Eidgenössische Technische Hochschule Zürich

Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs

A. Cohen*, R. DeVore† and C. Schwab

Research Report No. 2010-03
February 2010

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

*UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France; CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France.

†Department of Mathematics, Texas A& M University, College Station, TX 77843, USA.

Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs *

Albert Cohen, Ronald DeVore and Christoph Schwab

February 11, 2010

Abstract

Parametric partial differential equations are commonly used to model physical systems. They also arise when Wiener chaos expansions are used as an alternative to Monte Carlo when solving stochastic elliptic problems. This paper considers a model class of second order, linear, parametric, elliptic PDEs in a bounded domain D with coefficients depending on possibly countably many parameters. It shows that the dependence of the solution on the parameters in the diffusion coefficient is analytically smooth. This analyticity is then exploited to prove that under very weak assumptions on the diffusion coefficients, the entire family of solutions to such equations can be simultaneously approximated by multivariate polynomials (in the parameters) with coefficients taking values in the Hilbert space $V = H^1_0(D)$ of weak solutions of the elliptic problem with a controlled number of terms N. The convergence rate in terms of N does not depend on the number of parameters in V which may be countable, therefore breaking the curse of dimensionality. The discretization of the coefficients from a family of continuous, piecewise linear Finite Element functions in D is shown to yield finite dimensional approximations whose convergence rate in terms of the overall number N_{dof} of degrees of freedom is the minimum of the convergence rates afforded by the best N-term sequence approximations in the parameter space and the rate of Finite Element approximations in D for a single instance of the parametric problem.

1 Introduction

1.1 A class of parametric PDE's

This paper is concerned with simultaneously solving a family of elliptic equations on a bounded Lipschitz domain $D \subset \mathbb{R}^d$ of the form

$$-\nabla \cdot (a \nabla u) = f \quad \text{in} \quad D, \qquad u|_{\partial D} = 0, \tag{1.1}$$

where the diffusion coefficients $a(x, y)$ are functions of $x = (x_1, \cdots, x_d) \in D$ and of parameters $y = (y_1, y_2, \cdots)$ which may be finite or infinite in number. The right hand side f is a fixed function on D, and the gradient operator ∇ is taken with respect to x. The solution $u(x, y)$ also depends on these variables. Parametric problems of this type arise in modeling complex systems in various contexts:

*This research was supported by the Office of Naval Research Contracts ONR-N00014-08-1-1113, ONR N00014-09-1-0107, the AFOSR Contract FA95500910500, the ARO/DoD Contracts W911NF-05-1-0227 and W911NF-07-1-0185, the excellency chair of the Foundation "Science Mathématiques de Paris" awarded to Ronald DeVore in 2009. This publication is based on work supported by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST). Research supported in part by the Swiss National Science Foundation under Grant No. 200021-120290/1 and by the European Research Council under grant 247277

- Stochastic modelling: the parameters y are realizations of random variables which account for the fact that the diffusion coefficient is not known exactly and is therefore modelled as a random field. The user is interested in the resulting statistical properties of the solution u, such as the mathematical expectation $\bar{u}(x) := E(u(x))$ or the solution's two-point correlation $C_u(x, x') := E((u(x)-\bar{u}(x))(u(x')-\bar{u}(x')))$. This is the point of view adopted for example in [14, 16, 4, 13, 3, 18, 19].

- Deterministic modelling: the parameters y are known or controlled by the user, who is interested in studying the dependence of u with respect to these parameters for various purposes (for example, optimizing an output of the equation with respect to y). This is the point of view adopted for example in [7, 17].

A major idea for solving such families of equations is to exploit the fact that the solution depends smoothly on the coefficient a. Using this smooth dependence, the hope is to show that one can simultaneously solve the entire parametric family of equations to an accuracy ϵ, with reasonable computation cost regardless of the number of involved parameters, therefore not exhibiting the curse of dimensionality. The main goal of this paper is to give theoretical results which support this viewpoint. Namely, we shall prove that under very minimal assumptions on $a(x, y)$, the solution $u(x, y)$ has a power series representation in y with suitably decaying coefficients. We shall give a quantitative theory which describes how many terms of such an expansion are necessary to capture the solution to a given tolerance ϵ.

