

# Fully discrete approximation of parametric and stochastic elliptic PDEs

M. Bachmayr and A. Cohen and D. Düng and Ch. Schwab

Research Report No. 2017-08  
February 2017

Seminar für Angewandte Mathematik  
Eidgenössische Technische Hochschule  
CH-8092 Zürich  
Switzerland

# Fully discrete approximation of parametric and stochastic elliptic PDEs \*

Markus Bachmayr, Albert Cohen, Dinh Dũng and Christoph Schwab

February 13, 2017

## Abstract

It has recently been demonstrated that locality of spatial supports in the parametrization of coefficients in elliptic PDEs can lead to improved convergence rates of sparse polynomial expansions of the corresponding parameter-dependent solutions. These results by themselves do not yield practically realizable approximations, since they do not cover the approximation of the arising expansion coefficients, which are functions of the spatial variable. In this work, we study the combined spatial and parametric approximability for elliptic PDEs with affine or lognormal parametrizations of the diffusion coefficients and corresponding Taylor, Jacobi, and Hermite expansions, to obtain fully discrete approximations. Our analysis yields convergence rates of the fully discrete approximation in terms of the total number of degrees of freedom. The main vehicle consists in  $\ell^p$  summability results for the coefficient sequences measured in higher-order Hilbertian Sobolev norms. We also discuss similar results for non-Hilbertian Sobolev norms which arise naturally when using adaptive spatial discretizations.

## 1 Introduction

Parametric PDEs are of interest for modeling many complex phenomena, where the involved parameters may be of deterministic or stochastic nature. Their numerical treatment was initiated in the 1990s, see [18, 19, 22, 23] for general references. It has recently drawn much attention in the case where the number of involved parameters is very large [6, 7], or *countably infinite* [12, 13, 9].

In this paper, we consider the elliptic diffusion equation

$$-\operatorname{div}(a\nabla u) = f, \quad (1.1)$$

set on a given bounded Lipschitz domain  $D \subset \mathbb{R}^m$  (say with  $m = 1, 2$  or  $3$ ), for some fixed right-hand side  $f$ , homogeneous Dirichlet boundary conditions  $u|_{\partial D} = 0$ , and spatially variable scalar diffusion coefficient  $a$ . Using the notation  $V = H_0^1(D)$  and  $V' = H^{-1}(D)$ , for any  $f \in V'$ , the weak formulation of (1.1) in  $V$ ,

$$\int_D a \nabla u \cdot \nabla v = \langle f, v \rangle_{V', V}, \quad v \in H_0^1(D), \quad (1.2)$$

has a unique solution  $u \in V$  whenever the diffusion coefficient  $a$  satisfies  $0 < r < a < R < \infty$ .

---

\*Markus Bachmayr acknowledges support by the Hausdorff Center of Mathematics, University of Bonn. Albert Cohen is supported by the Institut Universitaire de France and by the European Research Council under grant ERC AdG 338977 BREAD. Dinh Dũng's research work is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2017.05. Christoph Schwab is supported in part by the Swiss National Science Foundation.

We consider diffusion coefficients having a parametrized form  $a = a(y)$ , where  $y = (y_j)_{j \geq 1}$  is a sequence of real-valued parameters ranging in some set  $U \subset \mathbb{R}^{\mathbb{N}}$ . The resulting solution map

$$u \mapsto u(y), \quad (1.3)$$

acts from  $U$  to the solution space  $V$ . The objective is to achieve numerical approximation of this complex map by a small number of parameters with some guaranteed error in a given norm. Two relevant types of parametrizations have been the object of intensive study.

The first one is the so-called affine form

$$a = a(y) = \bar{a} + \sum_{j \geq 1} y_j \psi_j, \quad (1.4)$$

where  $\bar{a}$  and  $(\psi_j)_{j \geq 1}$  are given functions in  $L^\infty(D)$ , and the parameters  $y_j$  range in  $[-1, 1]$ . The parameter domain is thus

$$U = [-1, 1]^{\mathbb{N}} \quad (1.5)$$

and the solution map is well-defined under the so-called *uniform ellipticity assumption*

$$\sum_{j \geq 1} |\psi_j| \leq \bar{a} - r, \quad \text{a.e. on } D, \quad (1.6)$$

for some  $r > 0$ . An equivalent condition is that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$  and

$$\left\| \frac{\sum_{j \geq 1} |\psi_j|}{\bar{a}} \right\|_{L^\infty} = \theta < 1. \quad (1.7)$$

In this case, the approximation error is often measured in  $L^\infty(U, V)$  or  $L^2(U, V, \sigma)$  where  $d\sigma$  is some product probability measure on  $U$ .

The second one is the so-called lognormal form

$$a = \exp(b), \quad b = b(y) = \sum_{j \geq 1} y_j \psi_j, \quad (1.8)$$

where the  $y_j$  are i.i.d. standard Gaussian random variables. In this case, we have  $U = \mathbb{R}^{\mathbb{N}}$  and the error is measured in  $L^2(U, V, \gamma)$  where

$$d\gamma(y) := \bigotimes_{j \geq 1} g(y_j) dy_j, \quad g(t) := \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. \quad (1.9)$$

Note that the above countably infinite-dimensional setting comprises its finite-dimensional counterpart by setting  $\psi_j = 0$  for  $j$  large enough. In both affine and lognormal cases, a powerful strategy for the approximation of the solution map is based on the truncation of polynomial expansions of the general form

$$u(y) = \sum_{\nu \in \mathcal{F}} u_\nu \phi_\nu(y). \quad (1.10)$$

Here  $\mathcal{F}$  is the set of finitely supported sequences of non-negative integers, and

$$\phi_\nu(y) = \prod_{j \geq 1} \varphi_{\nu_j}(y_j), \quad \nu = (\nu_j)_{j \geq 1}, \quad (1.11)$$

where  $(\varphi_n)_{n \geq 0}$  is a family of univariate polynomials such that  $\varphi_0 = 1$  and  $\deg(\varphi_n) = n$ . The approximation strategy consists in defining

$$u_n(y) = \sum_{\nu \in \Lambda_n} u_\nu \phi_\nu(y), \quad (1.12)$$

where  $\Lambda_n$  is a selection of  $n$  indices from  $\mathcal{F}$ . If the  $\phi_\nu$  are normalized so that  $|\phi_\nu(y)| \leq 1$  for all  $y \in U$ , then the truncation error can be controlled in  $L^\infty(U, V)$  by

$$\|u - u_n\|_{L^\infty(U, V)} \leq \sum_{\nu \notin \Lambda_n} \|u_\nu\|_V. \quad (1.13)$$

If the  $\phi_\nu$  constitute an orthonormal basis of  $L^2(U)$  for some given measure, then the truncation error in  $L^2(U, V)$  for the same measure is given by

$$\|u - u_n\|_{L^2(U, V)} = \left( \sum_{\nu \notin \Lambda_n} \|u_\nu\|_V^2 \right)^{1/2}. \quad (1.14)$$

These estimates justify the use of best  $n$ -term truncations, that is, taking  $\Lambda_n$  to be the index set corresponding to the  $n$  largest  $\|u_\nu\|_V$ . With such a choice, classical results on best  $n$ -term approximation in sequence spaces [14] show that if  $0 < p < q$ ,

$$(\|u_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F}) \implies \left( \sum_{\nu \notin \Lambda_n} \|u_\nu\|_V^q \right)^{1/q} \leq C n^{-s}, \quad s := \frac{1}{p} - \frac{1}{q}, \quad (1.15)$$

where  $C := \|(\|u_\nu\|_V)_{\nu \in \mathcal{F}}\|_{\ell^p}$ . In particular, the tails in the right-hand sides of (1.13) and (1.14) decrease like  $n^{-s}$  where  $s = \frac{1}{p} - 1$  and  $s = \frac{1}{p} - \frac{1}{2}$  respectively, provided that  $(\|u_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})$ , with  $p < 1$  and  $p < 2$ , respectively. A central objective is therefore to establish  $\ell^p$  summability results for the coefficient sequence in the given expansion, with  $p$  as small as possible.

In the affine case, we consider two types of expansions. The first type are the Taylor (or power) series of the form

$$\sum_{\nu \in \mathcal{F}} t_\nu y^\nu, \quad t_\nu := \frac{1}{\nu!} \partial^\nu u(y=0), \quad \nu! := \prod_{j \geq 1} \nu_j!, \quad (1.16)$$

with the convention that  $0! = 1$ . The second type are the orthogonal Jacobi series of the form

$$\sum_{\nu \in \mathcal{F}} v_\nu J_\nu(y), \quad J_\nu(y) = \prod_{j \geq 1} J_{\nu_j}(y_j), \quad v_\nu := \int_U u(y) J_\nu(y) d\sigma(y), \quad (1.17)$$

where  $(J_k)_{k \geq 0}$  is the sequence of Jacobi polynomials on  $[-1, 1]$  normalized with respect to the Jacobi probability measure

$$\int_{-1}^1 |J_k(t)|^2 d_{\alpha, \beta}(t) dt = 1, \quad (1.18)$$

where  $d_{\alpha, \beta}(t) = c_{\alpha, \beta} (1-t)^\alpha (1+t)^\beta$  and  $c_{\alpha, \beta} = (\int_{-1}^1 (1-t)^\alpha (1+t)^\beta dt)^{-1}$ . Therefore  $(J_\nu)_{\nu \in \mathcal{F}}$  is an orthonormal basis of  $L^2(U, \sigma)$ , where

$$d\sigma(y) := \bigotimes_{j \geq 1} d_{\alpha, \beta}(y_j) dy_j \quad (1.19)$$

One particular example corresponding to the values  $\alpha = \beta = 0$  and  $d_{0,0}(t) = \frac{1}{2}$  is the family of tensorized Legendre polynomials.

In the lognormal case, we consider Hermite series of the form

$$\sum_{\nu \in \mathcal{F}} w_\nu H_\nu(y), \quad H_\nu(y) = \prod_{j \geq 1} H_{\nu_j}(y_j), \quad w_\nu := \int_U u(y) H_\nu(y) d\gamma(y), \quad (1.20)$$

with  $(H_k)_{k \geq 0}$  being the sequence of Hermite polynomials normalized according to

$$\int_{\mathbb{R}} |H_k(t)|^2 g(t) dt = 1, \quad (1.21)$$

and  $d\gamma$  given by (1.9). In this case  $U = \mathbb{R}^{\mathbb{N}}$  and  $(H_\nu)_{\nu \in \mathcal{F}}$  is an orthonormal basis of  $L^2(U, \gamma)$ .

In the affine case, the first  $\ell^p$  summability results have been obtained in [13] for the Taylor and Legendre coefficient sequences, under the conditions that (1.6) holds and  $(\|\psi_j\|_{L^\infty})_{j \geq 1} \in \ell^p(\mathbb{N})$ , for some  $0 < p < 1$ . In the lognormal case, similar results have been first obtained in [21] for the Hermite coefficients under the conditions that  $(j\|\psi_j\|_{L^\infty})_{j \geq 1} \in \ell^p(\mathbb{N})$ , for some  $0 < p \leq 1$ . Such results yield algebraic convergence rates  $n^{-s}$  that are free from the curse of dimensionality, in the sense that they hold for countably many variables  $y_j$ . One of their intrinsic limitations is that the conditions on the functions  $\psi_j$  are expressed only through their  $L^\infty$  norms. In particular, they do not take into account their support properties, and the possible gain in summability when these supports have limited overlap.

A different approach to summability results that takes into account the support properties has recently been proposed in [3] for the affine case and [4] for the lognormal case. This approach gives significant improvements on the provable range of  $\ell^p$  summability when the  $\psi_j$  have disjoint supports, or limited overlap, such as splines, finite elements or wavelet bases. The main results are the following for the affine and lognormal case, respectively.

**Theorem 1.1.** *Let  $0 < q < \infty$  and  $0 < p < 2$  be such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ . Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of numbers strictly larger than 1 such that  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  and such that*

$$\theta := \left\| \frac{\sum_{j \geq 1} \rho_j |\psi_j|}{\bar{a}} \right\|_{L^\infty} < 1. \quad (1.22)$$

*Then  $(\|t_\nu\|_V)_{\nu \in \mathcal{F}}$  and  $(\|v_\nu\|_V)_{\nu \in \mathcal{F}}$  belong to  $\ell^p(\mathcal{F})$ .*

**Theorem 1.2.** *Let  $0 < q < \infty$  and  $0 < p < 2$  be such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ . Assume that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers such that  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  and such that*

$$\left\| \sum_{j \geq 1} \rho_j |\psi_j| \right\|_{L^\infty} < \infty. \quad (1.23)$$

*Then the sequence  $(\|w_\nu\|_V)_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ .*

While the above results allow us to establish algebraic convergence rates of the properly truncated series (1.12) in terms of the number  $n$  of retained coefficients, they do not yet yield practically realizable approximations. Indeed, the coefficients  $u_\nu \in \{t_\nu, v_\nu, w_\nu\}$  belong to the infinite dimensional space  $V$  and should themselves be approximated by means of spatial discretizations. Such discretizations are typically performed by means of finite elements or wavelets, and the number  $n_\nu$  of allocated degrees of freedom may vary with the retained index  $\nu$ .

In the present paper, we address this issue and establish several results that describe the rate of convergence in terms of the total number of degrees of freedom  $N = \sum_{\nu \in \Lambda_n} n_\nu$ . Our main vehicle is to investigate the  $\ell^p$  summability of the sequence  $(\|u_\nu\|)_{\nu \in \mathcal{F}}$  for the various coefficients  $u_\nu \in \{t_\nu, v_\nu, w_\nu\}$ , where  $\|\cdot\|$  is a norm associated to a higher-order smoothness class in  $V$ , for example the Sobolev space  $H^k(D)$  for some  $k > 1$ . These new summability results are stated in §2, and their implications on the convergence rate of spatial-parametric discretization

methods are discussed in §3 using finite element spaces for the spatial discretization. The proof of the convergence results is given in §4-5-6 for the Taylor, Jacobi and Hermite coefficients, respectively. The smoothness classes considered in these results are Hilbertian Sobolev spaces. In order to analyze the potential benefit of a fully adaptive space-parameter discretization, it is interesting to consider also non-Hilbertian Sobolev or Besov classes which are known to govern the rate of nonlinear approximation methods such as adaptive finite elements or wavelets. Some first results in this direction are given in §7.

Finally, we illustrate in §8 the various results in this paper for a specific affine parametrization of the diffusion coefficient by a wavelet decomposition. For this specific example, one main finding is that the convergence rate in terms of  $N$  for the space-parameter approximation is typically higher when using nonlinear approximation both in the spatial and parametric variable.

The approximation results obtained in this paper may be viewed as a benchmark for the performance of concrete numerical strategies, such as non-adaptive or adaptive space-parameter Galerkin methods.

## 2 Summability results

The results established in the further sections §4-5-6-7 of this paper give sufficient conditions for  $\ell^p$  summability of the sequences  $(\|u_\nu\|_X)_{\nu \in \mathcal{F}}$  where  $u_\nu \in \{t_\nu, v_\nu, w_\nu\}$  and  $\|\cdot\|_X$  is a norm of a relevant smoothness class  $X$  of  $V = H_0^1(D)$ .

One first such class that appears in a natural manner is the space

$$W := \{v \in V : \Delta v \in L^2(D)\}. \quad (2.1)$$

This space is equipped with the norm

$$\|v\|_W := \|\Delta v\|_{L^2}. \quad (2.2)$$

Note that the above is a norm since  $\Delta v = 0$  implies  $v = 0$  when  $v \in V$  due to the boundary condition. By elliptic regularity theory, it is also known that  $W$  coincides with the Sobolev space  $V \cap H^2(D)$  with equivalent norms, in the case where the domain  $D$  is convex or of  $C^{1,1}$  smoothness, see [17, Theorems 3.2.1.3 and 2.4.2.5]. Our first result, established in §4.1, concerns the Taylor coefficients in the case of the affine parametrization (1.4). Here, and in the rest of the paper, we use the notation

$$|\nabla\psi| = \left( \sum_{i=1}^m |D_{x_i}\psi|^2 \right)^{1/2}, \quad (2.3)$$

for the Euclidean norm of the gradient.

