|_H \leq 1$, and hence $f - f_n \in \mathcal{C}$. Thus it follows from (9) that

$$\|f - A_n^*(f)\|_W \ = \ \|f - f_n\|_W \ \leq \ e^{\text{wor}}(A_n; W).$$

The theorem now follows.   □

In the above result, we may, for example, take $W = L_p$ for any $1 \leq p \leq \infty$.

3 Lower and upper error bounds

3.1 Lower bound on the worst case L_p error ($1 \leq p \leq \infty$)

A recent paper [1] showed (with a different definition of the parameter α) that the worst case L_2 error for an approximation that uses the points of a rank-1 lattice cannot have an order of convergence better than $n^{-\alpha/4}$ (with our definition of α). Bearing in mind that H is a (Hilbert) space of functions of dominating mixed smoothness of order $\alpha/2$, this is just half the rate $n^{-\alpha/2}$ of the best approximation. Since the function space setting in that paper is rather different from ours (here we use a Fourier description and a so-called unanchored space, and have introduced weights) we briefly reprove the main result here, obtaining a sharp lower bound expressed in terms of the weights. Furthermore, in our setting we make the result stronger by showing that the same lower bound holds for the worse case L_1 error.

Theorem 3.1 *Let $s \geq 2$. Assume that the weights for the subsets of $\{1 : s\}$ containing a single element satisfy $\gamma_{\{j\}} > 0$ for all $j \in \{1 : s\}$, and that $z \in \{0, \ldots, n-1\}^s$ is given. Let $A_n : H \to L_p$ be an algorithm (linear or non-linear) that uses information only at the lattice points (5) and satisfies $A_n(0) = 0$. Then for $1 \leq p \leq \infty$ the worst case L_p error for algorithm A_n satisfies*

$$e^{\text{wor}}(A_n; L_p) \geq \sqrt{\frac{2}{1/\gamma_{\{j\}} + 1/\gamma_{\{k\}}}} \, n^{-\alpha/4} \quad \text{for any } j, k \in \{1 : s\} \text{ with } j \neq k.$$

In particular, if $\gamma_{\{1\}} \geq \gamma_{\{2\}} \geq \cdots > 0$ then

$$e^{\text{wor}}(A_n; L_p) \geq \sqrt{2/\left(\gamma_{\{1\}}^{-1} + \gamma_{\{2\}}^{-1}\right)} \, n^{-\alpha/4}.$$

Proof Without loss of generality we assume $\gamma_{\{1\}} \geq \gamma_{\{2\}} \geq \cdots > 0$. The heart of the matter is that there exists a non-zero integer vector $\boldsymbol{h}^*$ of length s in the 2-dimensional set

$$D_n := \left\{(h_1, h_2, 0, \ldots, 0) \ : \ h_j \in \mathbb{Z}, \ 0 \leq |h_j| \leq \lfloor \sqrt{n} \rfloor, \ j = 1, 2\right\},$$

such that

$$\boldsymbol{h}^* \cdot z \equiv 0 \pmod{n}. \tag{10}$$

(In the language of dual lattices, see [24], there exists a point of the dual lattice in $D_n \setminus \{\mathbf{0}\}$.) To prove this fact, we define $\widetilde{D}_n$, the positive quadrant of D_n, by

$$\widetilde{D}_n := \{(h_1, h_2, 0, \ldots, 0) : h_j \in \mathbb{Z}, \ 0 \le h_j \le \lfloor\sqrt{n}\rfloor, \ j = 1, 2\},$$

noting that if $\mathbf{h}, \mathbf{h}' \in \widetilde{D}_n$ then $\mathbf{h} - \mathbf{h}' \in D_n$. Now define

$$E_n(\mathbf{z}) := \{(\mathbf{h} \cdot \mathbf{z} \bmod n) \in \{0, \ldots, n-1\} : \mathbf{h} \in \widetilde{D}_n\}.$$

Since $|E_n(\mathbf{z})| \le n$ and $|\widetilde{D}_n| = (1 + \lfloor\sqrt{n}\rfloor)^2 > n$, it follows from the pigeonhole principle that two distinct elements of $\widetilde{D}_n$, say $\mathbf{h}$ and $\mathbf{h}'$, yield the same element of $E_n(\mathbf{z})$; from this it follows that $\mathbf{h}^* := \mathbf{h} - \mathbf{h}'$ satisfies (10). A "fooling function" is then defined by

$$q(\mathbf{y}) := e^{2\pi i h_1^* \mathbf{e}_1 \cdot \mathbf{y}} - e^{-2\pi i h_2^* \mathbf{e}_2 \cdot \mathbf{y}} = e^{-2\pi i h_2^* \mathbf{e}_2 \cdot \mathbf{y}}(e^{2\pi i \mathbf{h}^* \cdot \mathbf{y}} - 1), \qquad \mathbf{y} \in \mathbb{R}^s,$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the unit vectors corresponding to variables 1 and 2. By construction, q vanishes at all the lattice points (5). For this function, since the two terms in q are orthogonal with respect to the inner products in $H = H_{s,\alpha,\boldsymbol{\gamma}}$, the squared H norm satisfies

$$\|q\|_H^2 = r(h_1^* \mathbf{e}_1) + r(-h_2^* \mathbf{e}_2) = \frac{|h_1^*|^\alpha}{\gamma_{\{1\}}} + \frac{|-h_2^*|^\alpha}{\gamma_{\{2\}}} \le \left(\frac{1}{\gamma_{\{1\}}} + \frac{1}{\gamma_{\{2\}}}\right) n^{\alpha/2}.$$

On the other hand, the L_p norm is bounded from below by

$$\|q\|_{L_p} \ge \|q\|_{L_1} = \int_{[0,1]^2} |e^{2\pi i(h_1^* y_1 + h_2^* y_2)} - 1| \, \mathrm{d}y_1 \, \mathrm{d}y_2$$

$$= 2 \int_{[0,1]^2} |\sin(\pi(h_1^* y_1 + h_2^* y_2))| \, \mathrm{d}y_1 \, \mathrm{d}y_2.$$

This integrand is even with respect to h_1^* and h_2^* separately, so both h_1^* and h_2^* can be considered as non-negative. First assume that both h_1^* and h_2^* are positive, and partition the square into boxes of size $1/h_1^* \times 1/h_2^*$. It is easy to see that each box gives the same contribution to the integral, and hence

$$2 \int_{[0,1]^2} |\sin(\pi(h_1^* y_1 + h_2^* y_2))| \, \mathrm{d}y_1 \, \mathrm{d}y_2$$

$$= 2 h_1^* h_2^* \int_0^{1/h_1^*} \int_0^{1/h_2^*} |\sin(\pi(h_1^* y_1 + h_2^* y_2))| \, \mathrm{d}y_1 \, \mathrm{d}y_2$$

$$= 2 \int_0^1 \int_0^1 |\sin(\pi(z_1 + z_2))| \, \mathrm{d}z_1 \, \mathrm{d}z_2 = \frac{4}{\pi}.$$

(For the last step it may be useful to note that the integrand in the inner integral is 1-periodic, making the inner integral independent of z_2.) If we have $h_1^* > 0$ and $h_2^* = 0$

or vice versa, we again have $\|q\|_{L_1} = 4/\pi$. Since $\boldsymbol{h}^*$ is non-zero, we obtain

$$\frac{\|q\|_{L_p}}{\|q\|_H} \geq \frac{4/\pi}{\sqrt{1/\gamma_{\{1\}} + 1/\gamma_{\{2\}}}} \, n^{-\alpha/4}. \tag{11}$$

If we now define $g := q/\|q\|_H$, then g belongs to the unit ball in H and vanishes at all the points of the lattice (5), and $\|g\|_{L_p}$ is bounded below by the right-hand side of (11). Since $A_n(g)$ depends on g only through its values at the lattice points, and g vanishes at all those points, it follows that $A_n(g) = A_n(0) = 0$, with the last step following from the assumption on A_n. From the definition of worst case error we conclude that $e^{\mathrm{wor}}(A_n, L_p) \geq \|g - A_n(g)\|_{L_p} = \|g\|_{L_p}$, which is bounded below by the right-hand side of (11), completing the proof. $\square$

3.2 Upper bound on the worst case L_2 error

In this section, we obtain explicit L_2 error bounds for the kernel interpolant by using Theorem 2.2 combined with error bounds given for an explicit trigonometric polynomial approximation in [3,4] which extends the construction from [18,19] to general weights. (An alternative approach to obtain an upper bound would be to use a "reconstruction lattice", see, e.g., [1,13,16].)

The lattice algorithm $A_{n,M}^{\dagger}$ applied to a target function $f \in H$ takes the form

$$(A_{n,M}^{\dagger}(f))(\boldsymbol{y}) := \sum_{\boldsymbol{h} \in \mathcal{A}_s(M)} \left(\frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k\boldsymbol{z}}{n}\right) e^{-2\pi i k \boldsymbol{h} \cdot \boldsymbol{z}/n} \right) e^{2\pi i \boldsymbol{h} \cdot \boldsymbol{y}}, \tag{12}$$

which is obtained by applying a lattice integration rule to the Fourier coefficients in the orthogonal projection onto a finite index set defined for some parameter $M > 0$ by

$$\mathcal{A}_s(M) := \{\boldsymbol{h} \in \mathbb{Z}^s : r(\boldsymbol{h}) \leq M\}. \tag{13}$$

The error for this algorithm consists of the error from truncation to the index set $\mathcal{A}_s(M)$ together with the quadrature error from approximating those Fourier coefficients with indices $\boldsymbol{h} \in \mathcal{A}_s(M)$, leading to a worst case L_2 approximating error bound of the form

$$e^{\mathrm{wor}}(A_{n,M}^{\dagger}; L_2) \leq \left(\frac{1}{M} + M \, \mathcal{S}_s(\boldsymbol{z}) \right)^{1/2}. \tag{14}$$

The quantity $\mathcal{S}_s(\boldsymbol{z})$ (see [3] for details) can be used as a search criterion in a component-by-component (CBC) construction for finding suitable lattice generating vectors $\boldsymbol{z}$, and has the key advantage that it does *not* depend on the index set $\mathcal{A}_s(M)$. The analysis in [3] together with the optimality of the kernel interpolant (see Theorem 2.2) leads to the following theorem.

Theorem 3.2 *Given $s \geq 1$, $\alpha > 1$, weights $(\gamma_u)_{u \subset \mathbb{N}}$ with $\gamma_\emptyset := 1$, and prime n, the worst case L_2 approximation error of the kernel interpolant $A_n^*(f) = f_n$ defined by (4), (5) and (6), using the generating vector z obtained from the CBC construction with search criterion $\mathcal{S}_s(z)$ in [3,4], satisfies for all $\lambda \in (\frac{1}{\alpha}, 1]$,*

$$e^{\mathrm{wor}}(A_n^*; L_2) \leq \sqrt{2} \left[\mathcal{S}_s(z) \right]^{1/4}$$
$$\leq \frac{\kappa}{n^{1/(4\lambda)}} \left(\sum_{u \subseteq \{1:s\}} \max(|u|, 1) \, \gamma_u^\lambda \, [2\zeta(\alpha\lambda)]^{|u|} \right)^{1/(2\lambda)}, \qquad (15)$$

with $\kappa := \sqrt{2} \left[\max(6, 2.5 + 2^{2\alpha\lambda+1}) \right]^{1/(4\lambda)}$ and $\zeta(x) := \sum_{k=1}^\infty k^{-x}$ denoting the Riemann zeta function for $x > 1$. Hence

$$e^{\mathrm{wor}}(A_n^*; L_2) = \mathcal{O}(n^{-\alpha/4+\delta}) \quad \text{for every } \delta \in (0, \alpha/4),$$

where the implied constant depends on δ but is independent of s provided that

$$\sum_{\substack{u \subset \mathbb{N} \\ |u| < \infty}} \max(|u|, 1) \, \gamma_u^{\frac{1}{\alpha-4\delta}} \, [2\zeta(\frac{\alpha}{\alpha-4\delta})]^{|u|} < \infty.$$

Proof The optimality of the kernel interpolant established in Theorem 2.2 means that $e^{\mathrm{wor}}(A_n^*; L_2) \leq e^{\mathrm{wor}}(A_{n,M}^\dagger; L_2)$ for all M, and therefore the upper bound in (14) also serves as an upper bound for the kernel interpolant. It is easy to verify that the bound in (14) can be minimized by setting $M = \sqrt{1/\mathcal{S}_s(z)}$, leading to (15). The subsequent bound follows from [3, Theorem 3.5]. The big-$\mathcal{O}$ bound is then obtained by taking $\lambda = 1/(\alpha - 4\delta)$. $\qquad \square$

From this result (which by Theorem 3.1 is almost best possible with respect to the order of convergence) we immediately obtain an error bound for the kernel interpolant.

Theorem 3.3 *Under the conditions of Theorem 3.2, and with lattice generating vector z obtained by the CBC construction in [3,4], for any $f \in H$, we have for the kernel interpolant f_n defined by (4), (5) and (6),*

$$\|f - f_n\|_{L_2} \leq \frac{\kappa}{n^{1/(4\lambda)}} \left(\sum_{u \subseteq \{1:s\}} \max(|u|, 1) \, \gamma_u^\lambda \, [2\zeta(\alpha\lambda)]^{|u|} \right)^{1/(2\lambda)} \|f\|_H.$$

We stress again that the CBC construction in [3,4] does *not* require the explicit construction of the index set $\mathcal{A}_s(M)$ in order to determine an appropriate generating vector z. However, the expression $\mathcal{S}_s(z)$ (see [3] for details) used as the search criterion does depend in a complicated way on the weights γ_u, and therefore the target dimension s needs to be fixed at the start of the CBC construction (except for the case of product weights). For weights with no special structure, the computational cost will be exponentially large in s. We consider some special forms of weights:

- **Product weights**: $\gamma_{\mathrm{u}} = \prod_{j \in \mathrm{u}} \gamma_j$, specified by one sequence $(\gamma_j)_{j \geq 1}$.
- **POD weights** (product and order dependent): $\gamma_{\mathrm{u}} = \Gamma_{|\mathrm{u}|} \prod_{j \in \mathrm{u}} \gamma_j$, specified by two sequences $(\Gamma_\ell)_{\ell \geq 0}$ and $(\gamma_j)_{j \geq 1}$.
- **SPOD weights** (smoothness-driven product and order dependent) with degree $\sigma \geq 1$:

$$\gamma_{\mathrm{u}} = \sum_{\boldsymbol{\nu}_{\mathrm{u}} \in \{1:\sigma\}^{|\mathrm{u}|}} \Gamma_{|\boldsymbol{\nu}_{\mathrm{u}}|} \prod_{j \in \mathrm{u}} \gamma_{j, \nu_j},$$

specified by the sequences $(\Gamma_\ell)_{\ell \geq 0}$ and $(\gamma_{j,\nu})_{j \geq 1}$ for each $\nu = 1, \ldots, \sigma$, where $|\boldsymbol{\nu}_{\mathrm{u}}| := \sum_{j \in \mathrm{u}} \nu_j$.

Fast CBC construction of lattice generating vector for L_2 approximation has the cost of

$$\begin{aligned}
&\mathcal{O}\big(s\,n \log(n)\big) && \text{for product weights,} \\
&\mathcal{O}\big(s\,n \log(n) + s^2 \log(s)n\big) && \text{for POD weights,} \\
&\mathcal{O}\big(s\,n \log(n) + s^3 \sigma^2 n\big) && \text{for SPOD weights with degree } \sigma \geq 2,
\end{aligned}$$

plus storage cost and pre-computation cost for POD and SPOD weights, see [4].

4 Application to PDEs with random coefficients

As an application, we apply our kernel interpolation scheme to a forward uncertainty quantification problem, namely, a PDE problem with an uncertain, *periodically* parameterized diffusion coefficient, fitting the theoretical framework considered in the preceding sections. The kernel interpolant can be postprocessed with low computational cost to obtain statistics of the PDE solution itself or functionals of the solution for uncertainty quantification.