**Theorem 2.1.** *Let  $0 < q < \infty$  and  $0 < p < 2$  be such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ . Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of numbers strictly larger than 1 such that  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  and such that*

$$\theta := \left\| \frac{\sum_{j \geq 1} \rho_j |\psi_j|}{\bar{a}} \right\|_{L^\infty} < 1. \quad (2.4)$$

*Assume in addition that the right side  $f$  in (1.1) belongs to  $L^2(D)$  and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{1,\infty}(D)$  and that*

$$\left\| \sum_{j \geq 1} \rho_j |\nabla\psi_j| \right\|_{L^\infty} < \infty. \quad (2.5)$$

Then the sequence  $(\|t_\nu\|_W)_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ . In particular, when  $D$  is convex or of  $C^{1,1}$  smoothness, the same holds for  $(\|t_\nu\|_{H^2})_{\nu \in \mathcal{F}}$ .

**Remark 2.2.** The first condition (2.4) in the above theorem is the same as (1.22) in Theorem 1.1. However, the additional condition (2.5) puts further constraints on the choice of the sequence  $\rho$ . For this reason, we expect that the two values  $p_V := \inf\{p > 0 : (\|t_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})\}$  and  $p_W := \inf\{p > 0 : (\|t_\nu\|_W)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})\}$  may in general differ in the sense that  $p_V < p_W$ . This remark also applies to the next results dealing with higher-order smoothness and other types of polynomial expansions.

Higher-order smoothness can be treated under more stringent conditions on the smoothness of the functions  $f$ ,  $\bar{a}$ ,  $\psi_j$  and of the domain  $D$ . For any integer  $k \geq 2$  we introduce the space

$$W^k := \{v \in V : \Delta v \in H^{k-2}(D)\}. \quad (2.6)$$

In particular  $W^2 = W$  defined above. This space is equipped with the norm

$$\|v\|_{W^k} := \|\Delta v\|_{H^{k-2}}, \quad (2.7)$$

and coincides with the Sobolev space  $V \cap H^k(D)$  with equivalent norms if the domain  $D$  has  $C^{k-1,1}$  smoothness, see [17, Theorem 2.5.1.1]. The following result, established in §4.2, generalizes Theorem 2.1 to such spaces.

**Theorem 2.3.** *Let  $0 < q < \infty$  and  $0 < p < 2$  be such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ , and let  $k > 2$  be an integer. Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of numbers strictly larger than 1 such that  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  and such that (2.4) holds. Assume in addition that the right side  $f$  in (1.1) belongs to  $H^{k-2}(D)$ , that the domain  $D$  has  $C^{k-2,1}$  smoothness, that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{k-1,\infty}(D)$  and that*

$$\sup_{|\alpha| \leq k-1} \left\| \sum_{j \geq 1} \rho_j |D^\alpha \psi_j| \right\|_{L^\infty} < \infty. \quad (2.8)$$

Then the sequence  $(\|t_\nu\|_{W^k})_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ . In particular, when  $D$  has  $C^{k-1,1}$  smoothness, the same holds for  $(\|t_\nu\|_{H^k})_{\nu \in \mathcal{F}}$ .

We also present in §4.3 some variants of these results using fractional Sobolev spaces and weighted Sobolev spaces for polygonal domains. In the case of Jacobi expansions, the following analogous results are established in §5 under the exact same assumptions.

**Theorem 2.4.** *The sequence  $(\|v_\nu\|_{W^k})_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ , under the same assumptions as those of Theorem 2.1 when  $k = 2$  and of Theorem 2.3 when  $k > 2$ .*

Finally, in the case of the lognormal parametrization (1.8), we establish in §6 the following analog to Theorem 2.1.

**Theorem 2.5.** *Let  $0 < q < \infty$  and  $0 < p < 2$  be such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ . Assume that the right side  $f$  in (1.1) belongs to  $L^2(D)$  and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{1,\infty}(D)$ . Assume in addition that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers such that  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  and such that*

$$\left\| \sum_{j \geq 1} \rho_j |\psi_j| \right\|_{L^\infty} + \left\| \sum_{j \geq 1} \rho_j |\nabla \psi_j| \right\|_{L^\infty} < \infty. \quad (2.9)$$

Then  $u \in L^k(U, W, \gamma)$  for all  $k > 0$  and the sequence  $(\|w_\nu\|_W)_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ . In particular, when  $D$  is convex or of  $C^{1,1}$  smoothness, the same holds for  $(\|w_\nu\|_{H^2})_{\nu \in \mathcal{F}}$ .

The proof of this result is given in §6. For higher-order smoothness, an analogous result seems to be valid, namely: under the assumption that the domain  $D$  has  $C^{k-2,1}$  smoothness, that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{k-1,\infty}(D)$  and that (2.8) holds, then the sequence  $(\|w_\nu\|_{W^k})_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ . However, the proof of this result seems to require heavy technical and notational developments, and we therefore do not attempt to include it in this paper.

### 3 Space-parameter approximation

We recall that polynomial approximations are typically obtained by truncation of the general series (1.10) according to

$$u_n := \sum_{\nu \in \Lambda_n} u_\nu \phi_\nu. \quad (3.1)$$

When taking for  $\Lambda_n$  a set corresponding to the  $n$  largest  $\|u_\nu\|_V$ , the rate of convergence of such approximations is governed by the  $\ell^p$  summability of the  $(\|u_\nu\|_V)_{\nu \in \mathcal{F}}$ . However, the approximation  $u_n$  is still picked from an infinite-dimensional space, namely

$$V_{\Lambda_n} := V \otimes \text{span}\{\phi_\nu : \nu \in \Lambda_n\}. \quad (3.2)$$

The summability results for the  $W^k$  or  $H^k$  norms of the coefficients  $u_\nu$  stated in the previous section allow us to introduce further spatial discretization resulting in approximations picked from finite-dimensional spaces, and to analyze the rate of convergence in terms of the dimension of such spaces. As already noted in Remark 2.2, since the conditions in these results become more stringent as  $k$  increases, we generally obtain that

$$(\|u_\nu\|_{H^k})_{\nu \in \mathcal{F}} \in \ell^{p_k}(\mathcal{F}), \quad (3.3)$$

for some sequence

$$p_1 \leq p_2 \leq \dots \leq p_k. \quad (3.4)$$

Generally speaking, we assume that

$$(\|u_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^{p_V}(\mathcal{F}), \quad (3.5)$$

and that for some given regularity class  $X \subset V$  of interest, we have that

$$(\|u_\nu\|_X)_{\nu \in \mathcal{F}} \in \ell^{p_X}(\mathcal{F}), \quad (3.6)$$

for some  $p_V \leq p_X$ .

The spatial discretization is based on a sequence  $(V_n)_{n>0}$  of subspaces of  $V$  with dimension

$$\dim(V_n) = n. \quad (3.7)$$

One instance is obtained by considering Lagrange finite elements on a family of quasi-uniform regular simplicial partitions  $(\mathcal{T}_h)_{h>0}$  of mesh size  $h$ , up to isoparametric transformations in the case of curved domains. Since  $D$  is a bounded domain of  $\mathbb{R}^m$ , the dimension  $n$  of  $V_n$  is related to the corresponding mesh size

$$n \sim h^{-m}, \quad (3.8)$$

We assume that such spaces have the following *approximation property*: there is a constant  $C_X > 0$  and  $t > 0$ ,

$$\min_{v_n \in V_n} \|v - v_n\|_V \leq C_X n^{-t} \|v\|_X, \quad n > 0, \quad v \in X, \quad (3.9)$$



where  $X \subset V$  is the regularity class for which (3.6) holds.

For example, with  $X = H^k(D)$ , classical error estimates [10] yield the convergence rate  $t = \frac{k-1}{m}$  by using Lagrange finite elements of order at least  $k-1$  on quasi-uniform partitions. Note that the spaces  $W^k$  introduced in the previous section do not always coincide with  $H^k(D)$ . For example, for  $k=2$  and dimension  $m=2$ , we know that  $W$  is strictly larger than  $H^2(D)$  when  $D$  is a polygon with re-entrant corner. In this case, it is well known that the optimal rate  $t=1/2$  is yet attained with  $X=W$ , when using spaces  $V_n$  associated to meshes  $(\mathcal{T}_n)_{n>0}$  with proper refinement near the re-entrant corners where the functions  $v \in W$  might have singularities.

The relevant finite element discretization are therefore typically generated by uniform or non-uniform refinement from a coarse triangulation of  $D$ . Note that not all values of  $n > 0$  may be met in this process. However, up to completing the sequence by  $V_n := V_{\tilde{n}}$ , where  $\tilde{n}$  is the largest attained value below  $n$ , we reach the same approximation estimate (3.9) with spaces  $V_n$  of dimension at most  $n$ .

We apply spatial discretization in (3.1) by replacing each  $u_\nu \in V$  by some  $u_{\nu, n_\nu} \in V_{n_\nu}$ , where we allow the number of degrees of freedom  $n_\nu$  to depend on  $\nu$ . This variability is critical for the resulting convergence rate in terms of the total number of degrees of freedom. Thus, with

$$\mathbf{n} := (n_\nu)_{\nu \in \Lambda_n}, \quad (3.10)$$

the resulting approximant

$$u_{\mathbf{n}} := \sum_{\nu \in \Lambda_n} u_{\nu, n_\nu} \phi_\nu, \quad (3.11)$$

belongs to the space

$$V_{\mathbf{n}} := \bigoplus_{\nu \in \Lambda_n} (V_{n_\nu} \otimes \mathbb{R}\phi_\nu). \quad (3.12)$$

This space is characterized by the total number of degrees of freedom for each  $V_{n_\nu}$ :

$$N = \dim(V_{\mathbf{n}}) = \sum_{\nu \in \Lambda_n} n_\nu. \quad (3.13)$$

We again take for  $\Lambda_n$  a set corresponding to  $n$  largest  $\|u_\nu\|_V$ .

If the coefficients  $u_\nu$  belong to the smoothness space  $X$ , there exist  $u_{\nu, n_\nu} \in V_{n_\nu}$  such that

$$\|u_\nu - u_{\nu, n_\nu}\|_V \leq C_X n_\nu^{-t} \|u_\nu\|_X. \quad (3.14)$$

We may in particular take for  $u_{\nu, n_\nu}$  the  $V$ -orthogonal projection of  $u_\nu$  onto  $V_{n_\nu}$ .

We may then write for a given norm  $\mathcal{V}$ ,

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq \|u - u_n\|_{\mathcal{V}} + \left\| \sum_{\nu \in \Lambda_n} (u_\nu - u_{\nu, n_\nu}) \phi_\nu \right\|_{\mathcal{V}}. \quad (3.15)$$

In the case of Taylor series (or more general polynomial series where the  $\phi_\nu$  are uniformly bounded by 1 over  $U$ ), taking  $\mathcal{V} = L^\infty(U, V)$ , we control the first term by

$$\|u - u_n\|_{\mathcal{V}} \leq \sum_{\nu \notin \Lambda_n} \|u_\nu\|_V \leq C n^{-s}, \quad s := \frac{1}{p_V} - 1, \quad C := \|(\|u_\nu\|_V)_{\nu \in \mathcal{F}}\|_{\ell^{p_V}}, \quad (3.16)$$

and the second term by

$$\left\| \sum_{\nu \in \Lambda_n} (u_\nu - u_{\nu, n_\nu}) \phi_\nu \right\|_{\mathcal{V}} \leq C_X \sum_{\nu \in \Lambda_n} n_\nu^{-t} \|u_\nu\|_X. \quad (3.17)$$

We now allocate the degrees of freedom, that is, the values of  $n_\nu$ , in order to minimize  $N$  for a given total error. Leaving aside the multiplicative constants, this amounts to solving the constrained minimization problem

$$\min \left\{ \sum_{\nu \in \Lambda_n} n_\nu : \sum_{\nu \in \Lambda_n} n_\nu^{-t} \|u_\nu\|_X \leq n^{-s} \right\}. \quad (3.18)$$

Up to introducing multiplicative constants, we treat the  $n_\nu$  as real numbers, and using a Lagrange multiplier we obtain

$$n_\nu = \eta \|u_\nu\|_X^{\frac{1}{1+t}}, \quad (3.19)$$

for some  $\eta > 0$  independent of  $\nu \in \Lambda_n$ . Its value is determined by the saturated constraint

$$n^{-s} = \sum_{\nu \in \Lambda_n} n_\nu^{-t} \|u_\nu\|_X = \eta^{-t} \sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}}, \quad (3.20)$$

and therefore

$$\eta = n^{\frac{s}{t}} \left( \sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}} \right)^{\frac{1}{t}}. \quad (3.21)$$

Combining this with (3.19) and summing over  $\nu \in \Lambda_n$ , we find

$$N = n^{\frac{s}{t}} \left( \sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}} \right)^{\frac{1+t}{t}}. \quad (3.22)$$

We now distinguish between two cases.

1.  $p_X \leq \frac{1}{t+1}$  in which case  $\sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}}$  is bounded independently of the set  $\Lambda_n$ . Since the global error is controlled by  $n^{-s}$ , we obtain a bound of the form

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq CN^{-t}. \quad (3.23)$$

Note that this is the convergence rate for the spatial discretization of a single  $v \in X$  as given by (3.9).

2.  $p_X > \frac{1}{t+1}$ : in this case  $\sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}}$  may not be uniformly bounded and we estimate it using Hölder's inequality that gives

$$\sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{1}{1+t}} \leq Cn^\delta, \quad \delta := 1 - \frac{1}{p_X(1+t)} > 0.$$

According to (3.22), we thus have

$$N \leq Cn^{\frac{s+(1+t)\delta}{t}}. \quad (3.24)$$

Combining this with the fact that the global error is controlled by  $n^{-s}$ , we obtain a bound of the form

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq CN^{-r}, \quad r := \frac{st}{s + (1+t)\delta}. \quad (3.25)$$

Note that

$$r = \frac{st}{t + s + 1 - p_X^{-1}} \leq \frac{st}{t + s + 1 - p_V^{-1}} = s. \quad (3.26)$$

On the other hand, since the second expression increases with  $s$  as long as  $s \geq p_X^{-1} - 1$ , and since the actual value of  $s$  is  $p_V^{-1} - 1 \geq p_X^{-1} - 1$ , we find that  $r \geq \frac{1}{p_X} - 1$ . In summary, we find that

$$\frac{1}{p_X} - 1 \leq r \leq \frac{1}{p_V} - 1. \quad (3.27)$$

In the case of Jacobi or Hermite series (or more general orthonormal polynomial expansions), taking  $\mathcal{V} = L^2(U, V)$  with the appropriate probability measure, we control the first term by

$$\|u - u_n\|_{\mathcal{V}} = \left( \sum_{\nu \notin \Lambda_n} \|u_\nu\|_V^2 \right)^{1/2} \leq Cn^{-s}, \quad s := \frac{1}{p_V} - \frac{1}{2}, \quad C := \|(\|u_\nu\|_V)_{\nu \in \mathcal{F}}\|_{\ell^{p_V}}, \quad (3.28)$$

and the second term by

$$\left\| \sum_{\nu \in \Lambda_n} (u_\nu - u_{\nu, n_\nu}) \phi_\nu \right\|_{\mathcal{V}} = \left( \sum_{\nu \in \Lambda_n} \|u_\nu - u_{\nu, n_\nu}\|_V^2 \right)^{1/2} \leq C_X \left( \sum_{\nu \in \Lambda_n} n_\nu^{-2t} \|u_\nu\|_X^2 \right)^{1/2}. \quad (3.29)$$

Proceeding in a similar way as in the previous case for the optimal allocation of the degrees of freedom, we now obtain

$$N = n^{\frac{s}{t}} \left( \sum_{\nu \in \Lambda_n} \|u_\nu\|_X^{\frac{2}{1+2t}} \right)^{\frac{1+2t}{2t}}. \quad (3.30)$$

This leads to the following two cases which are discussed in the same way as before:

1.  $p_X \leq \frac{2}{2t+1}$  in which case we obtain the global error bound (3.23).
2.  $p_X > \frac{2}{2t+1}$  in which case we obtain the global error bound (3.25), with rate now given by

$$r := \frac{st}{s + t + \frac{1}{2} - p_X^{-1}}, \quad (3.31)$$

which satisfies

$$\frac{1}{p_X} - \frac{1}{2} \leq r \leq \frac{1}{p_V} - \frac{1}{2}. \quad (3.32)$$

We summarize the above discussion with the following result.

**Theorem 3.1.** *Assume that the discretization spaces  $(V_n)_{n>0}$  satisfy the approximation property (3.9) for a certain smoothness class  $X \subset V$  and that  $(\|u_\nu\|_X)_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$  for some  $0 < p < 2$ . Then, for each  $n$  there exists a vector  $\mathbf{n} := (n_\nu)_{\nu \in \Lambda_n}$  such that*

$$\min_{\mathbf{u}_n \in V_{\mathbf{n}}} \|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq CN^{-\min\{r, t\}}, \quad (3.33)$$

where  $N = \sum_{\nu \in \Lambda_n} n_\nu = \dim(V_{\mathbf{n}})$ . Here  $\mathcal{V}$  is  $L^\infty(U, V)$  in the case of Taylor series,  $L^2(U, V, \sigma)$  in the case of Jacobi series and  $L^2(U, V, \gamma)$  in the case of Hermite series. The rate  $t$  corresponds to the spatial approximation of a single  $v \in X$  as given by (3.9). The rate  $r$  is given by (3.25) in the first case and by (3.31) in the second case. The constant  $C$  in (3.33) depends on the quantities  $\|(\|u_\nu\|_V)_{\nu \in \mathcal{F}}\|_{\ell^{p_V}}$  and  $\|(\|u_\nu\|_X)_{\nu \in \mathcal{F}}\|_{\ell^{p_X}}$ , as well as on the constant  $C_X$  in (3.9).

**Remark 3.2.** The idea of optimizing the allocation of spatial degrees of freedom for the different coefficients  $u_\nu$  in the expansion (1.10) has also been developed in [15] using a slightly different approach. Here, one uses a nested hierarchy  $(V_{2^k})_{k \geq 0}$  of space discretization and  $P_{V_{2^k}}$  denotes a  $V$ -stable projector on the space  $V_{2^k}$ , so that the approximation property (3.9) now takes the form

$$\|v - P_{V_{2^k}} v\|_V \leq C_X 2^{-kt} \|v\|_X, \quad k \geq 0. \quad (3.34)$$

The approximations of the solution map are of the form

$$u_G = \sum_{(k, \nu) \in G} \delta_k(u_\nu) \phi_\nu(y) \quad (3.35)$$

where  $G$  is a finite subset of  $\mathbb{N}_0 \times \mathcal{F}$ , and where the  $\delta_k(u_\nu)$  are multiscale components defined by

$$\delta_k(v) := P_{V_{2^k}} v - P_{V_{2^{k-1}}} v, \quad k \geq 1, \quad \delta_0(v) := P_{V_1} v. \quad (3.36)$$

In [15], the proposed set  $G$  consists of pairs  $(k, \nu)$  that satisfy

$$2^k \leq \alpha_\nu, \quad (3.37)$$

for some appropriately chosen  $(\alpha_\nu)_{\nu \in \mathcal{F}}$ . This leads to an approximation of the form

$$u_G = \sum_{\nu \in \Lambda} P_{V_{2^{k(\nu)}}} u_\nu \phi_\nu(y), \quad (3.38)$$

where  $\Lambda$  is the set of  $\nu$  such that  $\alpha_\nu \geq 1$  and  $k(\nu)$  is the largest  $k$  such that (3.37) holds. This approximation is therefore similar to that proposed in (3.11), and optimizing the choice of the sequence  $(\alpha_\nu)_{\nu \in \mathcal{F}}$  leads to similar convergence rates as given in Theorem 3.1.

## 4 Summability of Taylor expansions

### 4.1 $H^2$ -regularity

Establishing the  $\ell^p$  summability in Theorem 2.1 will be based on a weighted  $\ell^2$  estimate expressed in the following result.

**Theorem 4.1.** *Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers such that (2.4) holds. Assume in addition that the right side  $f$  in (1.1) belongs to  $L^2(D)$  and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{1,\infty}(D)$  and that (2.5) holds. Then*

$$\sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_W)^2 < \infty, \quad \rho^\nu := \prod_{j \geq 1} \rho_j^{\nu_j}. \quad (4.1)$$

Before giving the proof of this theorem, let us explain why it implies Theorem 2.1. With  $0 < q < \infty$  and  $0 < p < 2$  such that  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ , we obtain by Hölder's inequality that

$$\left( \sum_{\nu \in \mathcal{F}} \|t_\nu\|_W^p \right)^{1/p} \leq \left( \sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_W)^2 \right)^{1/2} \left( \sum_{\nu \in \mathcal{F}} \rho^{-q\nu} \right)^{1/q}. \quad (4.2)$$

From the factorization

$$\sum_{\nu \in \mathcal{F}} \rho^{-q\nu} = \prod_{j \geq 1} \sum_{n \geq 0} \rho_j^{-nq} = \prod_{j \geq 1} \frac{1}{1 - \rho_j^{-q}}, \quad (4.3)$$

we easily conclude that the second factor on the right of (4.2) is finite under the assumption that the  $\rho_j$  are strictly larger than 1 and  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$ . Therefore  $(\|t_\nu\|_W)_{\nu \in \mathcal{F}}$  belongs to  $\ell^p(\mathcal{F})$ .

In the next proof we use the observation that for a given instance of the elliptic PDE (1.1), when  $f \in L^2$  and  $a \in W^{1,\infty}$  is a strictly positive function, the equation can be written in the strong form

$$-a\Delta u - \nabla a \cdot \nabla u = f, \quad (4.4)$$

where all terms in this identity belong to  $L^2(D)$ .

**Proof of Theorem 4.1:** We first observe that it suffices to prove the result in the case where  $\rho_j = 1$  for all  $j \geq 1$ . Indeed, the condition (2.4) for a general positive sequence  $\rho$  is

equivalent to the same condition with all  $\rho_j = 1$  when the functions  $\psi_j$  are replaced by the functions  $\rho_j \psi_j$ . This amounts to considering the solution map

$$y \mapsto \tilde{u}(y) := u(\rho y), \quad \rho y := (\rho_j y_j)_{j \geq 1}. \quad (4.5)$$

Therefore, the result established in this particular case gives that  $\sum_{\nu \in \mathcal{F}} \|\tilde{t}_\nu\|_W^2 < \infty$ , for the Taylor coefficients of  $\tilde{u}$  which are given by  $\tilde{t}_\nu = \rho^\nu t_\nu$ , which is equivalent to (4.1). We thus next assume that  $\rho_j = 1$  for all  $j$  and establish  $\ell^2$  summability of the  $\|t_\nu\|_W$ , under the condition that for some  $0 < \theta < 1$  holds

$$\left\| \sum_{j \geq 1} \frac{|\psi_j|}{\bar{a}} \right\|_{L^\infty} \leq \theta. \quad (4.6)$$

We first observe that since  $0 < \bar{a}_{\min} \leq \bar{a} \leq \bar{a}_{\max} < \infty$ , we have the norm equivalences

$$\bar{a}_{\min} \|v\|_V^2 \leq \int_D \bar{a} |\nabla v|^2 \leq \bar{a}_{\max} \|v\|_V^2, \quad (4.7)$$

and

$$\bar{a}_{\min} \|v\|_W^2 \leq \int_D \bar{a} |\Delta v|^2 \leq \bar{a}_{\max} \|v\|_W^2, \quad (4.8)$$

It will be more convenient to work with the above equivalent norms for  $V$  and  $W$ . We thus introduce the quantities

$$d_\nu := \int_D \bar{a} |\nabla t_\nu|^2 \quad \text{and} \quad c_\nu := \int_D \bar{a} |\Delta t_\nu|^2, \quad (4.9)$$

and prove that  $\sum_{\nu \in \mathcal{F}} c_\nu < \infty$ . For this purpose, we also introduce the quantities

$$D_n := \sum_{|\nu|=n} d_\nu \quad \text{and} \quad C_n := \sum_{|\nu|=n} c_\nu. \quad (4.10)$$

We first recall the argument from [3] showing that  $\sum_{\nu \in \mathcal{F}} d_\nu < \infty$ . Applying  $\frac{1}{|\nu|} \partial^\nu$  at  $y = 0$  to the variational formulation

$$\int_D a(y) \nabla u(y) \cdot \nabla v = \langle f, v \rangle_{V', V}, \quad v \in V, \quad (4.11)$$

yields the recursion identity

$$\int_D \bar{a} \nabla t_\nu \cdot \nabla v = - \sum_{j \in \text{supp}(\nu)} \int_D \psi_j \nabla t_{\nu - e_j} \cdot \nabla v, \quad v \in V, \quad (4.12)$$

with  $e_j = (0, \dots, 0, 1, 0, \dots)$  the Kronecker sequence of index  $j$ . Taking  $v = t_\nu$  as a test function and applying Young's inequality gives

$$d_\nu \leq \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\nabla t_\nu|^2 + \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\nabla t_{\nu - e_j}|^2. \quad (4.13)$$

From (4.6), the first term on the right does not exceed  $\frac{\theta}{2} d_\nu$ . Summing over  $|\nu| = n$ , we thus obtain

$$\begin{aligned} D_n &\leq \frac{\theta}{2} D_n + \frac{1}{2} \sum_{|\nu|=n} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\nabla t_{\nu - e_j}|^2 \\ &= \frac{\theta}{2} D_n + \frac{1}{2} \sum_{|\nu|=n-1} \int_D \left( \sum_{j \geq 1} |\psi_j| \right) |\nabla t_\nu|^2 \\ &\leq \frac{\theta}{2} D_n + \frac{\theta}{2} D_{n-1}, \end{aligned}$$

where we have again used (4.6). This shows that

$$D_n \leq \kappa D_{n-1} \leq \cdots \leq \kappa^n D_0, \quad \kappa = \frac{\theta}{2-\theta} < 1, \quad (4.14)$$

with  $D_0 = d_0 = \int_D \bar{a} |\nabla u(0)|^2 \leq \frac{\|f\|_{V'}^2}{\bar{a}_{\min}} < \infty$ , and in particular that  $\sum_{\nu \in \mathcal{F}} d_\nu = \sum_{n \geq 0} D_n < \infty$ .

We next turn to estimating the quantities  $C_n$ . For this we observe that for any  $y \in U$  the function  $a(y)$  belongs to  $W^{1,\infty}(D)$ . This allows to use the strong form (4.4)

$$-a(y)\Delta u(y) - \nabla a(y) \cdot \nabla u(y) = f, \quad (4.15)$$

where equality holds in  $L^2(D)$ . Applying  $\frac{1}{\nu!} \partial^\nu$  at  $y = 0$ , for any  $\nu \neq 0$ , gives

$$-\bar{a} \Delta t_\nu = \nabla \bar{a} \cdot \nabla t_\nu + \sum_{j \in \text{supp}(\nu)} (\psi_j \Delta t_{\nu-e_j} + \nabla \psi_j \cdot \nabla t_{\nu-e_j}), \quad (4.16)$$

which shows by recursion that  $\Delta t_\nu \in L^2(D)$  for all  $\nu \in \mathcal{F}$ . Integrating against  $\Delta t_\nu$ , and applying Young's inequality we find that, for any  $\varepsilon > 0$ ,

$$\begin{aligned} c_\nu &\leq \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\Delta t_\nu|^2 + \varepsilon \sum_{j \in \text{supp}(\nu)} \int_D |\nabla \psi_j| |\Delta t_\nu|^2 + \varepsilon \int_D |\nabla \bar{a}| |\Delta t_\nu|^2 \\ &\quad + \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\Delta t_{\nu-e_j}|^2 + \frac{1}{4\varepsilon} \sum_{j \in \text{supp}(\nu)} \int_D |\nabla \psi_j| |\nabla t_{\nu-e_j}|^2 + \frac{1}{4\varepsilon} \int_D |\nabla \bar{a}| |\nabla t_\nu|^2. \end{aligned}$$

The first three terms on the right are bounded by  $(\frac{\theta}{2} + \varepsilon B)c_\nu$ , where

$$B := \left\| |\nabla \bar{a}| + \sum_j |\nabla \psi_j| \right\|_{L^\infty} < \infty. \quad (4.17)$$

Summing over  $|\nu| = n$ , and exchanging the summations in  $j$  and  $\nu$  in the last three terms, we thus obtain

$$C_n \leq \left(\frac{\theta}{2} + \varepsilon B\right) C_n + \frac{\theta}{2} C_{n-1} + \frac{B}{4\varepsilon \bar{a}_{\min}} (D_n + D_{n-1}). \quad (4.18)$$

We next choose  $\varepsilon > 0$  small enough so that  $\frac{\theta}{2} + \varepsilon B < \frac{1}{2}$ , so that

$$C_n \leq \tau C_{n-1} + A(D_n + D_{n-1}), \quad \tau := \frac{\theta}{2-\theta-2\varepsilon B} < 1, \quad A := \frac{B}{2\varepsilon \bar{a}_{\min}(2-\theta-2\varepsilon B)}. \quad (4.19)$$

Since we have already seen that  $D_n \leq D_0 \kappa^n$ , we find that

$$C_n \leq \tau C_{n-1} + AD_0(1 + \kappa^{-1})\kappa^n. \quad (4.20)$$

We now choose a  $\delta$  such that

$$\kappa \leq \tau < \delta < 1, \quad (4.21)$$

and observe that  $C_{n-1} \leq C\delta^{n-1}$  implies that  $C_n \leq (\tau\delta^{-1}C + AD_0(1 + \kappa^{-1}))\delta^n$ . We thus find by induction that

$$C_n \leq \max \left\{ C_0, \frac{AD_0(1 + \kappa^{-1})}{1 - \tau\delta^{-1}} \right\} \delta^n, \quad (4.22)$$

and in particular that  $\sum_{\nu \in \mathcal{F}} c_\nu = \sum_{n \geq 0} C_n < \infty$ .  $\square$

## 4.2 Higher-order regularity

Similar to  $W$  or  $H^2$  regularity, the proof of  $\ell^p$  summability in Theorem 2.3 can be derived by the same argument from weighted  $\ell^2$  estimate expressed in the following result, analogous to Theorem 4.1.

**Theorem 4.2.** *Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess\,inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers such that (2.4) holds. Assume in addition that the right side  $f$  in (1.1) belongs to  $H^{k-2}(D)$  for integer  $k \geq 2$ , that the domain  $D$  has  $C^{k-2,1}$  smoothness, that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{k-1,\infty}(D)$ , and that (2.8) holds. Then*

$$\sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_{W^k})^2 < \infty, \quad \rho^\nu := \prod_{j \geq 1} \rho_j^{\nu_j}. \quad (4.23)$$

In proving this result, we use the following fact on elliptic regularity for a single instance of the elliptic PDE (1.1). Here, we use  $D^\alpha$  to denote the partial derivative of order  $\alpha$  in the space variable  $x$ , in order to avoid confusion with the partial derivative  $\partial^\nu$  with respect to the parametric variable  $y$ .

**Lemma 4.3.** *Let  $k \geq 2$  be an integer and assume that  $a \in W^{k-1,\infty}(D)$  is a strictly positive function,  $f \in H^{k-2}(D)$  and that the domain  $D$  has  $C^{k-2,1}$  smoothness. Then it holds that  $u \in H^{k-1}(D)$  and that for all  $|\alpha| \leq k-2$ , the equation*

$$-a \Delta D^\alpha u - \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \nabla D^\beta a \cdot \nabla D^{\alpha-\beta} u - \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} D^\beta a \Delta D^{\alpha-\beta} u = D^\alpha f, \quad (4.24)$$

holds in  $L^2(D)$ , where  $\binom{\alpha}{\beta}$  denotes the usual binomial coefficient of multi-indices.

**Proof:** We proceed by induction on  $k$ . The case  $k = 2$  is just the strong form (4.4). For  $k > 2$ , assuming the result at order  $k-1$ , we find that  $D^\alpha \Delta u \in L^2$  for all  $|\alpha| \leq k-3$ , or equivalently  $\Delta u \in H^{k-3}(D)$ . Since  $D$  has  $C^{k-2,1}$  smoothness, and  $f \in H^{k-2}(D) \subset H^{k-3}(D)$ , classical elliptic regularity theory [17, Theorem 2.5.1.1] implies that  $u \in H^{k-1}(D)$ . Then, since  $a \in W^{k-1,\infty}(D)$  we may use the Leibniz rule when applying  $D^\alpha$  to (4.4) for  $|\alpha| \leq k-2$  and we obtain (4.24) where each term belongs to  $L^2(D)$ .  $\square$

**Proof of Theorem 4.2:** We proceed by induction on  $k$ , reducing again to the case where  $\rho_j = 1$  for all  $j$ , as in the proof of Theorem 4.1 which gives the case  $k = 2$ . Since the arguments in the induction step are in essence similar to those used in the proof of this previous theorem, up to more involved notations, we mainly sketch them.