Letting $D \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, be a bounded domain with Lipschitz boundary, we consider the problem of finding $u \colon D \times \Omega \to \mathbb{R}$ that satisfies

$$-\nabla \cdot (a(\boldsymbol{x}, \omega)\, \nabla u(\boldsymbol{x}, \omega)) = q(\boldsymbol{x}), \qquad \boldsymbol{x} \in D, \tag{16}$$

$$u(\boldsymbol{x}, \omega) = 0, \qquad \boldsymbol{x} \in \partial D, \tag{17}$$

for almost all events $\omega \in \Omega$ in the probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with

$$a(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \frac{1}{\sqrt{6}} \sum_{j \geq 1} \sin(2\pi Y_j(\omega))\, \psi_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in D, \omega \in \Omega, \tag{18}$$

where $a_0 \in L_\infty(D)$, $\psi_j \in L_\infty(D)$ for all $j \geq 1$ are such that $\sum_{j \geq 1} |\psi_j(\boldsymbol{x})| < \infty$ for any $\boldsymbol{x} \in D$, and $Y_1, Y_2, \ldots$ are i.i.d. random variables uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. This type of random field is not new in the context of uncertainty quantification. Indeed, the random variable $\sin(2\pi Y_j(\omega))$ induces the arcsine measure as its distribution: for

if $Y(\omega)$ is uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$, then $Z(\omega) := \sin(2\pi Y(\omega))$ has the probability density $\frac{1}{\pi} \frac{1}{\sqrt{1-z^2}}$ on $[-1, 1]$. Thus, a is identical, up to the law to the random field

$$\hat{a}(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \frac{1}{\sqrt{6}} \sum_{j \geq 1} Z_j(\omega) \psi_j(\boldsymbol{x}) \tag{19}$$

with Z_j i.i.d. random variables with arcsine distribution on $[-1, 1]$. Expression (19) would be the starting point for deriving a polynomial chaos approximation [27] of the solution in terms of Chebyshev polynomials of the first kind [22]. In this paper, however, we want to exploit periodicity, hence we consider rather the formulation (18) and a different approximation method based on kernel interpolation.

Since the expression (18) is periodic in the random variable Y_j, we can shift those random variables so that their range is $[0, 1]$ instead of $[-\frac{1}{2}, \frac{1}{2}]$, i.e., we consider the equivalent parametric space $U := [0, 1]^{\mathbb{N}}$. Let $\mathcal{B}(U)$ be the Borel σ-algebra corresponding to the product topology on $U = [0, 1]^{\mathbb{N}}$, and equip $(U, \mathcal{B}(U))$ with the product uniform measure; see, for example, [23] for details. The weak formulation of (16)–(17) can then be stated parametrically as: for $\boldsymbol{y} \in U$, find $u(\cdot, \boldsymbol{y}) \in H_0^1(D)$ such that

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \nabla u(\boldsymbol{x}, \boldsymbol{y}) \cdot \nabla \phi(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \langle q, \phi \rangle_{H^{-1}(D), H_0^1(D)}, \quad \forall \phi \in H_0^1(D), \tag{20}$$

where the datum $q \in H^{-1}(D)$ is fixed and the diffusion coefficient is given by

$$a(\boldsymbol{x}, \boldsymbol{y}) = a_0(\boldsymbol{x}) + \frac{1}{\sqrt{6}} \sum_{j \geq 1} \sin(2\pi y_j) \, \psi_j(\boldsymbol{x}), \quad \boldsymbol{x} \in D, \; \boldsymbol{y} \in U. \tag{21}$$

Here $H_0^1(D)$ denotes the subspace of the L_2-Sobolev space $H^1(D)$ with vanishing trace on ∂D, and $H^{-1}(D)$ denotes the topological dual of $H_0^1(D)$, and $\langle \cdot, \cdot \rangle_{H^{-1}(D), H_0^1(D)}$ denotes the duality pairing between $H^{-1}(D)$ and $H_0^1(D)$. We endow the Sobolev space $H_0^1(D)$ with the norm $\|v\|_{H_0^1(D)} := \|\nabla v\|_{L_2(D)}$.

Since we now have two sets of variables $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in U$, from here on we will make the domain D and U explicit in our notation. We state the following assumptions and refer to them as they become needed:

(A1) $a_0 \in L_\infty(D)$, $\psi_j \in L_\infty(D)$ for all $j \geq 1$, and $\sum_{j \geq 1} \|\psi_j\|_{L_\infty(D)} < \infty$;
(A2) there exist positive constants $a_{\min}$ and $a_{\max}$ such that $0 < a_{\min} \leq a(\boldsymbol{x}, \boldsymbol{y}) \leq a_{\max} < \infty$ for all $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in U$;
(A3) $\sum_{j \geq 1} \|\psi_j\|_{L_\infty(D)}^p < \infty$ for some $0 < p < 1$;
(A4) $a_0 \in W^{1,\infty}(D)$ and $\sum_{j \geq 1} \|\psi_j\|_{W^{1,\infty}(D)} < \infty$, where

$$\|v\|_{W^{1,\infty}(D)} := \max\{\|v\|_{L_\infty(D)}, \|\nabla v\|_{L_\infty(D)}\};$$

(A5) $\|\psi_1\|_{L_\infty(D)} \geq \|\psi_2\|_{L_\infty(D)} \geq \cdots$;

(A6) the physical domain $D \subset \mathbb{R}^d, d \in \{1, 2, 3\}$, is a convex and bounded polyhedron with plane faces.

Let assumptions (A1) and (A2) be in effect. Then the Lax–Milgram lemma [2] implies unique solvability of the problem (20) for all $\boldsymbol{y} \in U$, with the solution satisfying the *a priori* bound

$$\|u(\cdot, \boldsymbol{y})\|_{H_0^1(D)} \leq \frac{\|q\|_{H^{-1}(D)}}{a_{\min}} \qquad \text{for all } \boldsymbol{y} \in U. \tag{22}$$

Moreover, from the recent paper [12, Theorem 2.3] we know, after differentiating the PDE (20), that the mixed derivatives of the PDE solution are 1-periodic and bounded by

$$\|\partial_{\boldsymbol{y}}^{\boldsymbol{\nu}} u(\cdot, \boldsymbol{y})\|_{H_0^1(D)} \leq \frac{\|q\|_{H^{-1}(D)}}{a_{\min}} (2\pi)^{|\boldsymbol{\nu}|} \sum_{\boldsymbol{m} \leq \boldsymbol{\nu}} |\boldsymbol{m}|! \prod_{j \geq 1} (b_j^{m_j} S(\nu_j, m_j)) \tag{23}$$

for all $\boldsymbol{y} \in U$ and all multi-indices $\boldsymbol{\nu} \in \mathbb{N}_0^{\infty}$ with finite order $|\boldsymbol{\nu}| := \sum_{j \geq 1} \nu_j < \infty$, and we define

$$b_j := \frac{1}{\sqrt{6}} \frac{\|\psi_j\|_{L_\infty(D)}}{a_{\min}} \qquad \text{for all } j \geq 1. \tag{24}$$

Furthermore, $S(\sigma, m)$ denotes the *Stirling number of the second kind* for integers $\sigma \geq m \geq 0$, with the convention that $S(\sigma, 0) = \delta_{\sigma, 0}$. In [12] we considered a function space with respect to $\boldsymbol{y}$ with a supremum norm rather than an L_2-based norm, so here we need to write down the relevant L_2-based norm bound instead. Moreover, we want to approximate the solution u directly, rather than a bounded linear functional $G(u)$ of the PDE solution.

For our proposed approximation scheme, we require the target function to be pointwise well-defined with respect to both the physical variable and the parametric variable. In terms of our PDE application, this can be achieved either by assuming additional regularity of both the diffusion coefficient a and the source term q or, alternatively, by analyzing instead the construction of the kernel interpolant for the finite element approximation of u (which is naturally pointwise well-defined everywhere). Here we focus on the latter case, in which the kernel interpolant is crafted for the finite element approximation of u. This is also the setting that arises in practical computations, where one only ever has access to a numerical approximation of the solution to (20), with the diffusion coefficient (21) truncated to a finite number of terms. To this end, we split our analysis into three parts: *dimension truncation error*, *finite element error*, and *kernel interpolation error*.

4.1 Dimension truncation error

In anticipation of the forthcoming discussion we define the dimensionally truncated solution of (20) as

$$u_s(\cdot, \boldsymbol{y}) := u_s(\cdot, (y_1, \ldots, y_s)) := u(\cdot, (y_1, \ldots, y_s, 0, 0, \ldots)), \qquad \boldsymbol{y} \in U.$$

Moreover, let us introduce the shorthand notations $U_s := U_{\leq s} := [0, 1]^s$, $U_{>s} := \{(y_j)_{j \geq s+1} : y_j \in [0, 1]\}$, and $\boldsymbol{y}_{>s} := (y_{s+1}, y_{s+2}, \ldots)$.

For an $\mathbb{R}$-valued function on U that is Lebesgue integrable with respect to the uniform measure on $\mathcal{B}(U)$, we use the notation $\int_U F(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}$ for the integral of F over U. Similarly, for an integrable function $\tilde{F}$ on $U_{>s}$, we denote the integral over $U_{>s}$ with respect to the uniform measure by $\int_{U_{>s}} \tilde{F}(\boldsymbol{y}_{>s}) \, \mathrm{d}\boldsymbol{y}_{>s}$.

Arguing as in [17, Theorem 5.1], it is not difficult to see that

$$\sup_{\boldsymbol{y} \in U} \|u(\cdot, \boldsymbol{y}) - u_s(\cdot, \boldsymbol{y})\|_{H_0^1(D)} = \mathcal{O}(s^{-1/p+1})$$

holds under assumptions (A1)–(A3) and (A5). In what follows, we consider the dimension truncation error in the L_2-norm in the stochastic parameter, and establish the rate $\mathcal{O}(s^{-1/p+1/2})$, which is one half order better. This case does not appear to have been considered in the existing literature. Notably, this rate is only half that of the rate proved in [12] for integration problem with respect to $\boldsymbol{y}$:

$$\left| \int_U G(u(\cdot, \boldsymbol{y}) - u_s(\cdot, \boldsymbol{y})) \, \mathrm{d}\boldsymbol{y} \right| = \mathcal{O}(s^{-2/p+1}), \qquad G \in H^{-1}(D).$$

We will establish a dimension truncation error for a general class of parametrized random fields that includes (21), without the periodicity assumption. Our proof adapts the argument by Gantner [6] to the $L^2(U; H_0^1(D))$-norm estimate.

Theorem 4.1 *Suppose that* (A1)*,* (A3) *and* (A5) *hold. Let* $\xi : [0, 1] \to \mathbb{R}$ *be an* $L_\infty([0, 1])$*-function such that*

$$\int_0^1 \xi(y) \, \mathrm{d}y = 0. \tag{25}$$

Suppose further that the function

$$a(\boldsymbol{x}, \boldsymbol{y}) = a_0(\boldsymbol{x}) + \sum_{j \geq 1} \xi(y_j) \, \psi_j(\boldsymbol{x}), \qquad \boldsymbol{x} \in D, \ \boldsymbol{y} \in U, \tag{26}$$

satisfies (A2)*. Then for any* $s \in \mathbb{N}$*, there exists a constant* $C > 0$ *such that*

$$\sqrt{\int_U \int_D (u(\boldsymbol{x}, \boldsymbol{y}) - u_s(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}} \leq c_D \sqrt{\int_U \int_D |\nabla(u - u_s)|^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}$$

$$\leq C \|q\|_{H^{-1}(D)} \, s^{-\left(\frac{1}{p} - \frac{1}{2}\right)},$$

where $u \in H_0^1(D)$ denotes the solution of the equation (20) but with $a(\mathbf{x}, \mathbf{y})$ given by (26), $u_s \in H_0^1(D)$ denotes the corresponding dimensionally truncated solution, $c_D > 0$ is the Poincaré constant of the embedding $H_0^1(D) \hookrightarrow L_2(D)$, and the constant $C > 0$ is independent of s and q.

Proof We begin by introducing some helpful notations. For $\mathbf{y} \in U$, let us define the operators $B, B^s \colon H_0^1(D) \to H^{-1}(D)$ by

$$B := B(\mathbf{y}) := B_0 + \sum_{k=1}^{\infty} \xi(y_k) B_k \quad \text{and} \quad B^s := B^s(\mathbf{y}) := B_0 + \sum_{k=1}^{s} \xi(y_k) B_k,$$

where the operators $B_k \colon H_0^1(D) \to H^{-1}(D)$ are defined by

$$\langle B_0 v, w \rangle_{H^{-1}(D), H_0^1(D)} := \langle a_0 \nabla v, \nabla w \rangle_{L_2(D)}$$

and $\langle B_k v, w \rangle_{H^{-1}(D), H_0^1(D)} := \langle \psi_k \nabla v, \nabla w \rangle_{L_2(D)}$ for $v, w \in H_0^1(D)$ and $k \geq 1$. This allows the equation (20) with the coefficient a given by (26) to be written as $Bu = q$. It is easy to see that the assumptions (A1) and (A2) ensure that both $B(\mathbf{y})$ and $B^s(\mathbf{y})$ are boundedly invertible linear maps for all $\mathbf{y} \in U$, with the norms of B and B^s both bounded by $a_{\max}$, and the norms of both B^{-1} and $(B^s)^{-1}$ bounded by $a_{\min}^{-1}$. Thus we can write $u := u(\mathbf{y}) := B^{-1} q$ and $u_s := u_s(\mathbf{y}) := (B^s)^{-1} q$ for all $\mathbf{y} \in U$.

Only in this proof, we redefine (24) by $b_j := \|\xi\|_{\infty} \|\psi_j\|_{L_{\infty}(D)} / a_{\min}$, with $\|\xi\|_{\infty} := \|\xi\|_{L_{\infty}([0,1])}$. Notice that with $\xi = \frac{1}{\sqrt{6}} \sin(2\pi \cdot)$ we recover (24). Let $s' \in \mathbb{Z}_+$ be such that

$$\sum_{j=s'+1}^{\infty} b_j < \frac{1}{2}. \tag{27}$$

Without loss of generality, we can assume that $s \geq s'$ since the assertion in the theorem can subsequently be extended to all values of s by making a simple adjustment of the constant $C > 0$ (see the end of the proof). Then for all $j \geq s' + 1$ and all $s \geq s'$ we have

$$b_j < \frac{1}{2}, \tag{28}$$

$$\sup_{\mathbf{y} \in U} \|(B^s)^{-1}(B - B^s)\|_{H_0^1(D) \to H_0^1(D)} \leq \sum_{j=s+1}^{\infty} b_j < \frac{1}{2} < 1. \tag{29}$$

The bound (29) permits the use of a Neumann series expansion

$$u - u_s = B^{-1} q - u_s = [I + (B^s)^{-1}(B - B^s)]^{-1}(B^s)^{-1} q - u_s$$

$$= \sum_{k=0}^{\infty} (-(B^s)^{-1}(B - B^s))^k (B^s)^{-1} q - u_s$$

$$= \sum_{k=1}^{\infty}(-(B^s)^{-1}(B - B^s))^k u_s = \sum_{k=1}^{\infty}(-1)^k \left(\sum_{i=s+1}^{\infty} \xi(y_i)(B^s)^{-1} B_i \right)^k u_s, \quad (30)$$

where it is assumed that the product symbol respects the non-commutative nature of the operators $(B^s)^{-1} B_j$, $j \geq 1$.