For a given order  $\alpha$  of spatial differentiation, we introduce

$$c_\nu^\alpha := \int_D \bar{a} |\Delta D^\alpha t_\nu|^2, \quad (4.25)$$

and prove that  $\sum_{\nu \in \mathcal{F}} c_\nu^\alpha < \infty$  for all  $|\alpha| = k-2$ . For this purpose, we introduce the quantities

$$C_n^\alpha := \sum_{|\nu|=n} c_\nu^\alpha, \quad (4.26)$$

and prove by induction on  $k$  that

$$C_n^\alpha \leq M_{|\alpha|} \kappa_{|\alpha|}^n, \quad n \geq 0, \quad (4.27)$$

for some constant  $M_k$  and  $0 < \kappa_k < 1$  that can be assumed to be non-decreasing with  $k$ . For  $k = 2$ , this was established in the proof of Theorem 4.1. Note that the induction assumption for all  $|\alpha| \leq k - 3$ , combined with the fact that  $D$  has  $C^{k-2,1}$  smoothness, also means, by elliptic regularity, that

$$\sum_{|\nu|=n} \|t_\nu\|_{H^{k-1}(D)}^2 = \sum_{|\nu|=n} \sum_{|\beta| \leq k-1} \|D^\beta t_\nu\|_{L^2(D)}^2 \leq M_{k-1} \kappa_{k-1}^n, \quad (4.28)$$

up to a multiplicative change in the constants  $M_{k-1}$ .

Using the strong form (4.24) for  $a = a(y)$  and  $u = u(y)$ , and applying  $\frac{1}{\nu!} \partial^\nu$  at  $y = 0$ , for any  $\nu \neq 0$ , gives

$$\begin{aligned} -\bar{a} \Delta D^\alpha t_\nu &= \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} D^\beta \bar{a} \Delta D^{\alpha-\beta} t_\nu + \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \nabla D^\beta \bar{a} \cdot \nabla D^{\alpha-\beta} t_\nu \\ &+ \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \sum_{j \in \text{supp}(\nu)} (D^\beta \psi_j \Delta D^{\alpha-\beta} t_{\nu-e_j} + \nabla D^\beta \psi_j \cdot \nabla D^{\alpha-\beta} t_{\nu-e_j}). \end{aligned}$$

Integrating against  $\Delta D^\alpha t_\nu$ , and applying Young's inequality we find that, for any  $\varepsilon > 0$ ,

$$\begin{aligned} c_\nu^\alpha &\leq \varepsilon \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} \int_D |D^\beta \bar{a}| |\Delta D^\alpha t_\nu|^2 + \frac{1}{4\varepsilon} \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} \int_D |D^\beta \bar{a}| |\Delta D^{\alpha-\beta} t_\nu|^2 \\ &+ \varepsilon \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \int_D |\nabla D^\beta \bar{a}| |\Delta D^\alpha t_\nu|^2 + \frac{1}{4\varepsilon} \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \int_D |\nabla D^\beta \bar{a}| |\nabla D^{\alpha-\beta} t_\nu|^2 \\ &+ \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\Delta D^\alpha t_\nu|^2 + \frac{1}{2} \sum_{j \in \text{supp}(\nu)} \int_D |\psi_j| |\Delta D^{\alpha-\beta} t_{\nu-e_j}|^2 \\ &+ \varepsilon \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} \sum_{j \in \text{supp}(\nu)} \int_D |D^\beta \psi_j| |\Delta D^\alpha t_\nu|^2 + \frac{1}{4\varepsilon} \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} \sum_{j \in \text{supp}(\nu)} \int_D |D^\beta \psi_j| |\Delta D^{\alpha-\beta} t_{\nu-e_j}|^2 \\ &+ \varepsilon \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \sum_{j \in \text{supp}(\nu)} \int_D |\nabla D^\beta \psi_j| |\Delta D^\alpha t_\nu|^2 + \frac{1}{4\varepsilon} \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} \sum_{j \in \text{supp}(\nu)} \int_D |\nabla D^\beta \psi_j| |\nabla D^{\alpha-\beta} t_{\nu-e_j}|^2 \\ &= T_1 + \dots + T_{10}. \end{aligned}$$

After summation over  $|\nu| = n$ , the left side is  $C_n^\alpha$  and the contribution of the odd numbered terms is bounded according to

$$\sum_{|\nu|=n} (T_1 + T_3 + T_5 + T_7 + T_9) \leq \left( \frac{\theta}{2} + B\varepsilon \right) C_n^\alpha, \quad (4.29)$$

where

$$B := \left\| \sum_{\beta \leq \alpha} \binom{\alpha}{\beta} (|\nabla D^\beta a| + \sum_{j \geq 1} |\nabla D^\beta \psi_j|) + \sum_{\beta \leq \alpha, \beta \neq 0} \binom{\alpha}{\beta} (|D^\beta a| + \sum_{j \geq 1} |D^\beta \psi_j|) \right\|_{L^\infty} < \infty. \quad (4.30)$$

The contribution of  $T_6$  is bounded by

$$\sum_{|\nu|=n} T_6 \leq \frac{\theta}{2} C_{n-1}^\alpha. \quad (4.31)$$



The contributions of the remaining terms can be bounded by

$$\sum_{|\nu|=n} (T_2 + T_4 + T_8 + T_{10}) \leq A \sum_{|\beta| \leq k-1} \left( \sum_{|\nu|=n} \|D^\beta t_\nu\|_{L^2}^2 + \sum_{|\nu|=n-1} \|D^\beta t_\nu\|_{L^2}^2 \right), \quad (4.32)$$

where  $A$  is a finite constant that depends on  $\varepsilon$ ,  $\bar{a}$ ,  $\psi_j$ ,  $k$  and  $m$ . From the induction assumption (4.28) we thus have

$$\sum_{|\nu|=n} (T_2 + T_4 + T_8 + T_{10}) \leq AM_{k-1}(\kappa_{k-1}^n + \kappa_{k-1}^{n-1}). \quad (4.33)$$

We take  $\varepsilon > 0$  small enough so that  $\frac{\theta}{2} + B\varepsilon < 1$  and therefore obtain that

$$C_n^\alpha \leq \kappa C_{n-1}^\alpha + \frac{2AM_{k-1}}{2-\theta-2B\varepsilon} (1 + \kappa_{k-1}^{-1}) \kappa_{k-1}^n. \quad \kappa := \frac{\theta}{2-\theta-2B\varepsilon} < 1. \quad (4.34)$$

By a similar argument as in the proof of Theorem 4.1, this yields the existence of constants  $M_k$  and  $\kappa_{k-1} \leq \kappa_k < 1$  such that

$$C_n^\alpha \leq M_k \kappa_k^n, \quad n \geq 0, \quad (4.35)$$

which is the claimed estimate at order  $k$ .  $\square$

### 4.3 Fractional and weighted regularity

The summability results expressed by Theorem 2.3 are derived for Sobolev spaces of integer order. In view of the induction argument on  $k$  used in the proof, we cannot treat fractional Sobolev spaces in a similar way. However, such spaces are well known to be obtainable by interpolation of Sobolev spaces of integer order. For example, when using the real interpolation method, for  $0 < \theta < 1$  we have

$$H^{1+\theta(k-1)}(D) = [H^1(D), H^k(D)]_{\theta,2}, \quad (4.36)$$

as well as

$$V \cap H^{1+\theta(k-1)}(D) = [V, V \cap H^k(D)]_{\theta,2}. \quad (4.37)$$

More generally, for any  $0 < \theta < 1$  and  $0 < r \leq \infty$ , we may consider the real interpolation space  $[V, W^k]_{\theta,r}$ , or the complex interpolation space  $[V, W^k]_\theta$ . We refer to [8] for a general introduction to interpolation spaces. If  $Z$  is any of these spaces, we have the interpolation inequality

$$\|t_\nu\|_Z \leq C \|t_\nu\|_V^{1-\theta} \|t_\nu\|_{W^k}^\theta. \quad (4.38)$$

The following result generalizes Theorem 2.3 to these smoothness spaces.

**Corollary 4.4.** *Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers such that (2.4) holds. Assume in addition that the right side  $f$  in (1.1) belongs to  $H^{k-2}(D)$ , that the domain  $D$  has  $C^{k-2,1}$  smoothness, and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{k-1,\infty}(D)$ . Assume that (2.8) holds with the sequence  $\rho$  replaced by a different sequence  $\bar{\rho} = (\bar{\rho}_j)_{j \geq 1}$  of positive numbers such that  $\bar{\rho}_j \leq \rho_j$ . Then for any  $0 < \theta < 1$  and  $0 < r \leq \infty$  we have*

$$\sum_{\nu \in \mathcal{F}} (\bar{\rho}^\nu \|t_\nu\|_Z)^2 < \infty, \quad \bar{\rho}^\nu := \prod_{j \geq 1} \bar{\rho}_j^{\nu_j}, \quad (4.39)$$

where  $Z = [V, W^k]_{\theta,r}$  or  $Z = [V, W^k]_\theta$ , and  $\bar{\rho}_j = \rho_j^{1-\theta} \bar{\rho}_j^\theta$ .

**Proof.** By our assumptions, we have

$$\sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_V)^2 < \infty \quad \text{and} \quad \sum_{\nu \in \mathcal{F}} (\bar{\rho}^\nu \|t_\nu\|_{W^k})^2 < \infty. \quad (4.40)$$

Using the interpolation inequality (4.38), we obtain

$$\sum_{\nu \in \mathcal{F}} (\bar{\rho}^\nu \|t_\nu\|_Z)^2 \leq C^2 \sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_V)^{2(1-\theta)} (\bar{\rho}^\nu \|t_\nu\|_{W^k})^{2\theta} < \infty, \quad (4.41)$$

by Hölder's inequality.  $\square$

As already noted, the spaces  $W^k$  coincide with the Sobolev space  $V \cap H^k$  only when the domain  $D$  has  $C^{k-1,1}$ -smooth boundary. In addition, Theorem 4.2 requires that  $D$  has at least  $C^{k-2,1}$ -smooth boundary. One way to circumvent this limitation is to search for analogous results when the spaces  $W^k$  are replaced by suitable *weighted Sobolev spaces*. Such spaces are particularly relevant to the case where  $D$  is a polygon or polyhedron, possibly with re-entrant corners. We discuss the adaptation of Theorems 4.1 and 4.2 to this situation.

To simplify the exposition, we confine the discussion to space dimension  $m = 2$  and shall assume in the following that  $D \subset \mathbb{R}^2$  is a bounded polygon with  $J$  vertices. For  $j = 1, \dots, J$ , we denote by  $\mathbf{c}_j$  these vertices and define for any  $x \in D$  the truncated distances

$$r_j(x) := 1 \wedge \text{dist}(x, \mathbf{c}_j). \quad (4.42)$$

For  $\beta \in \mathbb{R}$ , we define

$$\Phi_\beta(x) := \prod_{j \geq 1} r_j(x)^\beta. \quad (4.43)$$

For any integer  $k \geq 0$  and  $\theta \in \mathbb{R}$ , we define the Kondrat'ev spaces

$$K_\theta^k(D) := \{u : \Phi_{|\alpha|-\theta} \partial_x^\alpha u \in L^2(D) : |\alpha| \leq k\}. \quad (4.44)$$

For  $k = 0$ , we also write  $L_\theta^2(D) = K_\theta^0(D)$ . These weighted spaces are relevant to us for two reasons. First, the Dirichlet problem for the Poisson equation in a polygon  $D$  admits a shift theorem in these weighted spaces: for all  $k \geq 2$ , there exists  $\eta > 0$  depending on  $D$  and a constant  $C$  depending on  $D$  and  $k$ , such that, for  $|\theta| < \eta$ ,

$$\|u\|_{K_{\theta+1}^k} \leq C \|\Delta u\|_{K_{\theta-1}^{k-2}}. \quad (4.45)$$

Second, the approximation property (3.9) remains valid for the space  $X = K_{\theta+1}^k(D)$  with the same rate  $t = \frac{k-1}{2}$  as for  $X = H^k(D)$ , if we use Lagrange finite element spaces  $(V_n)_{n>0}$  of degree at least  $k-1$  and with appropriate mesh refinement near the corners of  $D$  (see [5, 16], [1, Eqn. (0.3)] and the references there). In the case  $k = 2$ , it is known that  $W \subset K_{\theta+1}^2(D)$  for  $|\theta|$  small enough. Specifically, for the homogeneous Dirichlet boundary conditions presently considered, when  $D$  is a polygon with straight sides,  $|\theta| < \pi/\omega_{max}$  where  $\omega_{max} \in ]0, 2\pi]$  denotes the maximal interior opening angle at the corners of  $D$ . Therefore, the same assumptions as those in Theorem 2.1 imply the result

$$\sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_{K_{\theta+1}^2})^2 < \infty, \quad (4.46)$$

in the case where  $D$  is a polygon. One can adapt the proof of Theorem 2.1 in order to show that (4.46) holds under the assumption

$$\left\| \sum_{j \geq 1} \rho_j \Phi_\gamma |\nabla \psi_j| \right\|_{L^\infty} < \infty, \quad (4.47)$$

where  $\gamma = \max\{1 - \theta, 0\}$ . For  $|\theta| < 1$  (e.g., for polygonal domains  $D$  with reentrant corners), this assumption is slightly weaker than (2.5) due to the presence of weight functions  $\Phi_\gamma$  which vanish at the corners of the domain, thereby allowing the  $\psi_j$  to be singular at these points. For  $k > 2$ , we conjecture that, under similar assumptions, Theorem 4.2 extends to a polygonal domain with  $W^k$  replaced by  $K_{\theta+1}^k(D)$  for  $|\theta|$  small enough.

## 5 Jacobi expansions

We now prove Theorem 2.4, transferring the results for Taylor expansions in the previous sections to Jacobi series. The corresponding univariate Jacobi polynomials  $(J_k)_{k \geq 0}$  are orthonormal with respect to the probability measure  $d_{\alpha,\beta}(t) dt$  on  $[-1, 1]$  with  $\alpha, \beta > -1$ , where

$$d_{\alpha,\beta}(t) = \frac{\Gamma(\alpha + \beta + 2)}{2^{\alpha+\beta+1} \Gamma(\alpha + 1) \Gamma(\beta + 1)} (1-t)^\alpha (1+t)^\beta. \quad (5.1)$$

For the corresponding orthonormal polynomials, one has the Rodrigues' formula

$$J_k(t) = \frac{c_k^{\alpha,\beta}}{k! 2^k} (1-t)^{-\alpha} (1+t)^{-\beta} \frac{d^k}{dt^k} ((t^2 - 1)^k (1-t)^\alpha (1+t)^\beta), \quad (5.2)$$

where  $c_0^{\alpha,\beta} = 1$  and

$$c_k^{\alpha,\beta} = \sqrt{\frac{(2k + \alpha + \beta + 1) k! \Gamma(k + \alpha + \beta + 1) \Gamma(\alpha + 1) \Gamma(\beta + 1)}{\Gamma(k + \alpha + 1) \Gamma(k + \beta + 1) \Gamma(\alpha + \beta + 2)}}, \quad k \in \mathbb{N}. \quad (5.3)$$

Notable special cases are Legendre polynomials for the uniform measure, where  $\alpha = \beta = 0$ , and Chebyshev polynomials, where  $\alpha = \beta = -\frac{1}{2}$ . On  $U$ , we consider the product measure

$$d\sigma(y) = \bigotimes_{j \geq 1} d_{\alpha,\beta}(y_j) dy_j, \quad (5.4)$$

and recall that the tensor product polynomials  $J_\nu(y) = \prod_{j \geq 1} J_{\nu_j}(y_j)$ ,  $\nu \in \mathcal{F}$ , are an orthonormal basis of  $L^2(U, \sigma)$ . Our aim is now to quantify the summability of  $W^k$  norms of Jacobi coefficients

$$v_\nu = \int_U u(y) J_\nu(y) d\sigma(y), \quad \nu \in \mathcal{F}, \quad (5.5)$$

as stated in Theorem 2.4. This assertion on  $\ell^p$  summability will again be derived from a result on weighted  $\ell^2$  summability.

**Theorem 5.1.** *Under the assumptions of Theorem 2.1 when  $k = 2$  and of Theorem 2.3 when  $k > 2$ ,*

$$\sum_{\nu \in \mathcal{F}} (a_\nu^{-1} \rho^\nu \|v_\nu\|_{W^k})^2 < \infty, \quad a_\nu := \prod_{j \geq 1} c_{\nu_j}^{\alpha,\beta}. \quad (5.6)$$

**Proof.** Closely following the proof of [3, Thm. 3.1], for  $y, z \in U$ , we set  $T_y z := (y_j + (1 - |y_j|)\rho_j z_j)_{j \geq 1}$  and  $w_y(z) := u(T_y z)$ . For each  $y \in U$ , let  $\bar{a}_y := a(y) = \bar{a} + \sum_{j \geq 1} y_j \psi_j$  and  $\psi_{y,j} := (1 - |y_j|)\rho_j \psi_j$ . Then  $w_y$  solves the modified affine-parametric problem

$$\int_D \left( \bar{a}_y + \sum_{j \geq 1} z_j \psi_{y,j} \right) \nabla w_y(z) \cdot \nabla v dx = \int_D f v dx, \quad v \in V, \quad (5.7)$$

with Taylor coefficients

$$t_{y,\nu} := \frac{1}{\nu!} \partial^\nu w_y(0) = \frac{1}{\nu!} \left( \prod_{j \geq 1} (1 - |y_j|)^{\nu_j} \right) \rho^\nu \partial^\nu u(y). \quad (5.8)$$

We have the following  $y$ -uniform bounds: on the one hand,

$$\|\bar{a}_y^{-1}\|_{L^\infty(D)} \leq \left\| \left( \bar{a} - \sum_{j \geq 1} |\psi_j| \right)^{-1} \right\|_{L^\infty(D)} < \infty, \quad (5.9)$$

as well as

$$\|D^\alpha \bar{a}_y\|_{L^\infty(D)} \leq \left\| D^\alpha \bar{a} + \sum_{j \geq 1} |D^\alpha \psi_j| \right\|_{L^\infty(D)} < \infty, \quad |\alpha| \leq k-1; \quad (5.10)$$

on the other hand,

$$\begin{aligned} \left\| \frac{\sum_{j \geq 1} |\psi_{y,j}|}{\bar{a}_y} \right\|_{L^\infty(D)} &\leq \left\| \frac{\sum_{j \geq 1} \rho_j |\psi_j| - \sum_{j \geq 1} \rho_j |y_j| |\psi_j|}{\bar{a} - \sum_{j \geq 1} |y_j| |\psi_j|} \right\|_{L^\infty(D)} \\ &\leq \left\| \frac{\sum_{j \geq 1} \rho_j |\psi_j|}{\bar{a}} \right\|_{L^\infty(D)} < 1, \end{aligned} \quad (5.11)$$

and

$$\left\| \sum_{j \geq 1} |D^\alpha \psi_{y,j}| \right\|_{L^\infty(D)} \leq \left\| \sum_{j \geq 1} \rho_j |D^\alpha \psi_j| \right\|_{L^\infty(D)} < \infty, \quad |\alpha| \leq k-1. \quad (5.12)$$

By Theorem 4.1 for  $k = 2$  and Theorem 4.2 for  $k > 2$ , we thus obtain

$$\sum_{\nu \in \mathcal{F}} \left\| \frac{1}{\nu!} \partial^\nu w_y(0) \right\|_{W^k}^2 = \sum_{\nu \in \mathcal{F}} \|t_{y,\nu}\|_{W^k}^2 \leq C < \infty, \quad (5.13)$$

with  $C > 0$  independent of  $y$ . Moreover, by (5.2) and integration by parts,

$$v_\nu = \int_U u(y) J_\nu(y) d\sigma(y) = a_\nu \int_U \frac{1}{\nu!} \partial^\nu u(y) \prod_{j \geq 1} \frac{(1 - y_j^2)^{\nu_j}}{2^{\nu_j}} d\sigma(y), \quad (5.14)$$

and consequently

$$\begin{aligned} \sum_{\nu \in \mathcal{F}} a_\nu^{-2} \rho^{2\nu} \|v_\nu\|_{W^k}^2 &\leq \int_U \sum_{\nu \in \mathcal{F}} \rho^{2\nu} \left\| \frac{1}{\nu!} \partial^\nu u(y) \right\|_{W^k}^2 \prod_{j \geq 1} (1 - |y_j|)^{2\nu_j} \frac{(1 + |y_j|)^{2\nu_j}}{2^{2\nu_j}} d\sigma(y) \\ &\leq \int_U \sum_{\nu \in \mathcal{F}} \left\| \frac{1}{\nu!} \partial^\nu w_y(0) \right\|_{W^k}^2 d\sigma(y) \leq C, \end{aligned} \quad (5.15)$$

with  $C$  as in (5.13). □

**Proof of Theorem 2.4.** From Theorem 5.1, we obtain the  $\ell^p$  summability of  $(\|v_\nu\|_{W^k})_{\nu \in \mathcal{F}}$  using a Hölder estimate analogous to (4.2) which was used in the proof of Theorems 2.1 and 2.3, but this time using

$$\left( \prod_{j \geq 1} a_\nu \rho_j^{-\nu_j} \right)_{\nu \in \mathcal{F}} \in \ell^q(\mathcal{F}). \quad (5.16)$$

To see that this holds true under the given assumptions, note first that for some  $c, t > 0$  depending on  $\alpha, \beta$ , we have  $c_k^{\alpha, \beta} \leq 1 + ck^t$  for  $k \in \mathbb{N}_0$ , and

$$\sum_{\nu \in \mathcal{F}} \rho^{-q\nu} \prod_{j \geq 1} (1 + c\nu_j^t)^q = \prod_{j \geq 1} \left( \sum_{k=0}^{\infty} \rho_j^{-qk} (1 + ck^t)^q \right) \leq \prod_{j \geq 1} (1 + C\rho_j^{-q}), \quad (5.17)$$

where in the last step we have used that  $\rho_{\min} := \min_{j \geq 1} \rho_j > 1$ , with  $C > 0$  depending on  $c, t, q$ , and  $\rho_{\min}$ . The infinite product on the right converges precisely when  $(\rho_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$ , and we thus have (5.16).  $\square$

**Remark 5.2.** The above arguments also apply to tensor product Jacobi measures with variable parameter sequences  $(\alpha_j)_{j \geq 1}, (\beta_j)_{j \geq 1}$  with  $\alpha_j, \beta_j > -1$ , that is,

$$d\sigma(y) = \bigotimes_{j \geq 1} d_{\alpha_j, \beta_j}(y_j) dy_j, \quad (5.18)$$

provided that there exists a fixed polynomial  $Q$  such that  $c_k^{\alpha_j, \beta_j} \leq Q(k)$ ,  $k \in \mathbb{N}_0$ . Since  $Q$  can be chosen to satisfy  $Q(0) = 1$  without loss (up to increasing the degree by one), the estimate (5.17) remains valid also in this case.

**Remark 5.3.** The summability results established in §4.3 for the fractional and weighted Sobolev norms of Taylor coefficients also easily extend to the Jacobi coefficients by similar arguments. In particular, under the assumptions of Theorem 4.4, we also obtain

$$\sum_{\nu \in \mathcal{F}} (a_\nu^{-1} \tilde{\rho}^\nu \|v_\nu\|_Z)^2 < \infty \quad (5.19)$$

with  $\tilde{\rho}_j$  as defined there.

## 6 Hermite expansions

We now turn to lognormal diffusion coefficients of the form  $a = \exp(b)$  and the proof of Theorem 2.5. This proof is more involved than those of the previous results, but follows in part a similar route as the proof of Theorem 1.2 from [4]. For this reason, we only sketch below the arguments when they are the same, and we detail the part that differs significantly.

By similar arguments as in [4, §2], one shows that under the assumptions of Theorem 2.5, one has for almost every  $y \in U = \mathbb{R}^N$  in the sense of the Gaussian product measure  $\gamma$ ,

$$\|b(y)\|_{L^\infty} < \infty \quad \text{and} \quad \|\nabla b(y)\|_{L^\infty} < \infty, \quad (6.1)$$

where we have used the notation  $\|\nabla \psi\|_{L^\infty} := \|\nabla \psi\|_{L^\infty}$ , from which one concludes  $u(y) \in W$  for such  $y$ . In addition, one has the moment bounds

$$\mathbb{E}(\exp(k\|b\|_{L^\infty})), \mathbb{E}(\exp(k\|\nabla b\|_{L^\infty})), \mathbb{E}(\|u\|_V^k), \mathbb{E}(\|u\|_W^k) < \infty \quad (6.2)$$

for all  $0 \leq k < \infty$ . Moreover, following the lines of [4, Theorem 3.3], we also obtain

$$\sum_{\nu \in \mathcal{F}} b_\nu \|u_\nu\|_W^2 = \sum_{\|\mu\|_{\ell^\infty} \leq r} \frac{\rho^{2\mu}}{\mu!} \int_U \|\partial^\mu u(y)\|_W^2 d\gamma(y), \quad (6.3)$$

where

$$b_\nu := \sum_{\|\mu\|_{\ell^\infty} \leq r} \binom{\nu}{\mu} \rho^{2\mu}. \quad (6.4)$$

Here, we recall from [4] the notation

$$\binom{\nu}{\mu} := \prod_{j \geq 1} \binom{\nu_j}{\mu_j}, \quad \mu, \nu \in \mathcal{F} \quad (6.5)$$

with the convention

$$\binom{n}{m} := 0, \quad \text{if } m > n. \quad (6.6)$$

The central estimate for the proof of Theorem 2.5 is the following.

**Theorem 6.1.** *Let  $r \geq 1$  be an integer, and let  $(\rho_j)_{j \geq 1}$  be a positive sequence such that*

$$K := \left\| \sum_{j \geq 1} \rho_j |\psi_j| \right\|_{L^\infty(D)} < \frac{\ln \theta}{\sqrt{r}}, \quad \theta := 1 + \left(1 - \frac{1}{\sqrt{2}}\right)^2, \quad (6.7)$$

and

$$\hat{K} := \left\| \sum_{j \geq 1} \rho_j |\nabla \psi_j| \right\|_{L^\infty(D)} < \infty. \quad (6.8)$$

Then

$$\sum_{\|\mu\|_{\ell^\infty} \leq r} \frac{\rho^{2\mu}}{\mu!} \int_U \|\partial^\mu u(y)\|_W^2 d\gamma(y) < \infty. \quad (6.9)$$

**Proof.** We first establish a bound pointwise a.e. in  $y$ . Let thus  $y \in U$  with  $\|b(y)\|_{L^\infty} < \infty$  and  $\|\nabla b(y)\|_{L^\infty} < \infty$  be fixed. Since  $a(y) \in W^{1,\infty}(D)$  and  $u(y) \in W$ , we have

$$a(y)\Delta u(y) = -f - \nabla a(y) \cdot \nabla u(y) \quad (6.10)$$

in  $L^2(D)$ . Similarly to [4, Lemma 3.1], using the notation

$$S_\mu := \{\nu \in \mathcal{F} : \nu \leq \mu \text{ and } \nu \neq \mu\}, \quad \mu \in \mathcal{F}, \quad (6.11)$$

for  $\mu \neq 0$  we obtain

$$a(y)\Delta \partial^\mu u(y) = - \sum_{\nu \in S_\mu} \binom{\mu}{\nu} \partial^{\mu-\nu} a(y) \Delta \partial^\nu u(y) - \sum_{0 \leq \nu \leq \mu} \binom{\mu}{\nu} \nabla \partial^{\mu-\nu} a(y) \cdot \nabla \partial^\nu u(y). \quad (6.12)$$

We now establish bounds for

$$\sigma_k := \sum_{\mu \in \Lambda_k} \frac{\rho^{2\mu}}{\mu!} \int_D a(y) |\nabla \partial^\mu u(y)|^2 dx, \quad \hat{\sigma}_k := \sum_{\mu \in \Lambda_k} \frac{\rho^{2\mu}}{\mu!} \int_D a(y) |\Delta \partial^\mu u(y)|^2 dx, \quad (6.13)$$

where

$$\Lambda_k := \{\mu \in \mathcal{F} : |\mu| = k, \|\mu\|_{\ell^\infty} \leq r\}. \quad (6.14)$$

Under the given assumptions, [4, Theorem 4.1] yields

$$\sigma_k \leq \delta^k \sigma_0 \quad \text{with } \delta := \sqrt{r}K / \ln \theta < 1, \quad (6.15)$$

and it remains to bound  $\hat{\sigma}_k$ . Note first that  $\partial^{\mu-\nu} a(y) = a(y)\psi^{\mu-\nu}$  and

$$\nabla(a(y)\psi^{\mu-\nu}) = a(y)\psi^{\mu-\nu} \nabla b(y) + a(y) \sum_{j \in \text{supp}(\mu-\nu)} (\mu_j - \nu_j) \psi^{\mu-\nu-e_j} \nabla \psi_j. \quad (6.16)$$

Here and further we use the notation

$$\psi^\nu := \prod_{j \geq 1} \psi_j^{\nu_j} \quad \text{and} \quad |\psi|^\nu := \prod_{j \geq 1} |\psi_j|^{\nu_j}. \quad (6.17)$$

Thus multiplication of (6.12) by  $\rho^{2\mu} \Delta \partial^\mu u(y) / \mu!$ , integration over  $D$ , and summation over  $\mu \in \Lambda_k$  yield

$$\hat{\sigma}_k \leq S_1 + S_2 + S_3 + S_4, \quad (6.18)$$

where

$$S_1 := \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \int_D \frac{\rho^{2\mu} |\psi|^{\mu-\nu}}{(\mu-\nu)! \nu!} a(y) |\Delta \partial^\nu u(y)| |\Delta \partial^\mu u(y)| dx, \quad (6.19)$$

$$S_2 := \sum_{\mu \in \Lambda_k} \frac{\rho^{2\mu}}{\mu!} \int_D a(y) |\nabla b(y)| |\nabla \partial^\mu u(y)| |\Delta \partial^\mu u(y)| dx, \quad (6.20)$$

$$S_3 := \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \int_D \frac{\rho^{2\mu} |\psi|^{\mu-\nu}}{(\mu-\nu)! \nu!} a(y) |\nabla b(y)| |\nabla \partial^\nu u(y)| |\Delta \partial^\mu u(y)| dx, \quad (6.21)$$

$$S_4 := \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \sum_{j \in \text{supp}(\mu-\nu)} \int_D \frac{\rho^{2\mu} |\psi|^{\mu-\nu-e_j}}{(\mu-\nu-e_j)! \nu!} a(y) |\nabla \psi_j| |\nabla \partial^\nu u(y)| |\Delta \partial^\mu u(y)| dx. \quad (6.22)$$

Introducing the functions

$$\varepsilon(\mu, \nu) = \sqrt{\frac{\mu!}{\nu!}} \frac{\rho^{\mu-\nu} |\psi|^{\mu-\nu}}{(\mu-\nu)!}, \quad (6.23)$$

we can proceed exactly as in the proof of [4, Theorem 4.1] to show

$$S_1 = \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \int_D \varepsilon(\mu, \nu) a(y) \frac{\rho^\nu |\Delta \partial^\nu u(y)|}{\sqrt{\nu!}} \frac{\rho^\mu |\Delta \partial^\mu u(y)|}{\sqrt{\mu!}} dx \leq \left( \sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell}}{(k-\ell)!} \hat{\sigma}_\ell \right)^{\frac{1}{2}} \hat{\sigma}_k^{\frac{1}{2}}. \quad (6.24)$$

For  $S_2$ , by the Cauchy-Schwarz inequality we immediately obtain

$$S_2 \leq \|\nabla b(y)\|_{L^\infty} \sqrt{\sigma_k \hat{\sigma}_k}. \quad (6.25)$$

Proceeding again as in [4, Theorem 4.1] also gives

$$\begin{aligned} S_3 &= \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \int_D \varepsilon(\mu, \nu) a(y) |\nabla b(y)| \frac{\rho^\nu |\nabla \partial^\nu u(y)|}{\sqrt{\nu!}} \frac{\rho^\mu |\Delta \partial^\mu u(y)|}{\sqrt{\mu!}} dx \\ &\leq \|\nabla b(y)\|_{L^\infty} \left( \sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell}}{(k-\ell)!} \sigma_\ell \right)^{\frac{1}{2}} \hat{\sigma}_k^{\frac{1}{2}}. \end{aligned} \quad (6.26)$$

By (6.15),

$$\sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell}}{(k-\ell)!} \sigma_\ell < \delta^k \sigma_0, \quad (6.27)$$

and consequently

$$S_2 + S_3 < 2 \|\nabla b(y)\|_{L^\infty} \delta^{\frac{k}{2}} (\sigma_0 \hat{\sigma}_k)^{\frac{1}{2}}. \quad (6.28)$$

In order to bound  $S_4$ , we introduce

$$\hat{\varepsilon}(\mu, \nu) := \sum_{j \in \text{supp}(\mu-\nu)} \sqrt{\frac{\mu!}{\nu!}} \frac{\rho^{\mu-\nu-e_j} |\psi|^{\mu-\nu-e_j}}{(\mu-\nu-e_j)!} \rho_j |\nabla \psi_j|, \quad (6.29)$$

and thus obtain

$$S_4 = \int_D \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \hat{\varepsilon}(\mu, \nu) a(y) \frac{\rho^\nu |\nabla \partial^\nu u(y)|}{\sqrt{\nu!}} \frac{\rho^\mu |\Delta \partial^\mu u(y)|}{\sqrt{\mu!}} dx. \quad (6.30)$$