Our strategy is to estimate first

$$S := \int_U \int_D a_s(\boldsymbol{x}, \boldsymbol{y}) |\nabla(u - u_s)|^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}$$

and then deduce by the Poincaré inequality $\|u\|_{L_2(D)} \leq c_D \|u\|_{H_0^1(D)}$, with $c_D > 0$ depending only on the domain D, together with uniform coercivity, that

$$\int_U \int_D (u - u_s)^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \leq c_D^2 \int_U \int_D |\nabla(u - u_s)|^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \leq \frac{c_D^2}{a_{\min}} S.$$

Let $\mathscr{B}_s : H_0^1(D) \to H_0^1(D)$ be defined by

$$\mathscr{B}_s(\boldsymbol{y}) := \sum_{i=s+1}^{\infty} \xi(y_i)(B^s(\boldsymbol{y}))^{-1} B_i,$$

and observe that $\mathscr{B}_s$ is self-adjoint with respect to the inner product

$$\langle v, w \rangle_{\boldsymbol{y}_{\leq s}} := \int_D a\big(\boldsymbol{x}, (\boldsymbol{y}_{\leq s}, 0, \ldots)\big) \nabla v(\boldsymbol{x}) \cdot \nabla w(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \langle B^s(\boldsymbol{y})v, w \rangle_{H^{-1}(D), H_0^1(D)}.$$

Indeed, for any $v, w \in H_0^1(D)$ we have

$$\langle \mathscr{B}_s(\boldsymbol{y})v, w \rangle_{\boldsymbol{y}_{\leq s}} = \sum_{i=s+1}^{\infty} \xi(y_i) \langle B^s(\boldsymbol{y})(B^s(\boldsymbol{y}))^{-1} B_i v, w \rangle_{H^{-1}, H_0^1}$$

$$= \sum_{i=s+1}^{\infty} \xi(y_i) \int_D \psi_i \nabla v \cdot \nabla w \, \mathrm{d}\boldsymbol{x} = \sum_{i=s+1}^{\infty} \xi(y_i) \langle B_i w, v \rangle_{H^{-1}, H_0^1} = \langle \mathscr{B}_s(\boldsymbol{y})w, v \rangle_{\boldsymbol{y}_{\leq s}}.$$

Hence, from (30) we have

$$\langle u - u_s, u - u_s \rangle_{\boldsymbol{y}_{\leq s}} = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} (-1)^{k+\ell} \langle \mathscr{B}_s^k u_s, \mathscr{B}_s^\ell u_s \rangle_{\boldsymbol{y}_{\leq s}}$$

$$= \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} (-1)^{k+\ell} \langle \mathscr{B}_s^{k+\ell} u_s, u_s \rangle_{\boldsymbol{y}_{\leq s}}$$

$$= \sum_{m=2}^{\infty} (-1)^m (m - 1) \langle \mathscr{B}_s^m u_s, u_s \rangle_{\boldsymbol{y}_{\leq s}}$$

$$= \sum_{m=2}^{\infty} (-1)^m (m-1)$$

$$\times \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} \left(\prod_{j=1}^{m} \frac{\xi(y_{\eta_j})}{\|\xi\|_\infty} \right) \left\langle \prod_{j=1}^{m} \|\xi\|_\infty (B^s(\boldsymbol{y})^{-1} B_{\eta_j}) u_s, u_s \right\rangle_{\boldsymbol{y}_{\leq s}},$$

(31)

where we used the notation $\sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} := \lim_{\tilde{s} \to \infty} \sum_{\boldsymbol{\eta} \in \{s+1:\tilde{s}\}^m}$, and the latter product is assumed to respect the non-commutative nature of the operators. Introducing

$$\boldsymbol{v}(\boldsymbol{\eta}) := (v_i(\boldsymbol{\eta}))_{i \geq 1} := (\#\{j = 1, \ldots, m : \eta_j = i\})_{i \geq 1}$$

for each $\boldsymbol{\eta} \in \{s+1, s+2, \ldots\}^m$, we have $v_i(\boldsymbol{\eta}) = 0$, $i = 1, \ldots, s$, $|\boldsymbol{v}(\boldsymbol{\eta})| := \sum_{i=1}^{\infty} v_i(\boldsymbol{\eta}) = m$, and

$$\prod_{j=1}^{m} \frac{\xi(y_{\eta_j})}{\|\xi\|_\infty} = \prod_{i=s+1}^{\infty} \left(\frac{\xi(y_i)}{\|\xi\|_\infty} \right)^{v_i(\boldsymbol{\eta})}.$$

Define

$$c_{\boldsymbol{v}} := \left| \int_{U_{>s}} \prod_{i \in \text{supp}(\boldsymbol{v})} \left(\frac{\xi(y_i)}{\|\xi\|_\infty} \right)^{v_i} \mathrm{d}\boldsymbol{y}_{>s} \right| \leq 1,$$

and note from (25) that $c_{\boldsymbol{v}} = 0$ if for some $i \in \text{supp}(\boldsymbol{v})$ we have $v_i = 1$. Then we have, using (31),

$$\int_U \int_D |\nabla(u - u_s)|^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y} \leq \frac{1}{a_{\min}} \int_U \langle u - u_s, u - u_s \rangle_{\boldsymbol{y}_{\leq s}} \, \mathrm{d}\boldsymbol{y}$$

$$= \frac{1}{a_{\min}} \sum_{m=2}^{\infty} (-1)^m (m-1) \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} \int_{U_{>s}} \left(\prod_{i=s+1}^{\infty} \left(\frac{\xi(y_i)}{\|\xi\|_\infty} \right)^{v_i(\boldsymbol{\eta})} \right) \mathrm{d}\boldsymbol{y}_{>s}$$

$$\times \int_{U_{\leq s}} \left\langle \prod_{j=1}^{m} \|\xi\|_\infty ((B^s)^{-1} B_{\eta_j}) u_s, u_s \right\rangle_{\boldsymbol{y}_{\leq s}} \mathrm{d}\boldsymbol{y}_{\leq s}$$

$$\leq \frac{1}{a_{\min}} \sum_{m=2}^{\infty} (m-1) \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} c_{\boldsymbol{v}(\boldsymbol{\eta})}$$

$$\times \left| \int_{U_{\leq s}} \left\langle \prod_{j=1}^{m} \|\xi\|_\infty ((B^s)^{-1} B_{\eta_j}) u_s, u_s \right\rangle_{\boldsymbol{y}_{\leq s}} \mathrm{d}\boldsymbol{y}_{\leq s} \right|,$$

which can be further bounded by

$$\frac{1}{a_{\min}} \sum_{m=2}^{\infty} (m-1) \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} c_{\boldsymbol{v}(\boldsymbol{\eta})} a_{\max} \left\| \prod_{j=1}^{m} \|\xi\|_{\infty} ((B^s)^{-1} B_{\eta_j}) u_s \right\|_{H_0^1} \|u_s\|_{H_0^1}$$

$$\leq \frac{1}{a_{\min}} \sum_{m=2}^{\infty} (m-1) \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} c_{\boldsymbol{v}(\boldsymbol{\eta})} a_{\max} \left(\prod_{j=1}^{m} b_{\eta_j} \right) \|u_s\|_{H_0^1}^2$$

$$\leq \left(\frac{\|q\|_{H^{-1}}}{a_{\min}} \right)^2 \frac{a_{\max}}{a_{\min}} \sum_{m=2}^{\infty} (m-1) \sum_{\boldsymbol{\eta} \in \{s+1:\infty\}^m} c_{\boldsymbol{v}(\boldsymbol{\eta})} \left(\prod_{j=1}^{m} b_{\eta_j} \right),$$

where the sum of $\boldsymbol{\eta}$ simplifies to

$$\sum_{\substack{|\boldsymbol{v}|=m \\ v_i=0,\ i \leq s}} \binom{m}{\boldsymbol{v}} c_{\boldsymbol{v}} \left(\prod_{i=s+1}^{\infty} b_i^{v_i} \right) \leq \sum_{\substack{|\boldsymbol{v}|=m \\ v_i=0,\ i \leq s \\ v_i \neq 1,\ i>s}} \binom{m}{\boldsymbol{v}} \left(\prod_{i=s+1}^{\infty} b_i^{v_i} \right).$$

The dimension truncation error is estimated by splitting the upper bound into two parts. Let $m^* \geq 3$ be an as yet undetermined index. Then

$$\int_U \int_D |\nabla(u - u_s)|^2 \, d\boldsymbol{x} \, d\boldsymbol{y}$$

$$\leq \left(\frac{\|q\|_{H^{-1}}}{a_{\min}} \right)^2 \frac{a_{\max}}{a_{\min}} \sum_{m=2}^{m^*-1} (m-1) \sum_{\substack{|\boldsymbol{v}|=m \\ v_i=0,\ i<s \\ v_i \neq 1,\ i>s}} \binom{m}{\boldsymbol{v}} \left(\sum_{i=s+1}^{\infty} b_i^{v_i} \right)$$

$$+ \left(\frac{\|q\|_{H^{-1}}}{a_{\min}} \right)^2 \frac{a_{\max}}{a_{\min}} \sum_{m=m^*}^{\infty} (m-1) \left(\sum_{i=s+1}^{\infty} b_i \right)^m. \tag{32}$$

We can estimate the sum in the first term of (32) by

$$\sum_{m=2}^{m^*-1} (m-1) \sum_{\substack{|\boldsymbol{v}|=m \\ v_i=0,\ i<s \\ v_i \neq 1,\ i>s}} \binom{m}{\boldsymbol{v}} \left(\sum_{i=s+1}^{\infty} b_i^{v_i} \right) \leq (m^*-2)(m^*-1)! \sum_{\substack{0 \neq |\boldsymbol{v}|_\infty \leq m^*-1 \\ v_i=0,\ i<s \\ v_i \neq 1,\ i \geq 1}} \boldsymbol{b}^{\boldsymbol{v}},$$

where $\boldsymbol{b}^{\boldsymbol{v}} := \prod_{i \in \mathrm{supp}(\boldsymbol{v})} b_i^{v_i}$. Furthermore, we obtain

$$\sum_{\substack{0 \neq |\boldsymbol{v}|_\infty \leq m^*-1 \\ v_i=0,\ i<s \\ v_i \neq 1,\ i \geq 1}} \boldsymbol{b}^{\boldsymbol{v}} = \prod_{j=s+1}^{\infty} \left(1 + \sum_{\ell=2}^{m^*-1} b_j^{\ell} \right) - 1 = \prod_{j=s+1}^{\infty} \left(1 + b_j^2 \frac{1 - b_j^{m^*-2}}{1 - b_j} \right) - 1$$

$$\leq \prod_{j=s+1}^{\infty} (1 + 2b_j^2) - 1 \leq \exp\left(2 \sum_{j=s+1}^{\infty} b_j^2\right) - 1 \leq 2(e-1) \sum_{j=s+1}^{\infty} b_j^2, \qquad (33)$$

where we used (28), (29), and the inequality $e^x \leq 1 + (e-1)x$ for all $x \in [0, 1]$.

Recalling (27), we find the following upper bound for the sum in the second term of (32):

$$\sum_{m=m^*}^{\infty} (m-1)\left(\sum_{j=s+1}^{\infty} b_j\right)^m \leq \sum_{m=m^*}^{\infty} \left(2 \sum_{j=s+1}^{\infty} b_j\right)^m \leq \frac{(2\sum_{j=s+1}^{\infty} b_j)^{m^*}}{1 - (2\sum_{j=s+1}^{\infty} b_j)}, \qquad (34)$$

where we used the estimates $m - 1 \leq 2^m$ and $2\sum_{j=s+1}^{\infty} b_j < 1$.

Observing that by [17, Theorem 5.1] it holds

$$\sum_{j=s+1}^{\infty} b_j \leq \left(\sum_{j=1}^{\infty} b_j^p\right)^{1/p} s^{-\frac{1}{p}+1}$$

and (with b_j replaced by b_j^2 and p replaced by $p/2$)

$$\sum_{j=s+1}^{\infty} b_j^2 \leq \left(\sum_{j=1}^{\infty} b_j^p\right)^{2/p} s^{-\frac{2}{p}+1},$$

so we see that the terms (33)–(34) can be balanced by choosing $m^* = \lceil\frac{2-p}{1-p}\rceil$. One arrives at the dimension truncation bound

$$\sqrt{\int_U \int_D (u - u_s)^2 \, dx \, dy}$$

$$\leq c_D \sqrt{\int_U \int_D |\nabla(u - u_s)|^2 \, dx \, dy} \leq C \|q\|_{H^{-1}(D)} s^{-\frac{1}{p}+\frac{1}{2}} \quad \text{for all } s \geq s',$$

where the constant $C > 0$ is independent of s and q. This proves the theorem for $s \geq s'$. The result can be extended to all $s \geq 1$ by noting that

$$\sqrt{\int_U \int_D |\nabla(u - u_s)|^2 \, dx \, dy} \leq \frac{2\|q\|_{H^{-1}(D)}}{a_{\min}} \leq \frac{2\|q\|_{H^{-1}(D)}}{a_{\min} \cdot (s')^{-1/p+1/2}} s^{-\frac{1}{p}+\frac{1}{2}}$$

for all $1 \leq s < s'$, where we used the *a priori* bound identical to (22). $\qquad \square$

Remark 4.2 Theorem 4.1 can be generalised further to include a more complex model

$$\tilde{a}(x, y) = a_0(x) + \sum_{j \geq 1} \xi_j(y_j) \psi_j(x), \qquad x \in D, \; y \in U,$$

where the function ξ in (26) is now replaced by an $L_\infty([0, 1])$ function ξ_j depending on j. Then, assuming that we have $\int_0^1 \xi_j(y)\,\mathrm{d}y = 0$, $j \geq 1$, and that $\tilde{b}_j := \|\xi_j\|_\infty \|\psi_j\|_{L_\infty(D)}/a_{\min}$ is non-increasing in j, and moreover that $\tilde{a}$ satisfies (A2), the same argument as above establishes the same estimate as in Theorem 4.1.

4.2 Finite element error

Let assumption (A6) be in effect. Let $\{V_h\}_h$ be a family of conforming finite element subspaces $V_h \subset H_0^1(D)$, parameterized by the one-dimensional mesh size $h > 0$, which are spanned by continuous, piecewise linear finite element basis functions. It is assumed that the triangulation corresponding to each V_h is obtained from an initial, regular triangulation of D by recursive, uniform partition of simplices.

For each $y \in U$, we denote by $u_h(\cdot, y) \in V_h$ the finite element solution to the system

$$\int_D a(x, y)\,\nabla u_h(x, y) \cdot \nabla v_h(x)\,\mathrm{d}x = \langle q, v_h\rangle_{H^{-1}(D), H_0^1(D)}, \quad \forall\, v_h \in V_h, \qquad (35)$$

where $q \in H^{-1}(D)$ and a is defined by (21). Under assumptions (A1)–(A2), this system is uniquely solvable and the finite element solution u_h satisfies both the *a priori* bound (22) as well as the partial derivative bounds (23). In analogy to the previous subsection, we also define the dimensionally truncated finite element solution by setting

$$u_{s,h}(\cdot, y) := u_{s,h}(\cdot, (y_1, \ldots, y_s)) := u_h(\cdot, (y_1, \ldots, y_s, 0, 0, \ldots)), \quad y \in U, \ (36)$$

where $u_h(\cdot, y) \in V_h$ is the solution of (35) for $y \in U$.

Theorem 4.3 *Under the assumptions* (A1), (A2), (A4) *and* (A6), *for every* $y \in U$ *and* $q \in H^{-1+t}(D)$ *with* $t \in [0, 1]$, *there holds the asymptotic convergence estimate*

$$\|u(\cdot, y) - u_h(\cdot, y)\|_{L_2(D)} \leq C\,h^{1+t}\,\|q\|_{H^{-1+t}(D)} \quad \text{as } h \to 0,$$

where the constant $C > 0$ *is independent of* h *and* y.

Proof Let $y \in U$. From [17, Theorem 7.2], under the assumptions (A1), (A2), (A4), and (A6), we have for every $g \in L_2(D)$ the following asymptotic convergence estimate as $h \to 0$

$$|\langle g, u(\cdot, y) - u_h(\cdot, y)\rangle_{L_2(D)}| \leq C\,h^{1+t}\,\|q\|_{H^{-1+t}(D)}\,\|g\|_{L^2(D)}, \qquad (37)$$

where the constant $C > 0$ is independent of h and y. Therefore

$$\|u(\cdot, y) - u_h(\cdot, y)\|_{L_2(D)} = \sup_{g \in L_2(D),\, \|g\|_{L_2(D)}=1} \langle g, u(\cdot, y) - u_h(\cdot, y)\rangle_{L_2(D)}$$

$$\leq C\,h^{1+t}\,\|q\|_{H^{-1+t}(D)},$$

and this concludes the proof. $\qquad\qquad\square$

4.3 Kernel interpolation error

We focus on approximating the finite element solution of the problem (20) in the following discussion, since it is essential for our approximation scheme that the function being approximated is pointwise well-defined in the physical domain D.

Let $H(U_s) = H$ denote the RKHS of functions with respect to the stochastic parameter $y \in U_s$, defined in Sect. 2.1. For every $x \in D$, let

$$u_{s,h,n}(x, \cdot) := A_n^*(u_{s,h}(x, \cdot)) \in H(U_s)$$

be the kernel interpolant of the dimensionally truncated finite element solution (36) at x as a function of y. We measure the L_2 approximation error $\|u_{s,h}(x, \cdot) - u_{s,h,n}(x, \cdot)\|_{L_2(U_s)}$ in y and then take the L_2 norm over x, to arrive at the error criterion

$$\sqrt{\int_D \|u_{s,h}(x, \cdot) - u_{s,h,n}(x, \cdot)\|_{L_2(U_s)}^2 \, dx}$$
$$= \sqrt{\int_{U_s} \int_D \left(u_{s,h}(x, y) - u_{s,h,n}(x, y)\right)^2 dx \, dy},$$

where, observing that $u_{s,h} - u_{s,h,n}$ is jointly measurable, we interchanged the order of integration by appeal to the Fubini's theorem.

Theorem 4.4 *Under the assumptions* (A1), (A2) *and* (A6), *let* $u_{s,h}(\cdot, y) \in H_0^1(D)$ *denote the dimensionally truncated finite element solution of* (35) *for* $y \in U_s$ *and let* $q \in H^{-1}(D)$ *be the corresponding source term. Moreover, for every* $x \in D$ *let* $u_{s,h,n}(x, \cdot) := A_n^*(u_{s,h}(x, \cdot))$ *be the kernel interpolant at* x *based on a lattice rule satisfying the assumptions of Theorem 3.2. Suppose that* $\alpha \in 2\mathbb{N}$ *and* $\sigma := \frac{\alpha}{2}$. *Then we have for all* $\lambda \in (\frac{1}{\alpha}, 1]$ *that*

$$\sqrt{\int_{U_s} \int_D \left(u_{s,h}(x, y) - u_{s,h,n}(x, y)\right)^2 dx \, dy} \leq \frac{\kappa}{n^{1/(4\lambda)}} \frac{c_D \|q\|_{H^{-1}}}{a_{\min}} C_s(\lambda),$$

where $c_D > 0$ *is the Poincaré constant of the embedding* $H_0^1(D) \hookrightarrow L_2(D)$, $\kappa > 0$ *is the constant defined in Theorem 3.2, and*

$$[C_s(\lambda)]^{2\lambda} := \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \max(|\mathfrak{u}|, 1) \gamma_{\mathfrak{u}}^\lambda \left[2\zeta(\alpha\lambda)\right]^{|\mathfrak{u}|} \right)$$
$$\times \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{\gamma_{\mathfrak{u}}} \left(\sum_{m_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |m_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j)\right) \right)^2 \right)^\lambda. \tag{38}$$

Proof We can express the squared L_2 error as

$$\int_{U_s} \int_D \left(u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y})\right)^2 \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}$$

$$= \int_D \|u_{s,h}(\boldsymbol{x}, \cdot) - u_{s,h,n}(\boldsymbol{x}, \cdot)\|_{L_2(U_s)}^2 \, \mathrm{d}\boldsymbol{x}$$

$$\leq \int_D \left(e^{\mathrm{wor}}(A_n^*; L_2(U_s)) \|u_{s,h}(\boldsymbol{x}, \cdot)\|_{H(U_s)}\right)^2 \mathrm{d}\boldsymbol{x}$$

$$= [e^{\mathrm{wor}}(A_n^*; L_2(U_s))]^2 \int_D \|u_{s,h}(\boldsymbol{x}, \cdot)\|_{H(U_s)}^2 \, \mathrm{d}\boldsymbol{x}.$$

The first factor is the squared worst case L_2 approximation error, which can be bounded using Theorem 3.2. The second factor can be estimated using (2) by

$$\int_D \|u_{s,h}(\boldsymbol{x}, \cdot)\|_{H(U_s)}^2 \, \mathrm{d}\boldsymbol{x}$$

$$= \int_D \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{(2\pi)^{\alpha|\mathfrak{u}|} \gamma_{\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \left(\int_{[0,1]^{s-|\mathfrak{u}|}} \left(\prod_{j \in \mathfrak{u}} \frac{\partial^\sigma}{\partial y_j^\sigma}\right) u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y}_{-\mathfrak{u}}\right)^2 \mathrm{d}\boldsymbol{y}_{\mathfrak{u}} \mathrm{d}\boldsymbol{x}$$

$$\leq \int_D \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{(2\pi)^{\alpha|\mathfrak{u}|} \gamma_{\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \int_{[0,1]^{s-|\mathfrak{u}|}} \left[\left(\prod_{j \in \mathfrak{u}} \frac{\partial^\sigma}{\partial y_j^\sigma}\right) u_{s,h}(\boldsymbol{x}, \boldsymbol{y})\right]^2 \mathrm{d}\boldsymbol{y}_{-\mathfrak{u}} \, \mathrm{d}\boldsymbol{y}_{\mathfrak{u}} \, \mathrm{d}\boldsymbol{x}$$

$$= \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{(2\pi)^{\alpha|\mathfrak{u}|} \gamma_{\mathfrak{u}}} \int_{[0,1]^s} \left\|\left(\prod_{j \in \mathfrak{u}} \frac{\partial^\sigma}{\partial y_j^\sigma}\right) u_{s,h}(\cdot, \boldsymbol{y})\right\|_{L_2(D)}^2 \mathrm{d}\boldsymbol{y}$$

$$\leq c_D^2 \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{(2\pi)^{\alpha|\mathfrak{u}|} \gamma_{\mathfrak{u}}} \int_{[0,1]^s} \left\|\left(\prod_{j \in \mathfrak{u}} \frac{\partial^\sigma}{\partial y_j^\sigma}\right) u_{s,h}(\cdot, \boldsymbol{y})\right\|_{H_0^1(D)}^2 \mathrm{d}\boldsymbol{y}$$

$$\leq c_D^2 \frac{\|q\|_{H^{-1}(D)}^2}{a_{\min}^2} \sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{\gamma_{\mathfrak{u}}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j)\right)\right)^2,$$

where we used the Cauchy–Schwarz inequality, Fubini's theorem, the Poincaré constant $c_D > 0$ for the embedding $H_0^1(D) \hookrightarrow L_2(D)$, together with the PDE derivative bound (23) applied with $\boldsymbol{v} = (\sigma, \ldots, \sigma) = (\frac{\alpha}{2}, \ldots, \frac{\alpha}{2})$. The theorem is proved by combining the above expressions with Theorem 3.2. $\qquad\square$

Next, we proceed to choose the weights $\gamma_{\mathfrak{u}}$ and the parameters λ and α to ensure that the constant $C_s(\lambda)$ can be bounded independently of s, with λ as small as possible to yield the best possible convergence rate.

4.3.1 Choosing SPOD weights

One way to choose the weights is to equate the terms inside the two sums over $\mathfrak{u}$ in the formula (38) for $C_s(\lambda)$. (The value of $C_s(\lambda)$ so obtained minimizes (38) with respect to

$\gamma_{\mathfrak{u}}$ for $\mathfrak{u} \subseteq \{1 : s\}$.) It will be shown that this yields the convergence rate $\mathcal{O}(n^{-(\frac{1}{2p}-\frac{1}{4})})$ with an implied constant independent of the dimension s. The rate is precisely the rate of convergence that we expect to get. However, this choice of weights is too complicated to allow for efficient CBC construction of the lattice generating vector. So in the theorem below we propose a choice of SPOD weights that achieves the same error bound.

Theorem 4.5 *Assume that* (A1)–(A3) *and* (A6) *hold, and that p is as in* (A3). *Take* $\alpha := 2\lfloor \frac{1}{p} + \frac{1}{2} \rfloor$, $\sigma := \frac{\alpha}{2}$, $\lambda := \frac{p}{2-p}$, *and define the weights to be*

$$\gamma_{\mathfrak{u}} := \left(\frac{1}{\max(|\mathfrak{u}|, 1)\, [2\zeta(\alpha\lambda)]^{|\mathfrak{u}|}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j) \right) \right)^2 \right)^{\frac{1}{1+\lambda}} \tag{39}$$

for $\emptyset \neq \mathfrak{u} \subset \mathbb{N}$, $|\mathfrak{u}| < \infty$, *or SPOD weights*

$$\gamma_{\mathfrak{u}} := \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} (|\boldsymbol{m}_{\mathfrak{u}}|!)^{\frac{2}{1+\lambda}} \prod_{j \in \mathfrak{u}} \left(\frac{b_j^{m_j} S(\sigma, m_j)}{\sqrt{2e^{1/e}\zeta(\alpha\lambda)}} \right)^{\frac{2}{1+\lambda}} \tag{40}$$

for $\emptyset \neq \mathfrak{u} \subset \mathbb{N}$, $|\mathfrak{u}| < \infty$, *with* $\gamma_{\emptyset} := 1$. *Then the kernel interpolant of the finite element solution in Theorem* 4.4 *satisfies*

$$\sqrt{\int_{U_s} \int_D \left(u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}) \right)^2 d\boldsymbol{x}\, d\boldsymbol{y}} \leq C\, \|q\|_{H^{-1}(D)}\, n^{-(\frac{1}{2p}-\frac{1}{4})},$$

where the constant $C > 0$ *is independent of the dimension s.*

Proof We will proceed to justify the two choices of weights (39) and (40), and show that in both cases the term $C_s(\lambda)$ appearing in Theorem 4.4 can be bounded independently of s, by specifying λ and α as in the theorem.

The first choice of weights (39) is obtained by equating the terms inside the two sums over $\mathfrak{u}$ in the formula (38). Substituting (39) into (38) yields

$$[C_s(\lambda)]^{\frac{2\lambda}{1+\lambda}}$$

$$= \sum_{\mathfrak{u} \subseteq \{1:s\}} \left(\max(|\mathfrak{u}|, 1)\, [2\zeta(\alpha\lambda)]^{|\mathfrak{u}|} \right)^{\frac{1}{1+\lambda}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j) \right) \right)^{\frac{2\lambda}{1+\lambda}}$$

$$\leq \sum_{\mathfrak{u} \subseteq \{1:s\}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j) [2e^{1/e}\zeta(\alpha\lambda)]^{\frac{1}{2\lambda}} \right) \right)^{\frac{2\lambda}{1+\lambda}}$$

$$\leq \sum_{\mathfrak{u} \subseteq \{1:s\}} \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} \left(|\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \beta_j^{m_j} \right)^{\frac{2\lambda}{1+\lambda}} = \sum_{\boldsymbol{m} \in \{0:\sigma\}^s} \left(|\boldsymbol{m}|! \prod_{j=1}^s \beta_j^{m_j} \right)^{\frac{2\lambda}{1+\lambda}}, \tag{41}$$

where we used $\max(|\mathfrak{u}|, 1) \leq [e^{1/e}]^{|\mathfrak{u}|} = (1.4446\cdots)^{|\mathfrak{u}|}$, and defined $S_{\max}(\sigma) := \max_{1 \leq m \leq \sigma} S(\sigma, m)$, and $\beta_j := S_{\max}(\sigma) [2e^{1/e}\zeta(\alpha\lambda)]^{\frac{1}{2\lambda}} b_j$ for all $j \geq 1$, while applying Jensen's inequality[1] with $0 < \frac{2\lambda}{1+\lambda} \leq 1$.

The second choice of weights (40) is inspired by the weights (39) but takes the SPOD form

$$\gamma_{\mathfrak{u}} := \frac{1}{\tau^{|\mathfrak{u}|}} \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2}{1+\lambda}}, \quad V(\boldsymbol{m}_{\mathfrak{u}}) := |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j) \right), \quad (42)$$

with $\tau > 0$ to be specified below. (The $\tau^{|\mathfrak{u}|}$ factor can be merged into the product over $\mathfrak{u}$, thus giving SPOD weights.) Estimating $\max(|\mathfrak{u}|, 1) \leq [e^{1/e}]^{|\mathfrak{u}|}$ in (38), plugging in the weights (42), applying the Cauchy–Schwarz inequality with $\frac{1}{1+\lambda} + \frac{\lambda}{1+\lambda} = 1$, and applying Jensen's inequality with $0 < \lambda \leq 1$, we obtain from (38)

$$[C_s(\lambda)]^{2\lambda}$$

$$\leq \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda} [2e^{1/e}\zeta(\alpha\lambda)]^{|\mathfrak{u}|} \right) \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{1}{\gamma_{\mathfrak{u}}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{1}{1+\lambda}+\frac{\lambda}{1+\lambda}} \right)^2 \right)^{\lambda}$$

$$\leq \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \left(\frac{2e^{1/e}\zeta(\alpha\lambda)}{\tau^{\lambda}} \right)^{|\mathfrak{u}|} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2}{1+\lambda}} \right)^{\lambda} \right)$$

$$\times \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{\tau^{|\mathfrak{u}|}}{\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2}{1+\lambda}}} \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2}{1+\lambda}} \right)$$

$$\times \left(\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2\lambda}{1+\lambda}} \right) \right)^{\lambda},$$

and further

$$[C_s(\lambda)]^{2\lambda} \leq \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \left(\frac{2e^{1/e}\zeta(\alpha\lambda)}{\tau^{\lambda}} \right)^{|\mathfrak{u}|} \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2\lambda}{1+\lambda}} \right)$$

$$\times \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \tau^{|\mathfrak{u}|} \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2\lambda}{1+\lambda}} \right)^{\lambda}$$

$$= \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \tau^{|\mathfrak{u}|} \sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} [V(\boldsymbol{m}_{\mathfrak{u}})]^{\frac{2\lambda}{1+\lambda}} \right)^{1+\lambda},$$

where equality holds provided that we now choose $\tau := [2e^{1/e}\zeta(\alpha\lambda)]^{\frac{1}{1+\lambda}}$. This leads to the same upper bound (41) as for the first choice of weights.

---

[1] Jensen's inequality states that $\sum_i a_i \leq (\sum_i a_i^t)^{1/t}$ for all $a_i \geq 0$ and $t \in (0, 1]$.

It remains to show that the upper bound (41) can be bounded independently of s. We define the sequence $d_j := \beta_{\lceil j/\sigma \rceil}$ for $j \geq 1$, so that $d_1 = \cdots = d_\sigma = \beta_1$, $d_{\sigma+1} = \cdots = d_{2\sigma} = \beta_2$, and so on. Then for $\boldsymbol{m} \in \{0 : \sigma\}^s$ we can write

$$\prod_{j=1}^{s} \beta_j^{m_j} = \underbrace{\beta_1 \cdots \beta_1}_{m_1 \text{ factors}} \cdots \underbrace{\beta_s \cdots \beta_s}_{m_s \text{ factors}} = \prod_{j \in \mathfrak{v}} d_j,$$

where $\mathfrak{v} := \{1, 2, \ldots, m_1, \sigma + 1, \sigma + 2, \ldots, \sigma + m_2, \ldots, (s-1)\sigma + 1, \ldots, (s-1)\sigma + m_s\}$. Clearly, the set $\mathfrak{v}$ is of cardinality $|\mathfrak{v}| = m_1 + \cdots + m_s = |\boldsymbol{m}|$. It follows that

$$\sum_{\boldsymbol{m} \in \{0:\sigma\}^s} \left(|\boldsymbol{m}|! \prod_{j=1}^{s} \beta_j^{m_j} \right)^{\frac{2\lambda}{1+\lambda}} \leq \sum_{\substack{\mathfrak{v} \subset \mathbb{N} \\ |\mathfrak{v}| < \infty}} \left(|\mathfrak{v}|! \prod_{j \in \mathfrak{v}} d_j \right)^{\frac{2\lambda}{1+\lambda}}$$

$$= \sum_{\ell \geq 0} (\ell!)^{\frac{2\lambda}{1+\lambda}} \sum_{\substack{\mathfrak{v} \subset \mathbb{N} \\ |\mathfrak{v}| = \ell}} \prod_{j \in \mathfrak{v}} d_j^{\frac{2\lambda}{1+\lambda}} \leq \sum_{\ell \geq 0} (\ell!)^{\frac{2\lambda}{1+\lambda}} \frac{1}{\ell!} \Big(\underbrace{\sum_{j \geq 1} d_j^{\frac{2\lambda}{1+\lambda}}}_{=: T} \Big)^\ell. \tag{43}$$

The final inequality holds because $(\sum_{j \geq 1} d_j^{\frac{2\lambda}{1+\lambda}})^\ell$ includes all the products of the form $\prod_{j \in \mathfrak{v}} d_j^{\frac{2\lambda}{1+\lambda}}$ with $|\mathfrak{v}| = \ell$, and moreover includes each such term $\ell!$ times.