By Cauchy-Schwarz on the summations over  $\mu$  and  $\nu$ , the expression on the right is bounded from above by

$$\int_D \left( \sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \hat{\varepsilon}(\mu, \nu) a(y) \frac{|\rho^\nu \nabla \partial^\nu u(y)|^2}{\nu!} \right)^{\frac{1}{2}} \left( \sum_{\mu \in \Lambda_k} a(y) \frac{|\rho^\mu \Delta \partial^\mu u(y)|^2}{\mu!} \sum_{\nu \in S_\mu} \hat{\varepsilon}(\mu, \nu) \right)^{\frac{1}{2}} dx. \quad (6.31)$$

On the one hand, using that  $\mu!/\nu! \leq r^{|\mu-\nu|}$ , we estimate

$$\begin{aligned} \sum_{\nu \in S_\mu} \hat{\varepsilon}(\mu, \nu) &\leq \sum_{\ell=1}^k r^{\ell/2} \sum_{\substack{\nu \in S_\mu \\ |\mu-\nu|=\ell}} \sum_{j \in \text{supp}(\mu-\nu)} \rho_j |\nabla \psi_j| \frac{\rho^{\mu-\nu-e_j} |\psi|^{\mu-\nu-e_j}}{(\mu-\nu-e_j)!} \\ &\leq \sum_{\ell=1}^k r^{\ell/2} \sum_{j \geq 1} \rho_j |\nabla \psi_j| \sum_{|\tau|=\ell-1} \frac{\rho^\tau |\psi|^\tau}{\tau!} \leq \hat{K} \sum_{\ell=1}^k r^{\ell/2} \frac{K^{\ell-1}}{(\ell-1)!} \leq \sqrt{r} \hat{K} e^{\sqrt{r}K}. \end{aligned} \quad (6.32)$$

On the other hand, with  $R_{\nu,k} := \{\mu \in \Lambda_k : \nu \in S_\mu\}$ , we have

$$\sum_{\mu \in \Lambda_k} \sum_{\nu \in S_\mu} \hat{\varepsilon}(\mu, \nu) a(y) \frac{|\rho^\nu \nabla \partial^\nu u(y)|^2}{\nu!} = \sum_{\ell=0}^{k-1} \sum_{\nu \in \Lambda_\ell} a(y) \frac{|\rho^\nu \nabla \partial^\nu u(y)|^2}{\nu!} \sum_{\mu \in R_{\nu,k}} \hat{\varepsilon}(\mu, \nu) \quad (6.33)$$

where, for each  $\nu \in \Lambda_\ell$ ,

$$\begin{aligned} \sum_{\mu \in R_{\nu,k}} \hat{\varepsilon}(\mu, \nu) &\leq \sum_{\mu \in R_{\nu,k}} r^{|\mu-\nu|/2} \sum_{j \in \text{supp}(\mu-\nu)} \rho_j |\nabla \psi_j| \frac{\rho^{\mu-\nu-e_j} |\psi|^{\mu-\nu-e_j}}{(\mu-\nu-e_j)!} \\ &\leq r^{(k-\ell)/2} \sum_{j \geq 1} \rho_j |\nabla \psi_j| \sum_{|\tau|=k-\ell-1} \frac{\rho^\tau |\psi|^\tau}{\tau!} \leq r^{(k-\ell)/2} \hat{K} \frac{K^{k-\ell-1}}{(k-\ell-1)!}. \end{aligned} \quad (6.34)$$

Combining (6.31) with (6.32) and (6.34) and using Cauchy-Schwarz for the integral over  $D$  yields

$$S_4 \leq \sqrt{r} \hat{K} e^{\frac{1}{2}\sqrt{r}K} \left( \sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell-1}}{(k-\ell-1)!} \sigma_\ell \right)^{\frac{1}{2}} \hat{\sigma}_k^{\frac{1}{2}} \leq \sqrt{\frac{\theta r}{\delta}} \hat{K} e^{\frac{1}{2}\sqrt{r}K} \sigma_0^{\frac{1}{2}} \delta^{\frac{k}{2}} \hat{\sigma}_k^{\frac{1}{2}}, \quad (6.35)$$

where we have used that by (6.15),

$$\sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell-1}}{(k-\ell-1)!} \sigma_\ell \leq \delta^{k-1} \sum_{\ell=0}^{k-1} \frac{(\ln \theta)^{k-\ell-1}}{(k-\ell-1)!} \sigma_0 \leq \delta^{k-1} \theta \sigma_0. \quad (6.36)$$

In summary, from (6.24), (6.28), and (6.36) we obtain the recursive estimate

$$\hat{\sigma}_k \leq \left( \sum_{\ell=0}^{k-1} \frac{(\sqrt{r}K)^{k-\ell}}{(k-\ell)!} \hat{\sigma}_\ell \right)^{\frac{1}{2}} \hat{\sigma}_k^{\frac{1}{2}} + C_0^{\frac{1}{2}} \delta^{\frac{k}{2}} \hat{\sigma}_k^{\frac{1}{2}}, \quad (6.37)$$

where

$$C_0 := \left( 2 \|\nabla b(y)\|_{L^\infty} + \sqrt{\frac{\theta r}{\delta}} \hat{K} e^{\frac{1}{2}\sqrt{r}K} \right)^2 \sigma_0. \quad (6.38)$$



We now prove by induction that

$$\hat{\sigma}_k \leq 2\hat{C}_0\delta^k, \quad \hat{C}_0 := \max\{\hat{\sigma}_0, C_0\}. \quad (6.39)$$

This evidently holds true for  $k = 0$ . Assuming that  $\hat{\sigma}_\ell \leq 2\hat{C}_0\delta^\ell$  for  $0 \leq \ell < k$ , either  $\hat{\sigma}_k = 0$ , in which case there is nothing to prove, or

$$\hat{\sigma}_k^{\frac{1}{2}} \leq \left( \sqrt{2} \left( \sum_{\ell=0}^{k-1} \frac{(\ln \theta)^{k-\ell}}{(k-\ell)!} \right)^{\frac{1}{2}} + 1 \right) \hat{C}_0^{\frac{1}{2}} \delta^{\frac{k}{2}}. \quad (6.40)$$

Since

$$\left( \sum_{\ell=0}^{k-1} \frac{(\ln \theta)^{k-\ell}}{(k-\ell)!} \right)^{\frac{1}{2}} \leq \sqrt{2}(e^{\ln \theta} - 1)^{\frac{1}{2}} = \sqrt{2} \left( 1 - \frac{1}{\sqrt{2}} \right) = \sqrt{2} - 1, \quad (6.41)$$

this shows (6.39). In summary, as a consequence of (6.39) we have, by summation over  $k$ ,

$$\begin{aligned} & \sum_{\|\mu\|_{\ell^\infty} \leq r} \frac{\rho^{2\mu}}{\mu!} \int_D a(y) |\Delta \partial^\mu u(y)|^2 dx \\ & \leq C \max \left\{ \int_D a(y) |\Delta u(y)|^2 dx, (1 + \|\nabla b(y)\|_{L^\infty})^2 \int_D a(y) |\nabla u(y)|^2 dx \right\}, \end{aligned} \quad (6.42)$$

where  $C > 0$  depends on  $r, K, \hat{K}, \delta$ , but not on  $y$ . The estimate (6.9) now follows with

$$\|\partial^\mu u(y)\|_W^2 \leq \exp(\|b(y)\|_{L^\infty(D)}) \int_D a(y) |\Delta \partial^\mu u(y)|^2 dx \quad (6.43)$$

and the boundedness of moments (6.2), similarly to [4, Theorem 4.2].  $\square$

Note that under the assumptions of Theorem 2.5, any  $(\rho_j)_{j \geq 1}$  such that (2.9) holds can be rescaled, without affecting the  $\ell^q$  summability of  $(\rho_j^{-1})_{j \geq 1}$ , such that the condition (6.7) in Theorem 6.1 holds. Using Theorem 6.1 and (6.3), we can thus conclude the proof of Theorem 2.5, by means of a Hölder inequality, exactly as in [4, Section 5], choosing  $r$  depending on  $p$ .

## 7 Towards space-parameter adaptivity

The space-parameter approximation results presented in §3 are based on analyzing the error resulting from space discretization of each coefficient  $u_\nu$  in (1.10). The error analysis is based on the available  $\ell^p$  summability for both sequences  $(\|u_\nu\|_V)_{\nu \in \mathcal{F}}$  and  $(\|u_\nu\|_X)_{\nu \in \mathcal{F}}$  where  $X$  is a regularity class that satisfies the spatial *approximation property*

$$\min_{v_n \in V_n} \|v - v_n\|_V \leq C_X n^{-t} \|v\|_X, \quad n > 0, \quad v \in X, \quad (7.1)$$

for a family of  $n$ -dimensional subspaces  $(V_n)_{n \geq 0}$ . We have focused our attention on the Hilbertian Sobolev spaces  $H^s$  for which (7.1) holds with  $t = \frac{s-1}{m}$  when using finite element spaces  $V_n$  of sufficiently high order on either uniformly refined meshes, or locally refined meshes in the case of polygonal domains.

We next would like to consider adaptive space discretizations, such as adaptive finite elements or wavelet methods. In this setting, the space  $V_n$  is no longer a linear space. For instance, in an adaptive wavelet method, it is described as the set of all possible  $n$ -term combinations in the given wavelet basis  $(\psi_\lambda)_{\lambda \in \mathcal{S}}$ , that is,

$$V_n := \left\{ \sum_{\lambda \in E} c_\lambda \psi_\lambda : (c_\lambda)_{\lambda \in E} \in \mathbb{R}^E, \quad \#(E) \leq n \right\}. \quad (7.2)$$

Therefore, when using such spaces, the resulting approximant  $u_{\mathbf{n}}$  given in (3.11), may be rewritten in the space-parameter basis  $(\psi_\lambda \otimes \phi_\nu)_{(\lambda,\nu) \in \mathcal{S} \times \mathcal{F}}$  according to

$$u_{\mathbf{n}} = \sum_{\nu \in \Lambda_n} \sum_{\lambda \in E_\nu} c_{\lambda,\nu} \psi_\lambda \otimes \phi_\nu, \quad (7.3)$$

where  $\#(E_\nu) = n_\nu$ . When imposing the number of degrees of freedom  $N = \sum_{\nu \in \Lambda_n} n_\nu$ , this means that  $u_{\mathbf{n}}$  can be picked as any  $N$ -term approximation

$$u_{\mathbf{n}} = \sum_{(\lambda,\nu) \in G_N} c_{\lambda,\nu} \psi_\lambda \otimes \phi_\nu, \quad (7.4)$$

where  $\#(G_N) = N$ . The error of best  $N$ -term approximation is a natural benchmark for space-parameter adaptive methods, as developed in [2, 20].

One advantage of nonlinear families  $(V_n)_{n \geq 1}$  such as given by (7.2) is that the smoothness conditions that govern a given approximation rate  $n^{-t}$  are substantially weaker than with their linear counterparts. Typical results, see e.g. [11, 14], are the following: if  $D \subset \mathbb{R}^m$  is a bounded Lipschitz domain, for  $s > 1$  and  $t = \frac{s-1}{m}$  the approximation property (7.1) holds for the Sobolev spaces

$$X = W^{s,\tau}(D), \quad \frac{1}{\tau} < \min \left\{ 1, \frac{1}{2} + \frac{s-1}{m} \right\}, \quad (7.5)$$

and more general Besov spaces

$$X = B_q^{s,\tau}(D), \quad \frac{1}{\tau} < \frac{1}{2} + \frac{s-1}{m}. \quad (7.6)$$

The limit case  $\frac{1}{\tau} = \frac{1}{2} + \frac{s-1}{m}$  also holds if  $q \leq \tau$ . These results hold provided that the degree of the chosen finite elements, or degree of polynomial reproduction of the chosen wavelet systems, is larger than  $s-1$ .

Since these spaces are larger than  $H^s$  when  $\tau < 2$ , we may hope that the summability index  $p_X$  of  $(\|u_\nu\|_X)_{\nu \in \mathcal{F}}$  is smaller when using such non-Hilbertian spaces for  $X$ , thereby leading to improved rates of space-parameter approximation when using best  $N$ -term approximations of the form (7.4).

In the following, we discuss this improved summability in a simple case. We give a result for the Taylor and Jacobi coefficients in the case of affine parameter dependence (1.4). For  $1 \leq \tau \leq 2$ , we introduce the Banach space

$$B^\tau := \{v \in V : \Delta u \in L^\tau(D)\}, \quad (7.7)$$

endowed with the norm and semi-norm

$$\|v\|_{B^\tau} = \|v\|_V + |v|_{B^\tau}, \quad |v|_{B^\tau} := \|\Delta v\|_{L^\tau}. \quad (7.8)$$

Note that if  $D$  is convex and  $1 < \tau \leq 2$ , we have by elliptic regularity that  $B^\tau = W^{2,\tau}(D)$ .

**Theorem 7.1.** *Let  $1 \leq \tau < 2$ . Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , and that there exists a sequence  $\rho = (\rho_j)_{j \geq 1}$  of positive numbers and a sequence  $\bar{\rho} = (\bar{\rho}_j)_{j \geq 1}$  of numbers strictly larger than 1, such that*

$$\left\| \frac{\sum_{j \geq 1} \bar{\rho}_j \rho_j |\psi_j|}{\bar{a}} \right\|_{L^\infty} = \theta < 1, \quad (7.9)$$

and that  $(\bar{\rho}_j^{-1})_{j \geq 1} \in \ell^{\bar{q}}(\mathbb{N})$  where  $\frac{1}{2} + \frac{1}{\bar{q}} = \frac{1}{\tau}$ . Assume in addition that the right side  $f$  in (1.1) belongs to  $H^{-1}(D) \cap L^\tau(D)$ , and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{1,\infty}(D)$  where

$$\left\| \sum_{j \geq 1} \rho_j |\nabla \psi_j| \right\|_{L^\infty} < \infty. \quad (7.10)$$

Then

$$\sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_{B^\tau})^\tau < \infty, \quad \rho^\nu := \prod_{j \geq 1} \rho_j^{\nu_j}. \quad (7.11)$$

**Proof.** We take  $\rho_j = 1$ , up to rescaling. We notice that when  $a \in W^{1,\infty}(D)$ , we can write the equation in the strong form  $-a\Delta u = \nabla a \cdot \nabla u + f$ , where all terms in the equality belong to  $H^{-1}(D)$  and to  $L^\tau(D)$ . Hence

$$-a(y)\Delta u(y) = \nabla a(y) \cdot \nabla u(y) + f, \quad (7.12)$$

for all  $y \in U$ , and we can differentiate at  $y = 0$ , which leads to the identities

$$-\bar{a}\Delta t_\nu = \nabla \bar{a} \cdot \nabla t_\nu + \sum_{j \in \text{supp}(\nu)} \nabla \psi_j \cdot \nabla t_{\nu - e_j} + \sum_{j \in \text{supp}(\nu)} \psi_j \Delta t_{\nu - e_j}, \quad (7.13)$$

where all terms in the equality belong to  $H^{-1}(D)$  and to  $L^\tau(D)$ . So pointwise

$$|\Delta t_\nu| \leq C \left( |\nabla t_\nu| + \sum_{j \in \text{supp}(\nu)} \kappa_j |\nabla t_{\nu - e_j}| \right) + \theta \sum_{j \in \text{supp}(\nu)} \omega_j |\Delta t_{\nu - e_j}|, \quad (7.14)$$

where  $\kappa_j := \frac{|\nabla \psi_j|}{\sum_{i \geq 1} |\nabla \psi_i|}$  and  $\omega_j := \frac{|\psi_j|}{\sum_{i \geq 1} |\psi_i|}$  so that  $\sum_{j \geq 1} \kappa_j = \sum_{j \geq 1} \omega_j = 1$ , and where  $C > 1$  is a fixed constant. We elevate to the power  $\tau$  and use the observation that for any  $\varepsilon > 0$ , there exists a constant  $C = C(\varepsilon, \tau) > 1$  such that for any  $a, b \geq 0$ ,

$$(a + b)^\tau \leq C a^\tau + (1 + \varepsilon) b^\tau. \quad (7.15)$$

Taking  $\varepsilon$  small enough, so that  $\bar{\theta} := (1 + \varepsilon)\theta^\tau < 1$ , and using the convexity of  $x \mapsto |x|^\tau$ , we obtain

$$|\Delta t_\nu|^\tau \leq C \left( |\nabla t_\nu|^\tau + \sum_{j \in \text{supp}(\nu)} \kappa_j |\nabla t_{\nu - e_j}|^\tau \right) + \bar{\theta} \sum_{j \in \text{supp}(\nu)} \omega_j |\Delta t_{\nu - e_j}|^\tau, \quad (7.16)$$

for some fixed  $C > 1$ . Integrating and summing over  $|\nu| = n$  thus gives

$$\sum_{|\nu|=n} \|\Delta t_\nu\|_{L^\tau}^\tau \leq C \left( \sum_{|\nu|=n} \|\nabla t_\nu\|_{L^\tau}^\tau + \sum_{|\nu|=n-1} \|\nabla t_\nu\|_{L^\tau}^\tau \right) + \bar{\theta} \sum_{|\nu|=n-1} \|\Delta t_\nu\|_{L^\tau}^\tau. \quad (7.17)$$