Recall from (24) and the assumption (A3) that $\sum_{j \geq 1} b_j^p < \infty$. We now choose

$$\frac{2\lambda}{1+\lambda} = p \iff \frac{1}{2\lambda} = \frac{1}{p} - \frac{1}{2} \iff \lambda = \frac{p}{2-p}.$$

For the inner sum in (43) we now have

$$T = \sum_{j \geq 1} d_j^p = \sigma \sum_{j \geq 1} \beta_j^p = \sigma \max\left(1, S_{\max}(\sigma) \left[2\zeta(\alpha\lambda)\right]^{\frac{1}{2\lambda}}\right)^p \sum_{j \geq 1} b_j^p < \infty,$$

provided that $\alpha\lambda > 1$, which is equivalent to $\alpha > \frac{2}{p} - 1$. This latter condition as well as the requirement that α be even can be satisfied by taking α such that $\frac{\alpha}{2} = \lfloor (\frac{1}{p} - \frac{1}{2}) + 1 \rfloor$, so we take $\alpha := 2\lfloor \frac{1}{p} + \frac{1}{2} \rfloor$. Finally, the ratio test implies convergence of the outer sum in (43), and consequently $C_s(\lambda)$ is bounded independently of s. Theorem 4.4 now ensures an error bound independent of s, and the convergence rate is $\mathcal{O}(n^{-(\frac{1}{2p} - \frac{1}{4})})$. This completes the proof. $\qquad\square$

4.3.2 Choosing POD weights

In the next theorem we prove that if the assumption (A3) holds for some $p \in \bigcup_{k=1}^{\infty} \left(\frac{2}{2k+1}, \frac{1}{k} \right)$, then it is possible to use POD weights to obtain the same rate of con-

vergence as in Theorem 4.5. For this and the next subsections we need the sequence of *Bell polynomials* (more precisely, *Touchard polynomials*), which we denote by

$$\mathrm{Bell}_\sigma(x) := \sum_{m=0}^{\sigma} S(\sigma, m)\, x^m, \quad \sigma \in \mathbb{N}_0,$$

where $S(\sigma, m)$ denotes the Stirling number of the second kind as before.

Theorem 4.6 *Assume that* (A1)–(A3), (A5) *and* (A6) *hold, and further assume that* $p \in \bigcup_{k=1}^{\infty} \left(\frac{2}{2k+1}, \frac{1}{k} \right)$ *in* (A3). *We take* $\alpha := 2\lfloor \frac{1}{p} \rfloor$, $\sigma := \frac{\alpha}{2}$, $\lambda := \frac{p}{2-p}$, *and define POD weights*

$$\gamma_{\mathfrak{u}} := \left(\frac{[(\sigma|\mathfrak{u}|)!]^2}{\max(|\mathfrak{u}|, 1)\,[2\zeta(\alpha\lambda)]^{|\mathfrak{u}|}} \prod_{j \in \mathfrak{u}} \mathrm{Bell}_\sigma(b_j)^2 \right)^{\frac{1}{1+\lambda}} \quad \textit{for}\ \emptyset \neq \mathfrak{u} \subseteq \{1 : s\}, \quad (44)$$

with $\gamma_\emptyset := 1$. *Then the kernel interpolant of the PDE solution in Theorem 4.4 satisfies*

$$\sqrt{\int_{U_s} \int_D \left(u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}) \right)^2 \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}} \leq C\, \|q\|_{H^{-1}(D)}\, n^{-\left(\frac{1}{2p} - \frac{1}{4} \right)},$$

where the constant $C > 0$ *is independent of the truncation dimension* s.

Proof In (38) we can apply the crude upper bound

$$\sum_{\boldsymbol{m}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} |\boldsymbol{m}_{\mathfrak{u}}|! \prod_{j \in \mathfrak{u}} \left(b_j^{m_j} S(\sigma, m_j) \right) \leq (\sigma|\mathfrak{u}|)! \prod_{j \in \mathfrak{u}} \left(\sum_{m=1}^{\sigma} b_j^m S(\sigma, m) \right)$$

$$= (\sigma|\mathfrak{u}|)! \prod_{j \in \mathfrak{u}} \mathrm{Bell}_\sigma(b_j),$$

which leads to

$$[C_s(\lambda)]^{2\lambda}$$
$$\leq \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \max(|\mathfrak{u}|, 1)\gamma_{\mathfrak{u}}^{\lambda}\,[2\zeta(\alpha\lambda)]^{|\mathfrak{u}|} \right) \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{[(\sigma|\mathfrak{u}|)! \prod_{j \in \mathfrak{u}} \mathrm{Bell}_\sigma(b_j)]^2}{\gamma_{\mathfrak{u}}} \right)^{\lambda}.$$
$$(45)$$

We equate the terms in the two sums in (45) to obtain the weights (44). Let us again define $S_{\max}(\sigma) := \max_{1 \leq m \leq \sigma} S(\sigma, m)$, so that $\mathrm{Bell}_\sigma(b_j) \leq S_{\max}(\sigma) \sum_{m=1}^{\sigma} b_j^m$. Plugging the weights back into (45) then yields

$$[C_s(\lambda)]^{\frac{2\lambda}{1+\lambda}} \leq \sum_{\mathfrak{u} \subseteq \{1:s\}} \left(\max(|\mathfrak{u}|, 1)\,[2\zeta(\alpha\lambda)]^{|\mathfrak{u}|} \right)^{\frac{1}{1+\lambda}} \left((\sigma|\mathfrak{u}|)! \prod_{j \in \mathfrak{u}} \mathrm{Bell}_\sigma(b_j) \right)^{\frac{2\lambda}{1+\lambda}}$$

$$\leq \sum_{\ell=0}^{\infty} \left(\max(\ell, 1) \, [2\zeta(\alpha\lambda)]^{\ell} \right)^{\frac{1}{1+\lambda}} \left((\sigma\ell)! \, [S_{\max}(\sigma)]^{\ell} \right)^{\frac{2\lambda}{1+\lambda}}$$

$$\times \underbrace{\sum_{\substack{u \subseteq \{1:s\} \\ |u|=\ell}} \prod_{j \in u} \left(\sum_{m=1}^{\sigma} b_j^m \right)^{\frac{2\lambda}{1+\lambda}}}_{=:V_\ell}.$$

To estimate V_ℓ, we have

$$V_\ell \leq \frac{1}{\ell!} \left(\sum_{j=1}^{\infty} \left(\sum_{m=1}^{\sigma} b_j^m \right)^{\frac{2\lambda}{1+\lambda}} \right)^{\ell} = \frac{1}{\ell!} \left(\sum_{j=1}^{\infty} \left(\sum_{m=1}^{\sigma} (1+b_1)^m \left(\frac{b_j}{1+b_1} \right)^m \right)^{\frac{2\lambda}{1+\lambda}} \right)^{\ell}$$

$$\leq \frac{1}{\ell!} \left((1+b_1)^{\frac{2\sigma\lambda}{1+\lambda}} \sum_{j=1}^{\infty} \left(\frac{b_j}{1+b_1-b_j} \right)^{\frac{2\lambda}{1+\lambda}} \right)^{\ell} \leq \frac{1}{\ell!} \left(\underbrace{(1+b_1)^{\frac{2\sigma\lambda}{1+\lambda}} \sum_{j=1}^{\infty} b_j^{\frac{2\lambda}{1+\lambda}}}_{=:T} \right)^{\ell},$$

where we estimated the sum over m by the geometric series formula and used $1 + b_1 - b_j \geq 1$ as a consequence of the assumption (A5).

In consequence, we have $[C_s(\lambda)]^{\frac{2\lambda}{1+\lambda}} \leq \sum_{\ell=0}^{\infty} a_\ell$, with

$$a_\ell := [\max(\ell, 1)]^{\frac{1}{1+\lambda}} \, [2\zeta(\alpha\lambda)]^{\frac{\ell}{1+\lambda}} \, [(\sigma\ell)!]^{\frac{2\lambda}{1+\lambda}} \, [S_{\max}(\sigma)]^{\frac{2\lambda\ell}{1+\lambda}} \frac{1}{\ell!} T^{\ell} > 0.$$

We can use the ratio test to determine sufficient conditions for the convergence of the infinite sum over ℓ. Letting $\ell > 0$, we find that

$$\frac{a_{\ell+1}}{a_\ell} = \left(\frac{\ell+1}{\ell} \right)^{\frac{1}{1+\lambda}} [2\zeta(\alpha\lambda)]^{\frac{1}{1+\lambda}} \, [(\sigma\ell+\sigma)\cdots(\sigma\ell+1)]^{\frac{2\lambda}{1+\lambda}} \, [S_{\max}(\sigma)]^{\frac{2\lambda}{1+\lambda}} \frac{T}{\ell+1}$$

$$\leq \left(\frac{\ell+1}{\ell} \right)^{\frac{1}{1+\lambda}} [2\zeta(\alpha\lambda)]^{\frac{1}{1+\lambda}} \, (\sigma\ell+\sigma)^{\frac{2\sigma\lambda}{1+\lambda}} \, [S_{\max}(\sigma)]^{\frac{2\lambda}{1+\lambda}} \frac{T}{\ell+1} \xrightarrow{\ell\to\infty} 0,$$

provided that $\frac{2\sigma\lambda}{1+\lambda} = \frac{\alpha\lambda}{1+\lambda} < 1$ and $\alpha\lambda > 1$. In conclusion, by choosing $\frac{2\lambda}{1+\lambda} = p \iff \lambda = \frac{p}{2-p}$, it follows from Theorem 4.4 that the convergence is independent of s with rate $\mathcal{O}(n^{-(\frac{1}{2p}-\frac{1}{4})})$, provided that

$$\frac{1}{\alpha} < \lambda < \frac{1}{\alpha-1} \iff \alpha - 1 < \frac{2}{p} - 1 < \alpha.$$

Unfortunately this condition cannot be fulfilled for all values of p, since $\alpha = 2\sigma$ needs to be an even integer. Indeed, the condition is equivalent to

$$\frac{2}{2\sigma + 1} < p < \frac{1}{\sigma} \iff \frac{1}{p} - \frac{1}{2} < \sigma < \frac{1}{p}.$$

Hence this condition is met if $p \in \bigcup_{k=1}^{\infty} \left(\frac{2}{2k+1}, \frac{1}{k} \right)$ by choosing $\alpha = 2\lfloor \frac{1}{p} \rfloor$. $\qquad \square$

The Lebesgue measure of the set of admissible values for p is precisely $\mu\left(\bigcup_{k=1}^{\infty} (\frac{2}{2k+1}, \frac{1}{k}) \right) = 2 - \log(4) \approx 0.61$. Nevertheless, even if $p \notin \bigcup_{k=1}^{\infty} \left(\frac{2}{2k+1}, \frac{1}{k} \right)$ we can always choose $\tilde{p} > p$ such that $\tilde{p} \in \bigcup_{k=1}^{\infty} \left(\frac{2}{2k+1}, \frac{1}{k} \right)$ and a correspondingly larger value of λ. The theorem then holds but with some loss in the rate of convergence.

4.3.3 Choosing product weights

In the next theorem we increase our error bounds to obtain product weights, which have the benefit of a lower computational cost (see Sect. 5), but with the disadvantage of a compromised theoretical convergence rate.

Theorem 4.7 *Assume that* (A1)–(A3), (A5) *and* (A6) *hold, and further assume that* $p < \frac{1}{2}$ *in* (A3). *If* $p \in \bigcup_{k=1}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}]$ *we take* $\alpha := 2\lfloor \frac{1}{2p} - \frac{1}{4} \rfloor$, $\sigma := \frac{\alpha}{2}$, *and* $\lambda := \frac{1}{2\sigma - 4\delta}$ *for arbitrary* $\delta \in (0, \frac{\sigma}{2} - \frac{1}{4})$. *If* $p \in (\frac{2}{5}, \frac{1}{2}) \cup \bigcup_{k=1}^{\infty} (\frac{2}{4k+5}, \frac{2}{4k+3})$ *we take* $\alpha := 2\lceil \frac{1}{2p} - \frac{1}{4} \rceil$, $\sigma := \frac{\alpha}{2}$, *and* $\lambda := \frac{1}{2/p - 1 - 2\sigma - 4\delta}$ *for arbitrary* $\delta \in (0, \frac{1}{2p} - \frac{1}{2} - \frac{\sigma}{2})$. *We define product weights*

$$\gamma_{\mathfrak{u}} := \prod_{j \in \mathfrak{u}} \left(\frac{\left[(j\sigma)^\sigma \operatorname{Bell}_\sigma (b_j) \right]^2}{2\mathrm{e}^{1/\mathrm{e}} \zeta(\alpha\lambda)} \right)^{\frac{1}{1+\lambda}} \qquad for \ \emptyset \neq \mathfrak{u} \subseteq \{1 : s\}, \qquad (46)$$

with $\gamma_\emptyset := 1$. *Then the kernel interpolant of the PDE solution in Theorem 4.4 satisfies*

$$\sqrt{\int_{U_s} \int_D \left(u_{s,h}(x, y) - u_{s,h,n}(x, y) \right)^2 \mathrm{d}x \, \mathrm{d}y}$$

$$\leq \begin{cases} C \, \|q\|_{H^{-1}(D)} \, n^{-(\frac{1}{2} \lfloor \frac{1}{2p} - \frac{1}{4} \rfloor - \delta)} & for \ p \in \bigcup_{k=1}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}], \\ C \, \|q\|_{H^{-1}(D)} \, n^{-(\frac{1}{2p} - \frac{1}{4} - \frac{1}{2} \lceil \frac{1}{2p} - \frac{1}{4} \rceil - \delta)} & for \ p \in (\frac{2}{5}, \frac{1}{2}) \cup \bigcup_{k=1}^{\infty} (\frac{2}{4k+5}, \frac{2}{4k+3}), \end{cases}$$

where the constant $C > 0$ *is independent of the truncation dimension* s.

Proof Starting again from the equation (45), we apply further crude upper bounds $\max(|\mathfrak{u}|, 1) \leq [\mathrm{e}^{1/\mathrm{e}}]^{|\mathfrak{u}|}$ and

$$(\sigma|\mathfrak{u}|)! = \prod_{j=1}^{|\mathfrak{u}|} \prod_{k=0}^{\sigma-1} (j\sigma - k) \leq \prod_{j=1}^{|\mathfrak{u}|} (j\sigma)^\sigma \leq \prod_{j \in \mathfrak{u}} (j\sigma)^\sigma,$$

to arrive at

$$[C_s(\lambda)]^{2\lambda} \leq \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda} \, [2\mathrm{e}^{1/\mathrm{e}} \zeta(\alpha\lambda)]^{|\mathfrak{u}|} \right) \left(\sum_{\mathfrak{u} \subseteq \{1:s\}} \frac{\prod_{j \in \mathfrak{u}} [(j\sigma)^{\sigma} \mathrm{Bell}_{\sigma}(b_j)]^2}{\gamma_{\mathfrak{u}}} \right)^{\lambda}.$$

$$(47)$$

We equate the terms in the two sums in (47) to obtain the product weights (46). Plugging the weights back into (47) and following the argument in the proof of Theorem 4.6, we obtain

$$
\begin{aligned}
[C_s(\lambda)]^{\frac{2\lambda}{1+\lambda}} &\leq \sum_{\mathfrak{u} \subseteq \{1:s\}} \prod_{j \in \mathfrak{u}} \left([2\mathrm{e}^{1/\mathrm{e}} \zeta(\alpha\lambda)]^{\frac{1}{1+\lambda}} \, [(j\sigma)^{\sigma} \mathrm{Bell}_{\sigma}(b_j)]^{\frac{2\lambda}{1+\lambda}} \right) \\
&\leq \sum_{\ell=0}^{\infty} [2\mathrm{e}^{1/\mathrm{e}} \zeta(\alpha\lambda)]^{\frac{\ell}{1+\lambda}} \left[\sigma^{\sigma} S_{\max}(\sigma) \right]^{\frac{2\lambda}{1+\lambda}\ell} \underbrace{\sum_{\substack{\mathfrak{u} \subseteq \{1:s\} \\ |\mathfrak{u}| = \ell}} \prod_{j \in \mathfrak{u}} \left(j^{\sigma} \sum_{m=1}^{\sigma} b_j^m \right)^{\frac{2\lambda}{1+\lambda}}}_{=: V_{\ell}},
\end{aligned}
$$

with

$$V_{\ell} \leq \frac{1}{\ell!} \left((1+b_1)^{\frac{2\sigma\lambda}{1+\lambda}} \sum_{j=1}^{\infty} (j^{\sigma} b_j)^{\frac{2\lambda}{1+\lambda}} \right)^{\ell},$$

Now one can easily check using the ratio test that the term $C_s(\lambda)$ can be bounded independently of s as long as the series $\sum_{j=1}^{\infty} (j^{\sigma} b_j)^{\frac{2\lambda}{1+\lambda}}$ is convergent.

From the monotonicity of $(b_j)_{j \geq 1}$ in the assumption (A5) it follows that $b_j \leq j^{-1/p} (\sum_{k=1}^{\infty} b_k^p)^{1/p}$ for all $j \geq 1$, implying

$$\sum_{j=1}^{\infty} (j^{\sigma} b_j)^{\frac{2\lambda}{1+\lambda}} \leq \left(\sum_{k=1}^{\infty} b_k^p \right)^{\frac{2\lambda p}{1+\lambda}} \sum_{j=1}^{\infty} j^{-\left(\frac{1}{p} - \sigma\right)\frac{2\lambda}{1+\lambda}},$$

which is finite provided that

$$\left(\frac{1}{p} - \sigma \right) \frac{2\lambda}{1+\lambda} > 1 \quad \Longleftrightarrow \quad \lambda > \frac{1}{\frac{2}{p} - 1 - 2\sigma}.$$

Taking into account also the requirement that $\frac{1}{\alpha} < \lambda \leq 1$ and that $\alpha = 2\sigma$ be an even integer, we have the constraint

$$\max \left(\frac{1}{2\sigma}, \frac{1}{\frac{2}{p} - 1 - 2\sigma} \right) < \lambda \leq 1.$$

$$(48)$$

We consider two scenarios below depending on the value of the maximum.

<u>Scenario A</u>. If $2\sigma \leq \frac{2}{p} - 1 - 2\sigma$ then $p \leq \frac{2}{4\sigma+1}$ and $\sigma \leq \frac{1}{2p} - \frac{1}{4}$, while the condition (48) simplifies to $\frac{1}{2\sigma} < \lambda \leq 1$. Since σ must be an integer and at least 1,

this scenario applies only when $p \in (0, \frac{2}{5}]$. In this case the best convergence rate is obtained by taking λ as close to $\frac{1}{2\sigma}$ as possible and σ as large as possible. Hence we take $\sigma := \lfloor \frac{1}{2p} - \frac{1}{4} \rfloor$ and $\lambda := \frac{1}{2\sigma - 4\delta}$ for arbitrary $\delta \in (0, \frac{\sigma}{2} - \frac{1}{4})$. By Theorem 4.4 this yields the convergence rate $\mathcal{O}(n^{-(\frac{1}{2}\lfloor \frac{1}{2p} - \frac{1}{4}\rfloor - \delta)})$ with the implied constant independent of the dimension s, but approaching ∞ as $\delta \to 0$.

<u>Scenario B</u>. On the other hand, if $2\sigma > \frac{2}{p} - 1 - 2\sigma$ then $p > \frac{2}{4\sigma + 1}$ and $\sigma > \frac{1}{2p} - \frac{1}{4}$, while the condition (48) becomes $\frac{1}{2/p - 1 - 2\sigma} < \lambda \leq 1$. Additionally, for the latter condition on λ to hold we require that $\frac{2}{p} - 1 - 2\sigma > 1$, which means $p < \frac{1}{\sigma + 1}$ and $\sigma < \frac{1}{p} - 1$. Combining all constraints we have

$$\frac{1}{\frac{2}{p} - 1 - 2\sigma} < \lambda \leq 1 \quad \text{and} \quad \frac{2}{4\sigma + 1} < p < \frac{1}{\sigma + 1} \quad \text{and} \quad \frac{1}{2p} - \frac{1}{4} < \sigma < \frac{1}{p} - 1.$$

Since σ must be an integer and at least 1, this scenario applies only when $p \in \bigcup_{k=1}^{\infty} (\frac{2}{4k+1}, \frac{1}{k+1}) = (0, \frac{1}{3}) \cup (\frac{2}{5}, \frac{1}{2})$. In this case the best convergence rate is obtained by taking λ close to $\frac{1}{2/p - 1 - 2\sigma}$ as possible but now with σ as small as possible. Hence we take $\sigma := \lceil \frac{1}{2p} - \frac{1}{4} \rceil$ and $\lambda := \frac{1}{2/p - 1 - 2\sigma - 4\delta}$ for arbitrary $\delta \in (0, \frac{1}{2p} - \frac{1}{2} - \frac{\sigma}{2})$. This yields the convergence rate $\mathcal{O}(n^{-(\frac{1}{2p} - \frac{1}{4} - \frac{1}{2}\lceil \frac{1}{2p} - \frac{1}{4}\rceil - \delta)})$, with the implied constant independent of the dimension s.

> If $p \in (\frac{2}{5}, \frac{1}{2})$ then only Scenario B applies.
> If $p \in [\frac{1}{3}, \frac{2}{5}]$ then only Scenario A applies.
> If $p \in (0, \frac{1}{3})$ then both scenarios apply, and it remains

to resolve which scenario to use in order to obtain the better convergence rate. For convenience we abbreviate $x := \frac{1}{2p} - \frac{1}{4}$ and $m := \lfloor \frac{1}{2p} - \frac{1}{4} \rfloor$, noting that $m \geq 1$ since $p < \frac{1}{3}$. Scenario B has a better convergence rate than Scenario A if and only if $\frac{1}{2}\lfloor x \rfloor < x - \frac{1}{2}\lceil x \rceil$. The latter condition is not satisfied if $x \in \mathbb{Z}$, while for $x \notin \mathbb{Z}$ the condition is equivalent to $\lfloor x \rfloor + \frac{1}{2} < x < \lceil x \rceil$. Hence the condition is equivalent to

$$m + \frac{1}{2} < \frac{1}{2p} - \frac{1}{4} < m + 1 \quad \Longleftrightarrow \quad \frac{2}{4m + 5} < p < \frac{2}{4m + 3}.$$

We conclude that for the case $p < \frac{1}{3}$ we should use Scenario B when $p \in \bigcup_{k=1}^{\infty} \left(\frac{2}{4k+5}, \frac{2}{4k+3} \right)$ and use Scenario A when $p \in [\frac{2}{7}, \frac{1}{3}) \cup \bigcup_{k=2}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}]$.

Combining the above analysis, we should apply Scenario B when $p \in (\frac{2}{5}, \frac{1}{2}) \cup \bigcup_{k=1}^{\infty} (\frac{2}{4k+5}, \frac{2}{4k+3})$ and apply Scenario A when $p \in [\frac{2}{7}, \frac{2}{5}] \cup \bigcup_{k=2}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}] = \bigcup_{k=1}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}]$. $\qquad \square$

4.4 Combined approximation error

The combined approximation error of the PDE problem (20) can be decomposed as

$$
\sqrt{\int_U \int_D (u(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}
$$

$$
\leq \sqrt{\int_U \int_D (u(\boldsymbol{x}, \boldsymbol{y}) - u_s(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}
$$

$$
+ \sqrt{\int_U \int_D (u_s(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h}(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}
$$

$$
+ \sqrt{\int_U \int_D (u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}},
$$

where the first term is the dimension truncation error, the second term is the finite element error, and the final term is the kernel interpolation error. Combining the results developed in Sects. 4.1–4.3, we arrive at the following result.

Theorem 4.8 *Assume that* (A1)–(A6) *hold. For any* $\boldsymbol{y} \in U$, *let* $u(\cdot, \boldsymbol{y}) \in H_0^1(D)$ *denote the solution to (20) with the source term* $q \in H^{-1+t}(D)$ *for some* $0 \leq t \leq 1$. *Let* $u_{s,h}(\cdot, \boldsymbol{y}) \in V_h$ *be the corresponding dimensionally truncated finite element solution and let* $u_{s,h,n}(\boldsymbol{x}, \cdot) = A_n^*(u_{s,h}(\boldsymbol{x}, \cdot))$ *be its kernel interpolant constructed using the weights described in Theorems 4.5, 4.6, or 4.7. Then we have the combined error estimate*

$$
\sqrt{\int_U \int_D (u(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}))^2 \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}
$$

$$
\leq C \left(\left(s^{-(\frac{1}{p} - \frac{1}{2})} + n^{-r} \right) \|q\|_{H^{-1}(D)} + h^{1+t} \|q\|_{H^{-1+t}(D)} \right),
$$

where $0 \leq t \leq 1$, *h denotes the mesh size of the piecewise linear finite element mesh, $C > 0$ is a constant independent of s, h, n, q, and*

$$
r = \begin{cases}
\frac{1}{2p} - \frac{1}{4} & \text{with weights (39) or SPOD weights (40),} \\
\frac{1}{2p} - \frac{1}{4} & \text{with POD weights (44) for } p \in \bigcup_{k=1}^{\infty} (\frac{2}{2k+1}, \frac{1}{k}), \\
\frac{1}{2} \lfloor \frac{1}{2p} - \frac{1}{4} \rfloor - \delta & \text{with product weights (46) for} \\
& \quad p \in \bigcup_{k=1}^{\infty} [\frac{2}{4k+3}, \frac{2}{4k+1}], \\
\frac{1}{2p} - \frac{1}{4} - \frac{1}{2} \lceil \frac{1}{2p} - \frac{1}{4} \rceil - \delta & \text{with product weights (46) for} \\
& \quad p \in (\frac{2}{5}, \frac{1}{2}) \cup \bigcup_{k=1}^{\infty} (\frac{2}{4k+5}, \frac{2}{4k+3}),
\end{cases}
$$

and $\delta > 0$ *is sufficiently small in each case.*

5 Cost analysis

5.1 What is the point set at which values are wanted?

In this section we consider the cost of evaluating the kernel interpolant

$$f_n(\mathbf{y}) = \sum_{k=1}^{n} a_k \, K(\mathbf{t}_k, \mathbf{y}),$$

as an approximation to the periodic function f, with lattice points $\mathbf{t}_k = \{\frac{k\mathbf{z}}{n}\}$, $k = 1, \ldots, n$, and $\mathbf{t}_n = \mathbf{t}_0 = \mathbf{0}$. Recall that all our functions including the kernel are 1-periodic with respect to $\mathbf{y}$. For the linear system (7), as observed already, the matrix $\mathcal{K} = [K(\mathbf{t}_k - \mathbf{t}_{k'}, \mathbf{0})]_{k,k'=1,\ldots,n}$ is circulant, thus we need to compute only its first column (see the cost for evaluating the kernel in the next subsection) and then solve for the coefficients a_k with a cost of $\mathcal{O}(n \log(n))$.

First, however, it turns out to be useful to ask: what is the set of points, say $\{\mathbf{y}_1, \mathbf{y}_2, \ldots\}$, at which the values of the interpolant are desired? If L such points $\mathbf{y}_\ell$, $\ell = 1, \ldots, L$, are chosen arbitrarily then the cost, naturally, is L times the cost of a single evaluation. On the other hand, for a set of Ln points formed by the *union of shifted lattices* $\mathbf{y}_\ell + \mathbf{t}_{k'}$, $\ell = 1, \ldots, L$, $k' = 1, \ldots n$, it turns out that the cost for Ln evaluations is little more than the cost of the L evaluations at arbitrary points.

The reason for the low cost lies in the shift invariance of the kernel and the group nature of the lattice. For a single given $\mathbf{y}$ the principal costs for evaluating the kernel interpolant come from evaluating $K(\mathbf{t}_k, \mathbf{0})$ and $f(\mathbf{t}_k)$ at the n lattice points; then solving the circulant linear system (7) for the n values of a_k; from evaluating $K(\mathbf{t}_k, \mathbf{y})$ at the n lattice points; and finally from assembling $f_n(\mathbf{y})$ with a cost of $\mathcal{O}(n)$. (The precise cost breakdown is given in Table 1 below after we discuss the cost for evaluating the kernel in the next subsection.)

But for evaluation of $K(\mathbf{t}_k, \mathbf{y} + \mathbf{t}_{k'})$ for all n values $k' = 1, \ldots, n$ we observe that $K(\mathbf{t}_k, \mathbf{y} + \mathbf{t}_{k'}) = K(\mathbf{t}_k - \mathbf{t}_{k'}, \mathbf{y})$, and hence

$$f_n(\mathbf{y} + \mathbf{t}_{k'}) = \sum_{k=1}^{n} a_k \, K(\mathbf{t}_k - \mathbf{t}_{k'}, \mathbf{y}). \tag{49}$$

Since the right-hand side has the form of a circulant $n \times n$ matrix multiplying a vector of length n, the n values $f_n(\mathbf{y} + \mathbf{t}_{k'})$ for $k' = 1, \ldots, n$ can be assembled with a cost of $\mathcal{O}(n \log(n))$, compared with the $\mathcal{O}(n)$ cost of assembling f_n at a single value of $\mathbf{y}$.

5.2 Cost for evaluating the kernel for a single y

Now consider the cost of computing $K(\mathbf{t}, \mathbf{y})$ for a single arbitrary value of $\mathbf{y}$ and arbitrary $\mathbf{t}$,

$$K(t, y) = \sum_{\mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}} \prod_{j \in \mathfrak{u}} \eta_\alpha(t_j, y_j) \qquad \text{for} \quad k = 1, \ldots, n.$$

In the following, we assume that evaluating η_α can be treated as having constant cost. For example, when α is even we have an analytic formula for η_α in terms of the Bernoulli polynomial.

If the weights have no special structure then the cost to evaluate $K(t, y)$ would be exponential in s because of the sum over subsets of $\{1 : s\}$, but the cost is much reduced in special cases:

- With product weights we have $K(t, y) = \prod_{j=1}^{s}(1 + \gamma_j \eta_\alpha(t_j, y_j))$, which can be evaluated for a pair (t, y) at the cost of $\mathcal{O}(s)$.
- With POD weights we have

$$K(t, y) = \sum_{\mathfrak{u} \subseteq \{1:s\}} \Gamma_{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \left(\gamma_j \, \eta_\alpha(t_j, y_j) \right) = \sum_{\ell=0}^{s} \Gamma_\ell \underbrace{\sum_{\substack{\mathfrak{u} \subseteq \{1:s\} \\ |\mathfrak{u}| = \ell}} \prod_{j \in \mathfrak{u}} \left(\gamma_j \, \eta_\alpha(t_j, y_j) \right)}_{=: P_{s,\ell}},$$

where $P_{s,\ell}$ is defined for $\ell = 0, \ldots, s$, and can be computed recursively using

$$P_{s,\ell} = P_{s-1,\ell} + \gamma_s \, \eta_\alpha(t_s, y_s) \, P_{s-1,\ell-1},$$

together with $P_{s,0} := 1$ for all s and $P_{s,\ell} := 0$ for all $\ell > s$. The cost to evaluate this for a pair (t, y) is $\mathcal{O}(s^2)$.

- With SPOD weights we have

$$\begin{aligned}
K(t, y) &= \sum_{\mathfrak{u} \subseteq \{1:s\}} \sum_{\boldsymbol{v}_{\mathfrak{u}} \in \{1:\sigma\}^{|\mathfrak{u}|}} \Gamma_{|\boldsymbol{v}_{\mathfrak{u}}|} \prod_{j \in \mathfrak{u}} \left(\gamma_{j,v_j} \, \eta_\alpha(t_j, y_j) \right) \\
&= \sum_{\boldsymbol{v} \in \{0:\sigma\}^s} \Gamma_{|\boldsymbol{v}|} \prod_{j : v_j > 0} \left(\gamma_{j,v_j} \, \eta_\alpha(t_j, y_j) \right) \\
&= \sum_{\ell=0}^{s\sigma} \Gamma_\ell \underbrace{\sum_{\substack{\boldsymbol{v} \in \{0:\sigma\}^s \\ |\boldsymbol{v}| = \ell}} \prod_{j : v_j > 0} \left(\gamma_{j,v_j} \, \eta_\alpha(t_j, y_j) \right)}_{=: P_{s,\ell}},
\end{aligned}$$

where $P_{s,\ell}$ is now defined for $\ell = 0, \ldots, s\sigma$, and can be computed recursively using

$$P_{s,\ell} = P_{s-1,\ell} + \eta_\alpha(t_s, y_s) \sum_{v=1}^{\min(\sigma,\ell)} \gamma_{s,v} \, P_{s-1,\ell-v},$$

together with $P_{s,0} := 1$ for all s and $P_{s,\ell} := 0$ for all $\ell > s\sigma$. The cost to evaluate this for a pair (t, y) is now $\mathcal{O}(s^2 \sigma^2)$.