Since we have

$$\left\| \frac{\sum_{j \geq 1} \bar{\rho}_j |\psi_j|}{\bar{a}} \right\|_{L^\infty} = \theta < 1, \quad (7.18)$$

with  $(\bar{\rho}_j^{-1})_{j \geq 1} \in \ell^{\bar{q}}(\mathbb{N})$  where  $\frac{1}{2} + \frac{1}{\bar{q}} = \frac{1}{\tau}$ , application of Theorem 1.1 gives us that  $(\|t_\nu\|_V)_{\nu \in \mathcal{F}} \in \ell^\tau(\mathcal{F})$ . Since we assumed  $\tau < 2$ , this implies that  $(\|\nabla t_\nu\|_{L^\tau})_{\nu \in \mathcal{F}} \in \ell^\tau(\mathcal{F})$ . Since  $\bar{\theta} < 1$ , we conclude by summing (7.17) over  $n$  that

$$\sum_{\nu \in \mathcal{F}} \|\Delta t_\nu\|_{L^\tau}^\tau < \infty,$$

which implies the  $\ell^\tau$  summability of  $(\|t_\nu\|_{B^\tau})_{\nu \in \mathcal{F}}$ .  $\square$

**Remark 7.2.** By proceeding in a similar way as in the proof of Theorem 5.1, we may extend the above result to Jacobi coefficients under the same assumptions: one has

$$\sum_{\nu \in \mathcal{F}} (a_\nu^{-1} \rho^\nu \|v_\nu\|_{B^\tau})^\tau < \infty \quad (7.19)$$

with  $a_\nu$  as in Theorem 5.1.

If  $\rho_j > 1$  and  $(\rho_j)_{j \geq 1} \in \ell^q(\mathbb{N})$ , by Hölder's inequality we obtain

$$\sum_{\nu \in \mathcal{F}} \|t_\nu\|_{B^\tau}^p \leq \left( \sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_{B^\tau})^\tau \right)^{\frac{q}{\tau+q}} \left( \sum_{\nu \in \mathcal{F}} \rho^{-q\nu} \right)^{\frac{\tau}{q+\tau}}, \quad \frac{1}{p} = \frac{1}{q} + \frac{1}{\tau}. \quad (7.20)$$

This needs to be compared to (4.2), which based on the same condition (7.10) constraining  $q$  only gives  $\frac{1}{p} = \frac{1}{q} + \frac{1}{2}$ . Concerning the spatial approximation rate  $t$ , we observe the following: for space dimension  $m = 1$ , we directly obtain the rate  $t = \frac{1}{m} = 1$  for nonlinear approximation in  $V$  of elements of  $B^\tau = W^{2,\tau}(D)$  with  $\tau \geq 1$ ; for  $m = 2$ , we still have  $B^\tau = W^{2,\tau}(D)$  for  $\tau > 1$  using the above elliptic regularity result, which then again gives  $t = \frac{1}{m} = \frac{1}{2}$ ; whereas for  $m = 3$ , we obtain  $t = \frac{1}{m} = \frac{1}{3}$  under the stronger condition  $\tau \geq \frac{6}{5}$ . In summary, nonlinear approximation in space gives us the same spatial approximation rate with better summability of the corresponding higher-order norms.

We next use interpolation, similarly to Theorem 4.4, to extend the above results to cases where (7.10) is not satisfied for any  $\rho_j > 1$ .

**Corollary 7.3.** *Let  $1 \leq \tau < 2$ . Assume that  $\bar{a} \in L^\infty(D)$  is such that  $\text{ess inf } \bar{a} > 0$ , that  $f$  in (1.1) belongs to  $H^{-1}(D) \cap L^\tau(D)$  and that  $\bar{a}$  and all functions  $\psi_j$  belong to  $W^{1,\infty}(D)$ . In addition, assume that there exist sequences  $(\rho_j)_{j \geq 1}$ ,  $(\hat{\rho}_j)_{j \geq 1}$  of positive numbers such that*

$$\left\| \frac{\sum_{j \geq 1} \hat{\rho}_j |\psi_j|}{\bar{a}} \right\|_{L^\infty} < 1, \quad \left\| \sum_{j \geq 1} \rho_j |\nabla \psi_j| \right\|_{L^\infty} < \infty, \quad (7.21)$$

where  $\hat{\rho}_j / \rho_j > 1$  and  $(\rho_j / \hat{\rho}_j)_{j \geq 1} \in \ell^q(\mathbb{N})$  with  $\frac{1}{2} + \frac{1}{q} = \frac{1}{\tau}$ . Then with  $\tilde{\rho}_j := \hat{\rho}_j^{1-\theta} \rho_j^\theta$  and  $\tilde{\rho}^\nu := \prod_{j \geq 1} \tilde{\rho}_j^{\nu_j}$ , for all  $0 < \theta < 1$  and  $0 < r \leq \infty$ ,

$$\sum_{\nu \in \mathcal{F}} (\tilde{\rho}^\nu \|t_\nu\|_Z)^\zeta < \infty \quad \text{and} \quad \sum_{\nu \in \mathcal{F}} (a_\nu^{-1} \tilde{\rho}^\nu \|v_\nu\|_Z)^\zeta < \infty, \quad \frac{1}{\zeta} = \frac{1}{2} + \left( \frac{1}{\tau} - \frac{1}{2} \right) \theta, \quad (7.22)$$

where  $Z = [V, B^\tau]_{\theta,r}$  or  $Z = [V, B^\tau]_\theta$ , and where  $a_\nu$  is defined in Theorem 5.1.

**Proof.** From the interpolation inequality (4.38), there exists  $C > 0$  such that

$$\|t_\nu\|_Z \leq C \|t_\nu\|_V^{1-\theta} \|t_\nu\|_{B^\tau}^\theta$$

for all  $\nu \in \mathcal{F}$ . By Hölder's inequality,

$$\begin{aligned} \left( \sum_{\nu \in \mathcal{F}} (\tilde{\rho}^\nu \|t_\nu\|_Z)^\zeta \right)^{\frac{1}{\zeta}} &\leq C \left( \sum_{\nu \in \mathcal{F}} (\hat{\rho}^\nu \|t_\nu\|_V)^{(1-\theta)\zeta} (\rho^\nu \|t_\nu\|_{B^\tau})^{\theta\zeta} \right)^{\frac{1}{\zeta}} \\ &\leq C \left( \sum_{\nu \in \mathcal{F}} (\hat{\rho}^\nu \|t_\nu\|_V)^2 \right)^{\frac{1-\theta}{2}} \left( \sum_{\nu \in \mathcal{F}} (\rho^\nu \|t_\nu\|_{B^\tau})^\tau \right)^{\frac{\theta}{\tau}}, \end{aligned}$$

and the right-hand side is finite by Theorems 1.1 and 7.1. The analogous statement for the Jacobi coefficients follows with Remark 7.2.  $\square$

We now verify that the interpolation spaces considered above indeed have the expected approximation properties.

**Proposition 7.4.** *Let  $0 < \theta < 1$  and  $0 < r \leq \infty$ . Consider a nonlinear family  $(V_n)_{n \geq 1}$  which satisfies the approximation property (7.1) for  $X = B^\tau$  and  $t = \frac{1}{m}$ , for some  $\tau \geq 1$ . Then, for any  $0 < \theta < 1$  and  $0 < r \leq \infty$ , the approximation property*

$$\min_{v_n \in V_n} \|v - v_n\|_V \leq C_Z n^{-\theta/m} \|v\|_Z, \quad n > 0, \quad (7.23)$$

holds for  $Z = [V, B^\tau]_{\theta, r}$  or  $Z = [V, B^\tau]_\theta$ .

**Proof.** Since for any admissible pair  $(V, X)$  of Banach spaces, one has

$$[V, X]_{\theta, 1} \subset [V, X]_\theta \subset [V, X]_{\theta, \infty}, \quad (7.24)$$

it is sufficient to consider  $Z = [V, B^\tau]_{\theta, \infty}$ . Now for any  $v \in V$  and  $w \in X$ , we may write

$$\min_{v_n \in V_n} \|v - v_n\|_V \leq \|v - w\|_V + \min_{v_n \in V_n} \|w - v_n\|_V \leq \|v - w\|_V + C_X n^{-1/m} \|w\|_X. \quad (7.25)$$

Therefore

$$\min_{v_n \in V_n} \|v - v_n\|_V \leq \min_{w \in X} \{\|v - w\|_V + C_X n^{-1/m} \|w\|_X\}. \quad (7.26)$$

The right-hand side is the  $K$ -functional  $K(v, C_X n^{-1/m}, V, X)$ , which by definition of interpolation spaces satisfies

$$K(v, C_X n^{-1/m}, V, X) \leq C_X^\theta n^{-\theta/m} \|v\|_Z \quad (7.27)$$

when  $v \in Z$ , thereby proving (7.23) with  $C_Z := C_X^\theta$ .  $\square$

## 8 Multiresolution representation of $a(y)$

We finally illustrate our results in the particular case of an affine wavelet-type parametrization of the coefficient  $a(y)$ . Our focus will be on Jacobi expansions in  $L^2(U, V, \sigma)$ . Concerning analogous results for Taylor approximation in  $L^\infty(U, V)$ , see Remark 8.6. We refer to [11] for a general treatment of wavelets and their adaptation to a bounded domain  $D \subset \mathbb{R}^m$ , and summarize below the main properties that are needed in our analysis.

**Assumptions 8.1.** *Let  $(\psi_\lambda)_{\lambda \in \mathcal{S}}$  be a family of wavelet basis functions, where the scale-space indices  $\lambda$  comprise dilation and translation parameters, with the convention that the scale  $l$  of  $\psi_\lambda$  is denoted by  $|\lambda| := l$ , and where the number of wavelets on level  $|\lambda| = l$  is proportional to  $2^{ml}$ . Moreover, we assume the wavelets at each given scale to have finite overlap, that is, there exists  $M > 0$  such that for all  $x \in D$  and  $l$ ,*

$$\#\{\lambda : |\lambda| = l \text{ and } \psi_\lambda(x) \neq 0\} \leq M. \quad (8.1)$$

For simplicity, we take  $\bar{a} = 1$ , assume  $D$  and  $f$  to be smooth, and fix an ordering  $(\lambda(j))_{j \geq 1}$  of the indices from coarser to finer scales, for which we set  $\psi_j := \psi_{\lambda(j)}$ . We consider wavelets with  $\psi_\lambda \in W^{\kappa, \infty}(D)$  for some  $\kappa \in \mathbb{N}$ , that are normalized such that

$$\|\psi_\lambda\|_{L^\infty} = C 2^{-\alpha|\lambda|}, \quad (8.2)$$

for an  $\alpha \in (0, \kappa)$  and  $C > 0$  chosen such that (1.6) holds. The partial derivatives behave like

$$\|D^\mu \psi_\lambda\|_{L^\infty} \sim 2^{-(\alpha - |\mu|)|\lambda|}, \quad (8.3)$$

for  $|\mu| \leq \kappa$ .

For ease of exposition, we assume in what follows that  $\alpha$  is not an integer. Under the above assumptions, for  $0 < \beta < \alpha$  we have

$$\sup_{x \in D} \sum_{\lambda} 2^{\beta|\lambda|} |\psi_{\lambda}(x)| < \infty \quad (8.4)$$

as a consequence of (8.1). Therefore, with

$$\rho_j := 1 + c2^{\beta|\lambda(j)|} \sim j^{\beta/m} \quad (8.5)$$

with  $c > 0$  is sufficiently small, we have (2.4), and thus obtain the following from Theorem 1.1.

**Proposition 8.2.** *Under Assumptions 8.1 one has  $(\|v_{\nu}\|_V)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})$  for any  $p > (\frac{\alpha}{m} + \frac{1}{2})^{-1}$ , and consequently*

$$\left\| u - \sum_{\nu \in \Lambda_n} v_{\nu} J_{\nu} \right\|_{L^2(U, V, \sigma)} \leq C n^{-s}, \quad (8.6)$$

holds for any  $s < \frac{\alpha}{m}$ . Here,  $C := \|(\|v_{\nu}\|_V)_{\nu \in \mathcal{F}}\|_{\ell^p}$  with  $s = \frac{1}{p} - \frac{1}{2}$ , and  $\Lambda_n$  is the set of indices  $\nu$  corresponding to the  $n$  largest  $\|v_{\nu}\|_V$ .

## 8.1 Sobolev regularity

In order to make use of our results for higher-order spatial regularity, we observe that for any  $\bar{\beta} < \alpha - (k - 1)$  and  $\bar{\rho}_j := 2^{\bar{\beta}|\lambda(j)|} \sim j^{\bar{\beta}/m}$ , also

$$\sup_{|\mu| \leq k-1} \left\| \sum_{j \geq 1} \bar{\rho}_j |D^{\mu} \psi_j(x)| \right\|_{L^{\infty}} < \infty. \quad (8.7)$$

Based on Theorems 2.3 and 2.4, which concern integer-order Sobolev regularity, and on Theorem 3.1, we obtain the following conclusions for the fully discrete approximations.

**Corollary 8.3.** *Let Assumptions 8.1 hold with  $\alpha > 1$ . Let  $k \in \{2, \dots, \lceil \alpha \rceil\}$ , and let the spatial approximation spaces  $(V_n)_{n>0}$  satisfy the approximation property*

$$\min_{w_n \in V_n} \|w - w_n\|_V \leq C_k n^{-t} \|w\|_{H^k}, \quad (8.8)$$

with  $t = \frac{k-1}{m}$ . Then for any  $r < \frac{\alpha}{2m}$ , there exists a constant  $C > 0$  such that the following holds: for each  $n$  there exists  $\mathbf{n} = (n_{\nu})_{\nu \in \Lambda_n}$  such that

$$\min_{u_{\mathbf{n}} \in V_{\mathbf{n}}} \|u - u_{\mathbf{n}}\|_{L^2(U, V, \sigma)} \leq C N^{-\min\{r, t\}}, \quad (8.9)$$

where  $N := \sum_{\nu \in \Lambda_n} n_{\nu} = \dim(V_{\mathbf{n}})$ .

**Proof.** Theorems 2.3 and 2.4 can be applied when  $k - 1 < \alpha$  with integer  $k \geq 2$ . Since by assumption  $\alpha$  is not an integer, the largest value for  $k$  that we can use is thus  $k = \lceil \alpha \rceil$ . Then Theorems 2.3 and 2.4 yield  $(\|v_{\nu}\|_{H^k})_{\nu \in \mathcal{F}} \in \ell^{\bar{p}}(\mathcal{F})$  for  $\bar{p} > (\frac{\bar{\beta}}{m} + \frac{1}{2})^{-1} > (\frac{\alpha - k + 1}{m} + \frac{1}{2})^{-1}$ . By Theorem 3.1 with  $\mathcal{V} = L^2(U, V, \sigma)$ , we obtain the overall convergence rate  $\min\{r, t\}$ , where

$$r = \frac{st}{t + s + \frac{1}{2} - \bar{p}^{-1}}. \quad (8.10)$$

Taking  $s \rightarrow \frac{\alpha}{m}$  and  $\bar{p} \rightarrow (\frac{\alpha - k + 1}{m} + \frac{1}{2})^{-1}$ , we obtain  $r \rightarrow \frac{\alpha}{2m}$ .  $\square$

**Remark 8.4.** Provided that the approximation property in (8.8) holds for sufficiently large  $k$ , one has  $\min\{r, t\} = r$ . For instance, as soon as  $1 + \frac{1}{2}\alpha \leq k \leq \lceil \alpha \rceil$  (where  $\alpha > 1$ ), the conclusion in the above corollary reads as follows: for any  $\delta > 0$ , there exists a constant  $C > 0$  such that, for each  $n$ , there exists  $\mathbf{n} = (n_\nu)_{\nu \in \Lambda_n}$  such that

$$\min_{\mathbf{n} \in V_{\mathbf{n}}} \|u - u_{\mathbf{n}}\|_{L^2(U, V, \sigma)} \leq CN^{-\frac{\alpha}{2m} + \delta}, \quad (8.11)$$

where  $N := \sum_{\nu \in \Lambda_n} n_\nu = \dim(V_{\mathbf{n}})$ . Conversely, one has  $\min\{r, t\} = t$  in (8.9) only when an artificially low value of  $k$  is used. This corresponds to the situation when one is limited to a certain order of spatial approximation  $t$  that is lower than what could in principle be exploited for the given  $\alpha$ .