Table 1 Cost breakdown for the kernel interpolant f_n based on n lattice points t_k in s dimensions, evaluated at L arbitrary points y_ℓ. Here X is the cost for one evaluation of f.

Operation \ Weights	Product	POD	SPOD
Fast CBC construction for z	$s\,n\log(n)$	$s\,n\log(n) + s^2\log(s)\,n$	$s\,n\log(n) + s^3\sigma^2\,n$
Compute $K(t_k, \mathbf{0})$ for all k	$s\,n$	$s^2\,n$	$s^2\sigma^2\,n$
Evaluate $f(t_k)$ for all k	$X\,n$	$X\,n$	$X\,n$
Linear solve for all coefficients a_k	$n\log(n)$	$n\log(n)$	$n\log(n)$
Compute $K(t_k, y_\ell)$ for all k, ℓ	$s\,n\,L$	$s^2\,n\,L$	$s^2\sigma^2\,n\,L$
Assemble $f_n(y_\ell)$ for all ℓ	$n\,L$	$n\,L$	$n\,L$
OR Assemble $f_n(y_\ell + t_k)$ for all ℓ, k	$n\,\log(n)\,L$	$n\,\log(n)\,L$	$n\,\log(n)\,L$

Table 2 Cost breakdown for the kernel interpolant $u_{s,h,n}$ based on n lattice points t_k in s dimensions, evaluated at M finite element nodes x_i and L arbitrary points y_ℓ. Here M^a for some positive a is the cost for one finite element solve with M nodes.

Operation \ Weights	Product	POD	SPOD
Fast CBC construction for z	$s\,n\log(n)$	$s\,n\log(n) + s^2\log(s)\,n$	$s\,n\log(n) + s^3\sigma^2\,n$
Compute $K(t_k, \mathbf{0})$ for all k	$s\,n$	$s^2\,n$	$s^2\sigma^2\,n$
Evaluate $u_{s,h}(x_i, t_k)$ for all i, k	$M^a\,n$	$M^a\,n$	$M^a\,n$
Linear solve for all coeff. $a_k(x_i)$	$M\,n\log(n)$	$M\,n\log(n)$	$M\,n\log(n)$
Compute $K(t_k, y_\ell)$ for all k, ℓ	$s\,n\,L$	$s^2\,n\,L$	$s^2\sigma^2\,n\,L$
Assemble $u_{s,h,n}(x_i, y_\ell)$ for all i, ℓ	$M\,n\,L$	$M\,n\,L$	$M\,n\,L$
OR Assemble $u_{s,h,n}(x_i, y_\ell + t_k)$ for all i, ℓ, k	$M\,n\,\log(n)\,L$	$M\,n\,\log(n)\,L$	$M\,n\,\log(n)\,L$

5.3 Cost for the kernel interpolant

We now summarize the cost for the kernel interpolant and different weights using the results of the preceding two subsections. Let X denote the cost for one evaluation of f. The cost breakdown is shown in Table 1. The first four rows are considered to be pre-computation cost while the last three rows are the running cost for sampling. The cost for the fast CBC construction based on the criterion $\mathcal{S}_s(z)$ with different weight parameters is analyzed in [4].

For the PDE application, our kernel method is

$$u_{s,h}(x_i, y) \approx u_{s,h,n}(x_i, y) = \sum_{k=1}^{n} a_k(x_i)\, K(t_k, y),$$

where $\{x_i : i = 1, \ldots, M\} \subset D$ is the set of finite element nodes in the physical domain, and $a_k(x_i)$ for $k = 1, \ldots n$ is the solution for fixed x_i of the linear system

$$\sum_{k=1}^{n} \mathcal{K}_{k,k'} \, a_k(x_i) = u_{s,h}(x_i, t_{k'}), \quad k' = 1, \ldots, n.$$

Let M^a for some $a \geq 1$ denote the cost of the finite element solve to obtain all x_i for one y. The cost breakdown for obtaining the kernel interpolant at all M nodes for all L samples is shown in Table 2. Note in this case that the coefficients $a_k(x_i)$ need to be computed for every finite element node x_i, hence the scaling of the cost in line 4 of Table 2 by M. If the quantity of interest is a *linear functional* of the PDE finite element solution (no need for the solution at every node), then the cost is reduced to be as in Table 1 with $X = M^a$.

6 Numerical experiments

We consider the parametric PDE problem (16)–(17) in the physical domain $D = (0, 1)^2$ with the source term $q(x) = x_2$ and the diffusion coefficient periodic in the parameters y given by (21), with $a_0(x) = 1$ for $x \in D$.

For each fixed $y \in U_s$ (i.e. with the sum in (21) truncated to s terms), we solve the PDE using a piecewise linear finite element method with $h = 2^{-5}$ as the finite element mesh size. As the stochastic fluctuations, we consider the functions

$$\psi_j(x) := c \, j^{-\theta} \sin(j\pi x_1) \sin(j\pi x_2), \qquad x = (x_1, x_2) \in D, \ j \geq 1,$$

where $c > 0$ is a constant, $\theta > 1$ is the decay rate of the stochastic fluctuations, and $s \in \mathbb{N}$ is the truncation dimension. Following (24), the sequence $(b_j)_{j\geq 1}$ is taken to be

$$b_j := \frac{c \, j^{-\theta}}{\sqrt{6} \, a_{\min}}, \quad \text{with} \ a_{\min} := 1 - \frac{c}{\sqrt{6}} \zeta(\theta) \ \text{as well as} \ a_{\max} := 1 + \frac{c}{\sqrt{6}} \zeta(\theta),$$

and $c < \frac{\sqrt{6}}{\zeta(\theta)}$, ensuring that the assumption (A2) is satisfied.

We approximate the dimensionally truncated finite element solution $u_{s,h}$ of the PDE (16)–(17) by constructing a kernel interpolant $u_{s,h,n}(x, y) := A_n^*(u_{s,h}(x, y))$, $x \in D$ and $y \in U_s$ using SPOD weights, POD weights, and product weights chosen according to Theorems 4.5, 4.6, and 4.7, respectively. The same weights appear in the formula for the kernel as well as the search criterion for finding good lattice generating vectors. The kernel interpolant is constructed over a lattice point set $t_k := \{kz/n\}$, $k \in \{1, \ldots, n\}$, where the generating vector $z \in \{1, \ldots, n-1\}^s$ has been obtained separately for each weight type using the fast CBC algorithm detailed in [4]. We assess the kernel interpolation error by computing
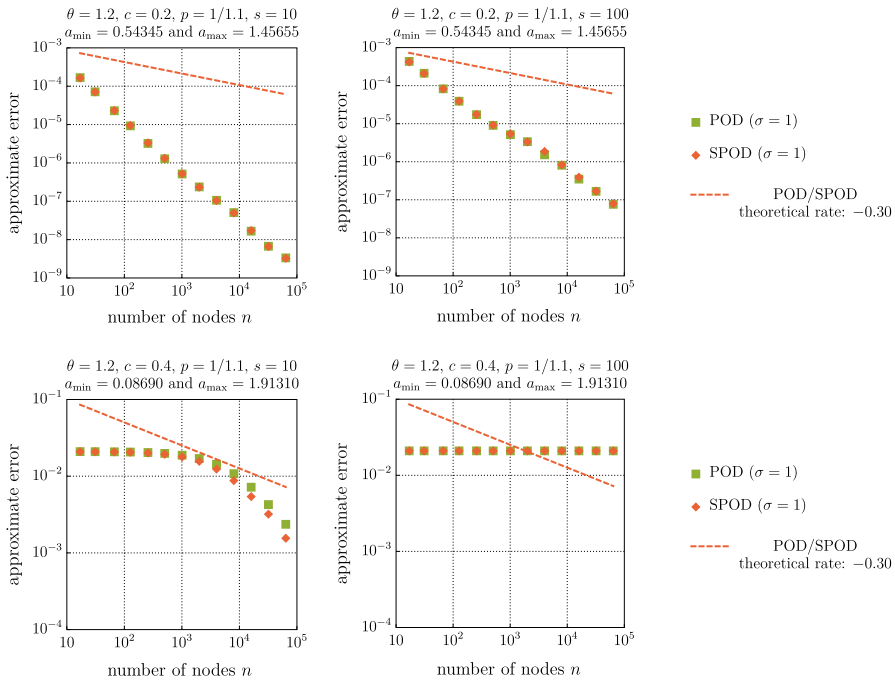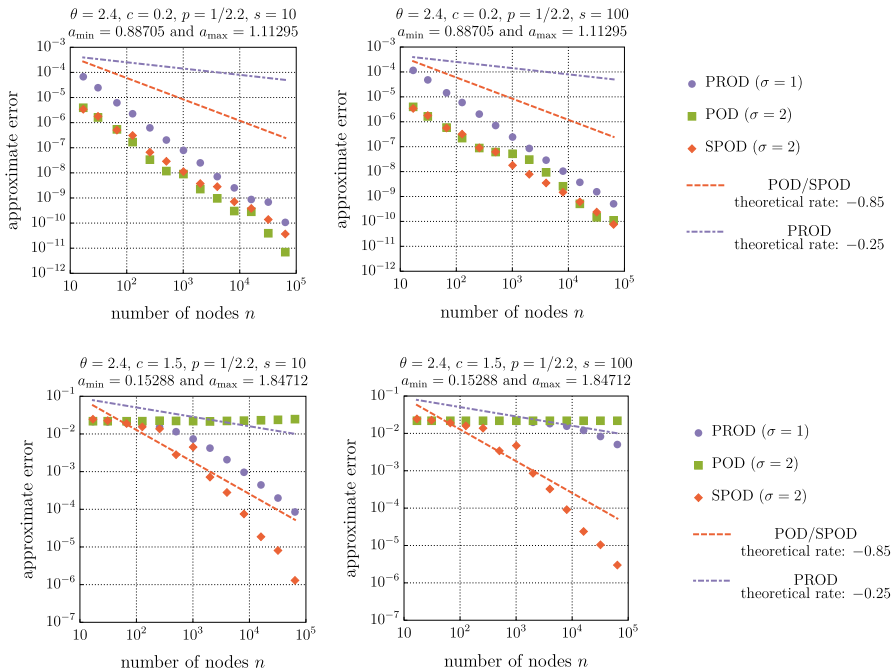
Fig. 1 The kernel interpolation errors of the PDE problem (16)–(17) with $\theta = 1.2$, $p = 1/1.1$, $c \in \{0.2, 0.4\}$, and $s \in \{10, 100\}$. Results are displayed for kernel interpolants constructed using POD and SPOD weights. (Product weights (46) are not well-defined in this case)

$$\text{error} = \sqrt{\int_{U_s} \int_D \left(u_{s,h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}) \right)^2 \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{y}}$$

$$\approx \sqrt{\frac{1}{Ln} \sum_{\ell=1}^{L} \sum_{k=1}^{n} \int_D \left(u_{s,h}(\boldsymbol{x}, \boldsymbol{y}_\ell + \boldsymbol{t}_k) - u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y}_\ell + \boldsymbol{t}_k) \right)^2 \mathrm{d}\boldsymbol{x}},$$

where $\boldsymbol{y}_\ell$ for $\ell = 1, \ldots, L$ is a sequence of Sobol′ nodes in $[0, 1]^s$, with $L = 100$, and we recall that all our functions including $u_{s,h}(\boldsymbol{x}, \boldsymbol{y})$ and $u_{s,h,n}(\boldsymbol{x}, \boldsymbol{y})$ are 1-periodic with respect to $\boldsymbol{y}$. The kernel interpolant in the formula above can be evaluated efficiently over the union of shifted lattices $\boldsymbol{y}_\ell + \boldsymbol{t}_k$, $\ell = 1, \ldots, L$, $k = 1, \ldots, n$, by making use of formula (49) in conjunction with the fast Fourier transform, requiring only the evaluation of the values $K(\boldsymbol{t}_k, \boldsymbol{y}_\ell)$.

We compute the approximation error when $\theta \in \{1.2, 2.4, 3.6\}$, choosing $p \in \{\frac{1}{1.1}, \frac{1}{2.2}, \frac{1}{3.3}\}$, respectively, which are all p values ensuring that (A3) is satisfied. We also use several values of the parameter $c \in \{0.2, 0.4, 1.5\}$ to control the difficulty of the problem. We set $\delta = 0.1$ in the product weights (46). The numerical experiments have been carried out by using both $s = 10$ and $s = 100$ as the truncation dimensions. Selected results are displayed in Figures 1, 2, 3, where the corresponding values of $a_{\min}$ and $a_{\max}$ are listed to give insights to the difficulty of the problem in each case,
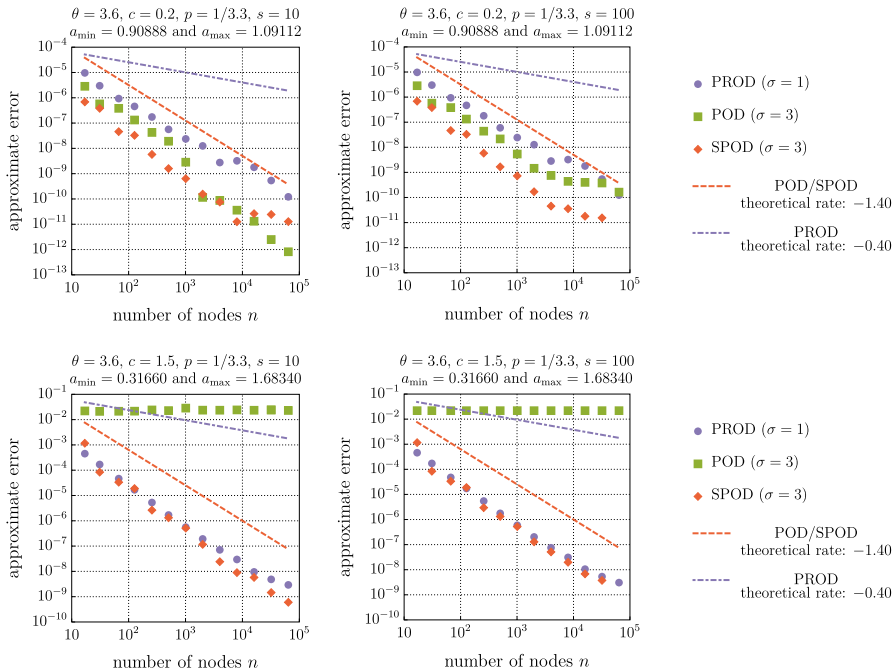
Fig. 2 The kernel interpolation errors of the PDE problem (16)–(17) with $\theta = 2.4$, $p = 1/2.2$, $c \in \{0.2, 1.5\}$, and $s \in \{10, 100\}$. Results are displayed for kernel interpolants constructed using product (PROD), POD, and SPOD weights

as well as the parameter σ which shows the "order" of the lattice rule. Note that as s increases the problem does not change, but the computation becomes harder because the diffusion coefficient takes a wider range of values, with small values of $a(\boldsymbol{x}, \boldsymbol{y})$ being especially challenging.

The empirically obtained convergence rates appear to exceed the theoretically expected rates once the kernel interpolant enters the asymptotic regime of convergence. The convergence behavior of the kernel interpolant with SPOD weights is good across all experiments, except for the most difficult PDE problem of the lot corresponding to parameters $\theta = 1.2$ and $c = 0.4$, illustrated in the bottom row of Fig. 1. On the other hand, the POD weights and, to a lesser extent, the product weights appear to be somewhat sensitive to the effective dimension of the PDE problem, either leading to a longer pre-asymptotic regime compared to SPOD weights (see "PROD" in the bottom row of Fig. 2) or no apparent convergence (see "POD" in the bottom rows of Figs. 2 and 3).

In the top graph of Fig. 4 we compare the results in Figs. 1, 2, 3 from SPOD weights with truncation dimension $s = 100$ for the same damping parameter $c = 0.2$ and different $\theta \in \{1.2, 2.4, 3.6\}$, listing the estimated convergence rate in each case. In the middle graph of Fig. 4 we show the results of an additional experiment, namely, that for $s = 100$ where we fix the decay rate $\theta = 3.6$ of the stochastic fluctuations and solve the parametric PDE problem using different $\sigma \in \{1, 2, 3\}$ in the formula for

Fig. 3 The kernel interpolation errors of the PDE problem (16)–(17) with $\theta = 3.6$, $p = 1/3.3$, $c \in \{0.2, 1.5\}$, and $s \in \{10, 100\}$. Results are displayed for kernel interpolants constructed using product (PROD), POD, and SPOD weights

SPOD weights, which correspond to $p \in \{\frac{1}{1.1}, \frac{1}{2.2}, \frac{1}{3.3}\}$, see Theorem 4.5. Finally, in the bottom graph of Fig. 4, we return to the experimental setup illustrated in Fig. 1 except this time we carry out the experiment using the truncation dimensions $s \in \{10, 20, 40, 80, 160\}$.

In all cases displayed in Fig. 4, the observed error decays faster than the rate implied by Theorem 4.5. We also see that increasing σ improves the error and mildly improves the rate of convergence. Moreover, we observe from the graph in the middle that the parameter θ that governs the decay of $\|\psi_j\|_{L_\infty(D)}$ is more important in determining the rate than the choice of σ. This observation suggests that the kernel interpolation with the rank-1 lattice points are robust in σ. Notice that σ appears in the definition (40) of the SPOD weights; and that the weights are an input of the CBC construction, and are used to define the kernel $K(\cdot, \cdot)$. These observed error decay rates and the robustness are encouraging, but also suggest that the worst-case error estimates may be pessimistic in practical situations. The bottom graph in Fig. 4 illustrates the effect that the truncation dimension has on the obtained convergence rates: we see that the observed convergence rate remains reasonable even when $s = 160$.

Finally, we present numerical experiments that assess the dimension truncation error rate given in Theorem 4.1. We consider the same PDE and stochastic fluctuations $(\psi_j)_{j\geq 1}$ which were stated at the beginning of this section. We choose the parameters $c = 0.4$ with $\theta \in \{1.2, 2.4, 3.6\}$ and $c = 1.5$ with $\theta \in \{2.4, 3.6\}$. The PDE is discretized using piecewise linear finite element method with mesh size $h = 2^{-6}$
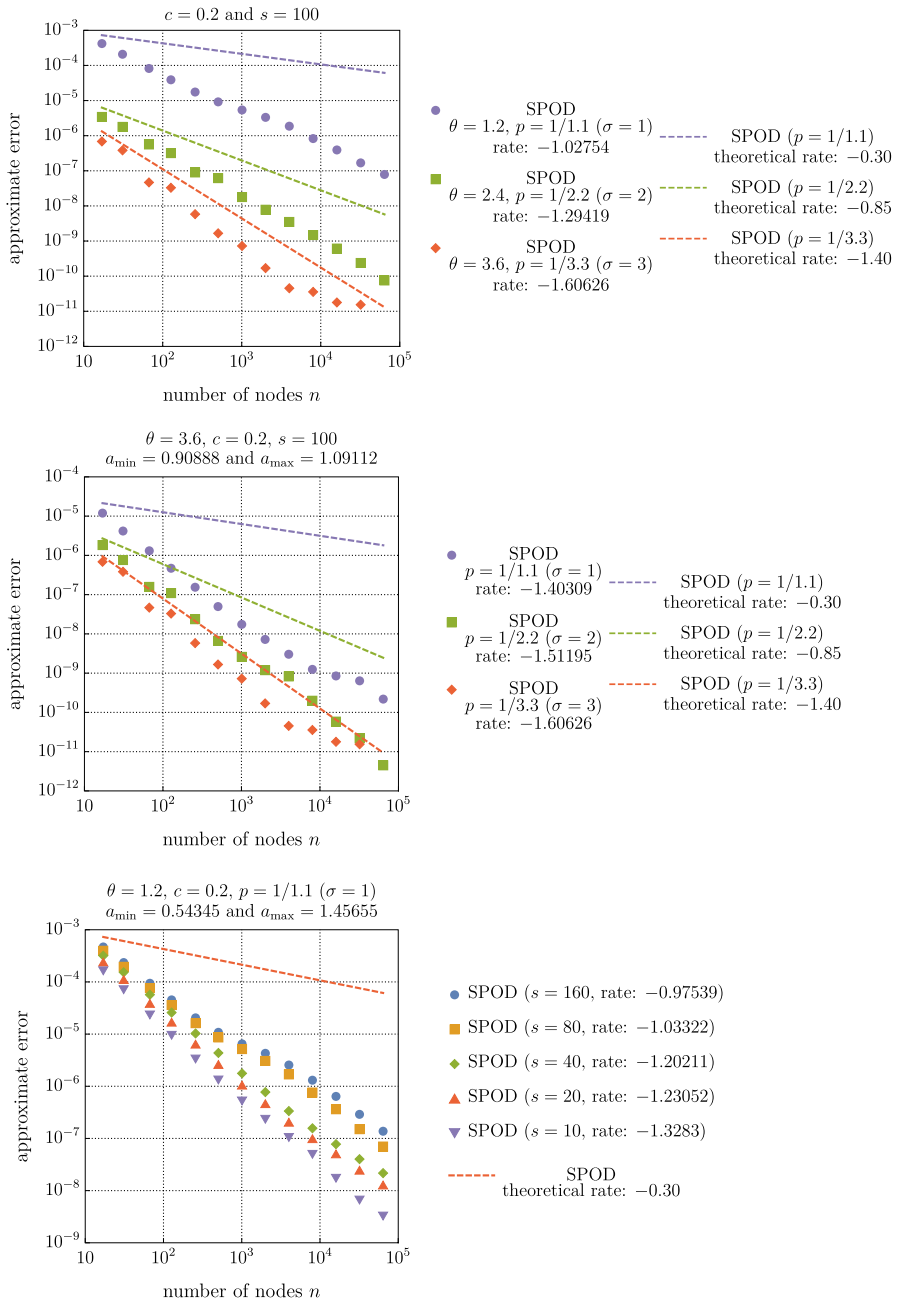
Fig. 4 The kernel interpolation errors of the PDE problem (16)–(17) for kernel interpolants constructed using SPOD weights and varying parameters. Top: *fixed* $s = 100$ and $c = 0.2$ and different values of θ. Middle: *fixed* $s = 100$ and $\theta = 3.6$, different values of p, and corresponding $\sigma = \sigma(p)$. Theoretical error-decay rate is $-\frac{1}{2p} + \frac{1}{4} = -0.3, -0.85, -1.4$ for $p = \frac{1}{1.1}, \frac{1}{2.2}, \frac{1}{3.3}$. Bottom: *fixed* $\theta = 1.2$, $c = 0.2$, $p = 1/1.1$, and $\sigma = 1$ with different values of $s \in \{10, 20, 40, 80, 160\}$
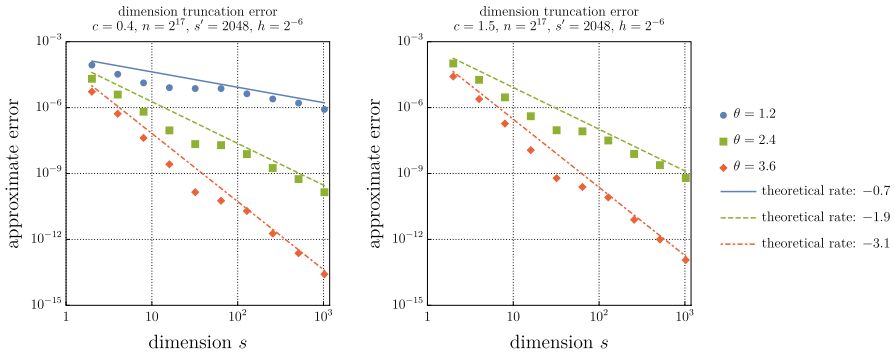
Fig. 5 The dimension truncation errors of the PDE problem (16)–(17). Left: $c = 0.4$ and $\theta \in \{1.2, 2.4, 3.6\}$. Right: $c = 1.5$ and $\theta \in \{2.4, 3.6\}$

and the integral over the computational domain D is computed exactly for the finite element solutions. As the reference solution, we use the finite element solution with truncation dimension $s' = 2^{11}$. The dimension truncation error is then estimated by computing

$$\sqrt{\int_{U_{s'}} \int_D (u_{s',h}(\boldsymbol{x}, \boldsymbol{y}) - u_{s,h}(\boldsymbol{x}, \boldsymbol{y}))^2 \, d\boldsymbol{x} \, d\boldsymbol{y}}$$

for $s = 2^k$, $k = 2, \ldots, 10$, where the value of the parametric integral is computed approximately by means of a rank-1 lattice rule based on the off-the-shelf generating vector `lattice-39101-1024-1048576.3600` downloaded from https://web.maths.unsw.edu.au/~fkuo/lattice/ with $n = 2^{17}$ nodes. The results are displayed in Fig. 5. The theoretically expected rate, which is essentially $\mathcal{O}(s^{-\theta+1/2})$, is clearly observed in all cases.

7 Conclusions

In this paper we have developed an approximation scheme for periodic multivariate functions based on kernel approximation at lattice points, in the setting of weighted Hilbert spaces of dominating mixed smoothness. We have developed L_2 error estimates that are independent of dimension, for three classes of weights: product weights, POD (product and order dependent) weights and SPOD (smoothness driven product and order dependent) weights. Numerical experiments for 10 and 100 dimensions give results that (with the possible exception of POD weights) are generally satisfactory, and that exhibit better than predicted rates of convergence.

Nevertheless, there is room for future improvement. First, the error analysis is based on the principle that the L_2 error is bounded above by the worst case L_2 error multiplied by the norm of the function being approximated; yet it is known (see Sect. 3.1) that the worst-case error has a poor rate of convergence. It may be possible to obtain improved error rates by making better use of the special properties of the minimum norm interpolant in conjunction with the analytic parameter dependence of the PDE solution of (16)–(17).

References

1. Byrenheid, G., Kämmerer, L., Ullrich, T., Volkmer, T.: Tight error bounds for rank-1 lattice sampling in spaces of hybrid mixed smoothness. Numer. Math. **136**, 993–1034 (2017)
2. Ciarlet, P.G.: The Finite Element Method for Elliptic Problems. North-Holland Publishing Company, Amsterdam (1978)
3. Cools, R., Kuo, F. Y., Nuyens, D., Sloan, I. H.: Lattice algorithms for multivariate approximation in periodic spaces with general weight parameters, In: Brenner, S. C., Shparlinski, I., Shu, C.-W., Szyld, D. (eds.) Celebrating 75 Years of Mathematics of Computation, Contemporary Mathematics, 754, AMS, pp. 93–113 (2020)
4. Cools, R., Kuo, F.Y., Nuyens, D., Sloan, I.H.: Fast component-by-component construction of lattice algorithms for multivariate approximation with POD and SPOD weights. Math. Comp. **90**, 787–812 (2021)
5. de Boor, C., Lynch, R.E.: On splines and their minimum properties. J. Math. Mech. **15**, 953–969 (1966)
6. Gantner, R. N.: Dimension truncation in QMC for affine-parametric operator equations. In: Monte Carlo and Quasi-Monte Carlo Methods 2016, pp. 249–264, Stanford, CA, August 14–19 (2018)
7. Gantner, R.N., Herrmann, L., Schwab, C.: Quasi-Monte Carlo integration for affine-parametric, elliptic PDEs: local supports and product weights. SIAM J. Numer. Anal. **56**(1), 111–135 (2018)
8. Golomb R. N., Weinberger, H. F.: Optimal approximation and error bounds. In: Langer, R. E. (ed.) Numerical Approximation Proceedings a Symposium, Madison, April 21–23, 1958. Publication No. 1 of the Mathematics Research Center, US Army, the University of Wisconsin, pp. 117–190. The University of Wisconsin Press, Madison, Wis (1959)
9. Graham, I.G., Kuo, F.Y., Nichols, J.A., Scheichl, R., Schwab, C., Sloan, I.H.: Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. Numer. Math. **131**(2), 329–368 (2015)
10. Griebel, M., Rieger, C.: Reproducing kernel Hilbert spaces for parametric partial differential equations. SIAM/ASA J. Uncertain. Quantif. **5**(1), 111–137 (2017)
11. Herrmann, L., Schwab, C.: QMC integration for lognormal-parametric, elliptic PDEs: local supports and product weights. Numer. Math. **141**(1), 63–102 (2019)
12. Kaarnioja, V., Kuo, F.Y., Sloan, I.H.: Uncertainty quantification using periodic random variables. SIAM J. Numer. Anal. **58**(2), 1068–1091 (2020)
13. Kämmerer, L., Potts, D., Volkmer, T.: Approximation of multivariate periodic functions by trigonometric polynomials based on rank-1 lattice sampling. J. Comp. **31**, 543–576 (2015)
14. Kazashi, Y.: Quasi-Monte Carlo integration with product weights for elliptic PDEs with log-normal coefficients. IMA J. Numer. Anal. **39**(3), 1563–1593 (2019)
15. Kempf, R., Wendland, H., Rieger, C.: Kernel-based reconstructions for parametric PDEs. In: Griebel, M., Schweitzer, M. (eds.) Meshfree Methods Partial Differ. Equations IX. IWMMPDE 2017, pp. 53–71. Springer (2019)
16. Kuo, F.Y., Migliorati, G., Nobile, F., Nuyens, D.: Function integration, reconstruction and approximation using rank-1 lattices. Math. Comp. **90**, 1861–1897 (2021)

17. Kuo, F.Y., Schwab, C., Sloan, I.H.: Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients. SIAM J. Numer. Anal. **50**(6), 3351–3374 (2012)
18. Kuo, F. Y., Sloan, I. H., Woźniakowski, H.: Lattice rules for multivariate approximation in the worst case setting. In: Niederreiter, H., Talay, D. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004, pp. 289–330, Springer (2006)
19. Kuo, F.Y., Sloan, I.H., Woźniakowski, H.: Lattice rule algorithms for multivariate approximation in the average case setting. J. Comp. **24**, 283–323 (2008)
20. Micchelli, C. A., Rivlin, T. J.: A survey of optimal recovery. Optim. Estim. Approx. theory (Proc. Internat. Sympos., Freudenstadt, 1976), pp. 1–54 (1977)
21. Micchelli, C. A., Rivlin, T. J.: Lectures on optimal recovery. In: Turner P.R. (eds) Numerical Analysis Lancaster 1984. Lecture Notes in Mathematics, vol 1129. Springer (1985)
22. Rauhut, H., Schwab, C.: Compressive sensing Petrov–Galerkin approximation of high-dimensional parametric operator equations. Math. Comp. **86**(304), 661–700 (2016)
23. Rogers, L. C. G., Williams, D.: Diffusions, Markov Processes, and Martingales. Vol. 1., 2nd edition, Cambridge University Press (2017)
24. Sloan, I.H., Joe, S.: Lattice Methods for Multiple Integration. Oxford University Press, Oxford (1994)
25. Sloan, I.H., Woźniakowski, H.: Tractability of multivariate integration for weighted Korobov classes. J. Comp. **17**, 697–721 (2001)
26. Wendland, H.: Scattered Data Approximation. Cambridge University Press, Cambridge (2005)
27. Xiu, D., Karniadakis, G.E.: The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. **24**, 619–644 (2002)
28. Zeng, X. Y., Leung, K. T., Hickernell, F. J.: Error analysis of splines for periodic problems using lattice designs. In: Niederreiter, H., Talay, D. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004, pp. 501–514, Springer (2006)
29. Zeng, X.Y., Kritzer, P., Hickernell, F.J.: Spline methods using integration lattices and digital nets. Constr. Approx. **30**, 529–555 (2009)