We now show that the restriction to  $\alpha > 1$  in Corollary 8.3 can be removed using the interpolation argument from Corollary 4.4. Under Assumptions 8.1, the hypotheses of this corollary are satisfied for  $k := \lceil \alpha \rceil + 1$ , and by our smoothness assumption on  $D$ , for  $0 < \theta < 1$  we have

$$Z := [V, W^k]_{\theta, 2} = [V, V \cap H^k(D)]_{\theta, 2} = V \cap H^{1+\theta(k-1)}(D).$$

On the one hand, we again choose  $\rho_j$  as in (8.5) with positive  $\beta < \alpha$  to ensure (2.4). On the other hand, for  $\bar{\beta} < \alpha - \lceil \alpha \rceil < 0$ , taking  $\bar{\rho}_j := 2^{\bar{\beta}|\lambda(j)|} \leq 1$ , we also have

$$\left\| \sum_{j \geq 1} \bar{\rho}_j |D^\mu \psi_j(x)| \right\|_{L^\infty} < \infty, \quad |\mu| \leq k - 1 = \lceil \alpha \rceil. \quad (8.12)$$

Applying Corollary 4.4 and Remark 5.3, we obtain  $\tilde{\rho}_j > 1$  in (5.19) provided that  $\theta$  is chosen to satisfy  $(1 - \theta)\beta + \theta\bar{\beta} > 0$ , which is ensured to be achievable provided that

$$0 < \theta < \alpha / \lceil \alpha \rceil. \quad (8.13)$$

With suitably chosen  $\beta, \bar{\beta}$  we then obtain  $(\|v_\nu\|_Z)_{\nu \in \mathcal{F}} \in \ell^{\bar{p}}(\mathcal{F})$  for any  $\bar{p}$  such that

$$\bar{p}^{-1} < \frac{\alpha - \theta \lceil \alpha \rceil}{m} + \frac{1}{2}.$$

In Theorem 3.1, for elements of  $Z$  we can achieve the spatial rate  $t = \theta \lceil \alpha \rceil / m$ . As  $\theta \rightarrow \alpha / \lceil \alpha \rceil$ , we thus obtain the following conclusion, extending Corollary 8.3 to  $\alpha \in ]0, 1[$ .

**Corollary 8.5.** *Let (8.8) hold with  $k \geq 2$ , and let Assumptions 8.1 hold with  $0 < \alpha < 1$ . Then, for any  $\delta > 0$ , there exists a constant  $C > 0$  such that the following holds: for each  $n$  there exists  $\mathbf{n} = (n_\nu)_{\nu \in \Lambda_n}$  such that*

$$\min_{\mathbf{n} \in V_{\mathbf{n}}} \|u - u_{\mathbf{n}}\|_{L^2(U, V, \sigma)} \leq CN^{-\frac{\alpha}{2m} + \delta} \quad (8.14)$$

where  $N := \sum_{\nu \in \Lambda_n} n_\nu = \dim(V_{\mathbf{n}})$ .

**Remark 8.6.** For obtaining rates in  $L^\infty(U, V)$ , we need to additionally require  $\alpha > \frac{m}{2}$ . One then has (8.6) for any  $s < \frac{\alpha}{m} - \frac{1}{2}$ , and analogously to Corollaries 8.3 and 8.5, one obtains (8.9) also for Taylor approximation with any  $r < \frac{1}{2} \left( \frac{\alpha}{m} - \frac{1}{2} \right)$ .

**Remark 8.7.** The limiting rate  $\frac{\alpha}{2m}$  for approximation in  $L^2(U, V, \sigma)$ , which we obtain here using optimized  $\nu$ -dependent spatial discretizations, can also be achieved by a simpler construction using a single  $\nu$ -independent spatial discretization space. This can be seen as follows: Under

our present assumptions,  $\sup_{y \in U} \|a(y)\|_{C^\beta} < \infty$  for any positive  $\beta < \alpha$ , and consequently  $\sup_{y \in U} \|u(y)\|_{H^{1+\beta}} < \infty$ . Hence for the Jacobi coefficients, we have

$$\sum_{\nu \in \mathcal{F}} \|v_\nu\|_{H^{1+\beta}}^2 = \int_U \|u(y)\|_{H^{1+\beta}}^2 d\sigma(y) < \infty.$$

For  $n \in \mathbb{N}$ , as in §3 let  $\Lambda_n \subset \mathcal{F}$  be the subset of the  $n$  largest  $\|v_\nu\|_V$ . Now choose the vector  $\mathbf{n}$  as in (3.10) as  $n_\nu = \hat{n}$  with some fixed  $\hat{n}$  for all  $\nu \in \Lambda_n$ . With an appropriate sequence of spatial approximation spaces  $(V_n)_{n>0}$ , we then obtain

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq \left\| \sum_{\nu \in \mathcal{F}} (v_\nu - v_{\nu, \hat{n}}) J_\nu \right\|_{\mathcal{V}} + \left\| \sum_{\nu \notin \Lambda_n} v_{\nu, \hat{n}} J_\nu \right\|_{\mathcal{V}} \leq C \hat{n}^{-t} \left( \sum_{\nu \in \mathcal{F}} \|v_\nu\|_{H^{1+\beta}}^2 \right)^{\frac{1}{2}} + C n^{-s},$$

with  $t = \frac{\beta}{m} < \frac{\alpha}{m}$  and any  $s < \frac{\alpha}{m}$ . In this case, the total number of coefficients is simply  $N = n\hat{n}$ , which means that with appropriately chosen  $n, \hat{n}$  we obtain

$$\|u - u_{\mathbf{n}}\|_{\mathcal{V}} \leq CN^{-\frac{\alpha}{2m} + \delta} \quad (8.15)$$

for any  $\delta > 0$ , and hence the same asymptotic convergence as in Corollaries 8.3 and 8.5. In other words, in the present setting, individual optimization of discretization spaces for each Jacobi coefficient does not improve the achievable convergence rate compared to a single spatial discretization for all coefficients.

In summary, under Assumptions 8.1, provided that spatial approximation by linear methods of sufficiently high order is used (e.g., based on a predefined hierarchy of finite element meshes), the limiting spatial-parametric rate is *half* the limiting convergence rate of the parametric expansion.

## 8.2 Space-parameter adaptivity

We next consider the convergence rates resulting from a combination of Theorem 3.1 with Theorem 7.1 and Corollary 7.3, under the particular Assumptions 8.1 on a wavelet-type expansion of the coefficient  $a$ . To simplify the exposition, we focus here on Jacobi expansions in  $L^2(U, V, \sigma)$  with spatial dimensions  $m = 2, 3$ .

**Corollary 8.8.** *Let Assumptions 8.1 hold, and let  $\tau > 1$  if  $m = 2$  and  $\tau \geq \frac{6}{5}$  if  $m = 3$ . Assume that the approximation property (7.1) holds for a nonlinear family  $(V_n)_{n \geq 1}$  such as (7.2), with  $X = B^\tau$  and  $t = \frac{1}{m}$  when  $\alpha > 1$ , or otherwise that the approximation property (7.23) holds with  $Z = [V, B^\tau]_\theta$  or  $[V, B^\tau]_{\theta, r}$  for  $0 < \theta < \alpha < 1$  and  $0 < r \leq \infty$ . Then, for any  $\delta > 0$  and  $\alpha > 0$ , there exists a constant  $C > 0$  such that the following holds: for each  $n$  there exists  $\mathbf{n} = (n_\nu)_{\nu \in \Lambda_n}$ , such that, when  $0 < \alpha < 1$ ,*

$$\|u - u_{\mathbf{n}}\|_{L^2(U, V, \sigma)} \leq CN^{-\frac{\alpha}{m} + \delta}, \quad (8.16)$$

and when  $\alpha > 1$ ,

$$\|u - u_{\mathbf{n}}\|_{L^2(U, V, \sigma)} \leq CN^{-\frac{1}{m}}, \quad (8.17)$$

where  $N := \sum_{\nu \in \Lambda_n} n_\nu$ .

**Proof.** In the case  $\alpha > 1$ , we combine Theorem 7.1 with the bound (7.20) to obtain  $(\|t_\nu\|_{B^\tau})_{\nu \in \mathcal{F}} \in \ell^p$  for any  $p$  such that  $\frac{1}{p} < \frac{\alpha-1}{m} + \frac{1}{\tau}$ . By Proposition 8.2, we have convergence of the parametric



expansion with any  $s < \frac{\alpha}{m}$ . We now again apply Theorem 3.1. As  $s$ ,  $p$ , and  $\tau$  approach their limiting values,

$$\frac{st}{s + t + \frac{1}{2} - p^{-1}} \rightarrow \frac{\alpha}{m}, \quad (8.18)$$

resulting in the convergence rate  $\min\{\frac{\alpha}{m}, t\} = \frac{1}{m}$ .

In the case  $0 < \alpha < 1$ , we apply Corollary 7.3 with  $0 < \theta < \alpha$ , where we may choose  $\rho_j = 2^{\beta|\lambda(j)|}$  with  $\beta < \alpha - 1$  and  $\hat{\rho}_j = 1 + c2^{\hat{\beta}|\lambda(j)|}$  with  $\beta + 1 < \hat{\beta} < \alpha$  and sufficiently small  $c > 0$ . By taking  $\beta, \hat{\beta}$  sufficiently close to their limiting values, we obtain  $(\tilde{\rho}_j^{-1})_{j \geq 1} \in \ell^q(\mathbb{N})$  for  $\frac{1}{q} < \frac{\alpha - \theta}{m}$ . By (7.20), with  $\zeta$  as in (7.22), we have  $(\|t_\nu\|_Z)_{\nu \in \mathcal{F}} \in \ell^p(\mathcal{F})$  for any  $p$  such that

$$\frac{1}{p} < \frac{\alpha - \theta}{m} + \frac{1}{\zeta} = \frac{\alpha}{m} + \frac{1}{2} + \left(\frac{1}{\tau} - \frac{1}{2} - \frac{1}{m}\right)\theta.$$

By Proposition 7.4, under our conditions on  $\tau$ , we have  $t = \frac{\theta}{m}$  for elements of  $Z = [V, B^\tau]_\theta$  or  $[V, B^\tau]_{\theta, r}$ . As  $s$ ,  $p$ ,  $\tau$  approach their respective limiting values we again obtain (8.18), independently of  $\theta$ . For the resulting limiting convergence rate, we thus obtain  $\min\{\frac{\alpha}{m}, \frac{\theta}{m}\} \rightarrow \frac{\alpha}{m}$  as  $\theta \rightarrow \alpha$ .  $\square$

**Remark 8.9.** Following the same lines for  $m = 1$ , we instead obtain the rate  $\min\{\frac{2\alpha}{3} - \delta, 1\}$ . Improving this to  $\min\{\alpha - \delta, 1\}$  analogously to our results for  $m > 1$  would require summability results for appropriate Besov norms with  $\tau = \frac{2}{3}$ .

In summary, in contrast to the case of linear approximation in the spatial variable, for  $0 < \alpha < 1$  we arrive at the same rate as for the separate approximations in spatial or parametric variables alone. For  $\alpha > 1$ , the convergence rates obtained here are limited to  $\frac{1}{m}$ , since the results obtained in §7 do not cover the  $\ell^p$  summability of the sequence  $(\|u_\nu\|_X)_{\nu \in \mathcal{F}}$  for the higher-order smoothness classes  $X = W^{s, \tau}(D)$  or  $B_q^{s, \tau}(D)$  when  $s > 2$ .

**Remark 8.10.** In [2], it is shown for Legendre expansions that when a suitable wavelet basis is used for the spatial discretization, under Assumption 8.1 with sufficiently regular  $\psi_j$  one can achieve  $s^*$ -compressibility of the resulting representation of the parametric diffusion operator on  $L^2(U, V, \sigma)$  with  $s^*$  arbitrarily close to  $\frac{\alpha}{m}$ . This implies that approximations  $u_{\mathbf{n}}$  as in Corollary 8.8 can be computed by standard adaptive methods with the number of operations scaling almost linearly in the total number of degrees of freedom  $N$  for any  $\alpha > 0$ .

## References

- [1] James H. Adler and V. Nistor, *Graded mesh approximation in weighted Sobolev spaces and elliptic equations in 2d*, Mathematics of Computation **84**, Number 295, 2191–2220 (2015).
- [2] M. Bachmayr, A. Cohen, W. Dahmen, *Parametric PDEs: Sparse or low-rank approximations?*, arXiv:1607.04444, 2016.
- [3] M. Bachmayr, A. Cohen, G. Migliorati, *Sparse polynomial approximation of parametric elliptic PDEs. Part I: affine coefficients*, ESAIM: M2AN **51**, 321–339, 2017.
- [4] M. Bachmayr, A. Cohen, R. DeVore, G. Migliorati, *Sparse polynomial approximation of parametric elliptic PDEs. Part II: lognormal coefficients*, ESAIM: M2AN **51**, 341–363, 2017.

- [5] C. Băcuță, V. Nistor and L.T. Zikatanov, *Improving the rate of convergence of ‘high order finite elements’ on polygons and domains with cusps*, Numerische Mathematik **100** (2) (2005) 165-184.
- [6] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, *On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods*, Mathematical Models and Methods in Applied Sciences, **22**, 1-33, 2012.
- [7] J. Beck, F. Nobile, L. Tamellini, and R. Tempone, *Convergence of quasi-optimal stochastic Galerkin methods for a class of PDEs with random coefficients*, Computers and Mathematics with Applications, **67**, 732-751, 2014.
- [8] J. Bergh, J. Löfström, *Interpolation Spaces, An Introduction*, Springer, 1976.
- [9] A. Chkifa, A. Cohen and C. Schwab, *Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs*, Journal de Mathématiques Pures et Appliqués **103**(2) 400-428, 2015.
- [10] P. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland Publ., 1978.
- [11] A. Cohen, *Numerical analysis of wavelet methods*, Studies in mathematics and its applications, Elsevier, Amsterdam, 2003.
- [12] A. Cohen and R. DeVore, *Approximation of high-dimensional parametric PDEs*, Acta Numerica 24, 1-159, 2015.
- [13] A. Cohen, R. DeVore and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic PDEs*, Anal. and Appl. 9, 11-47, 2011.
- [14] R. DeVore, *Nonlinear Approximation*, Acta Numerica 7, 51-150, 1998
- [15] Dinh Dũng, *Linear collective collocation and Galerkin approximations for parametric and stochastic elliptic PDEs*, arXiv:1511.03377, 2015, v4 23Jun2016.
- [16] F.D. Gaspoz and P. Morin, *Convergence rates for adaptive finite elements* IMA Journal of Numerical Analysis **29**(4) (2009), 917-936.
- [17] P. Grisvard, *Elliptic problems in nonsmooth domains*. Monographs and Studies in Mathematics, 24. Pitman (Advanced Publishing Program), Boston, MA, 1985.
- [18] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, 2nd edn., Dover, 2007.
- [19] R. G. Ghanem and P. D. Spanos, *Spectral techniques for stochastic finite elements*, Archive of Computational Methods in Engineering, **4**, 63-100, 1997.
- [20] C. J. Gittelsohn, *An adaptive stochastic Galerkin method for random elliptic operators*, Mathematics of Computation, **82**, 1515-1541, 2013.
- [21] V.H. Hoang and C. Schwab, *N-term Galerkin Wiener chaos approximation rates for elliptic PDEs with lognormal Gaussian random inputs*, M3AS **24** (2014) 797–826.
- [22] O. Knio and O. Le Maitre, *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*, Springer, 2010.
- [23] D. Xiu, *Numerical methods for stochastic computations: a spectral method approach*, Princeton University Press, 2010.