In order to formulate our results, we begin by recalling a few standard results for elliptic equations of the form (1.1). To begin our discussion, let us first consider the case where we want to solve only one equation, i.e. $a(x, z) = \alpha(x)$ depends only on $x \in D$. Central to elliptic theory is the Sobolev space $V := H_0^1(D)$, called the *energy space*, which is the set of all functions v whose trace vanishes on the boundary of D and whose *energy norm* $\|v\|_V := \|\nabla v\|_{L^2(D)}$ is finite. The dual of V is denoted by $V^* = H^{-1}(D)$. The solution of (1.1) is defined for any $f \in V^*$ in weak form as a function $u \in V$ which satisfies

$$\int_D \alpha(x)\nabla u(x) \cdot \nabla v(x)dx = \int_D f(x)v(x)dx. \text{ for all } v \in V, \tag{1.2}$$

Under the assumption that

$$0 < r \leq \alpha(x) \leq R < \infty, \quad x \in D, \tag{1.3}$$

the Lax-Milgram Lemma ensures the existence and uniqueness of the solution u of (1.2) in V. Moreover, this solution satisfies the a-priori estimate

$$\|u\|_V \leq \frac{\|f\|_{V^*}}{r}. \tag{1.4}$$

Of key importance to further development is the fact that Lax-Milgram theory can be extended to the case where the coefficient function α is complex valued. In this case, the ellipticity assumption (1.3) should be replaced by

$$0 < r \leq \Re(\alpha(x)) \leq |\alpha(x)| \leq R < \infty, \quad x \in D. \tag{1.5}$$

and all the above results remain valid with the usual extension of Sobolev spaces to complex valued functions.

With these facts in hand, let us return to our main interest, which is to solve the family of elliptic equations (1.1). Rather than striving for atmost generality, we consider *affine dependence* of a with respect to y, which means that the parameters y_j are the coefficients of the function a in some formal series expansion

$$a(x, y) = \bar{a}(x) + \sum_{j=1}^{\infty} y_j \psi_j(x), \quad x \in D, \tag{1.6}$$

where $\bar{a} \in L^\infty(D)$ and $\{\psi_j\}_{j\geq 1} \subset L^\infty(D)$. The sequence $\{\psi_j\}_{j\geq 1}$ could either be given to us by the physical system we are modeling or its choice could be at our discretion. For example, a typical choice for the $\{\psi_j\}$ in the stochastic context are the elements of the Karhúnen-Loève basis, which means that $\bar{a}$ is the average of a and that the y_j are pairwise decorrelated random variables, but we may as well choose other bases such as wavelets. In certain models where the parameters are finitely many, the sequence $\{\psi_j\}_{j\geq 1}$ may not be a complete basis of $L^2(D)$. For example in the case where a is piecewise constant with respect to a partition of the domain $D = D_1 \cup \cdots \cup D_K$ into measurable sets, it is then natural to represent it as

$$a(x,y) = \bar{a}(x) + \sum_{j=1}^{K} y_j \psi_j(x), \quad x \in D, \tag{1.7}$$

where $\bar{a}(x)$ is itself piecewise constant on this partition and $\psi_j := \chi_{D_j}$.

We are interested in simultaneously approximating the solutions $u(y)$ to the family of elliptic equations with the above input parameters. In the decomposition (1.6), we have the choice to either normalize the basis (e.g., assume they all have norm one in some space) or to normalize the parameters. It is more convenient for us to do the latter. This leads us to the following assumptions which shall be made throughout:

i) For all $j \in \mathbb{N} : \psi_j \in L^\infty(D)$ and $\psi_j(x)$ is defined for all $x \in D$,

ii) the $y = (y_1, y_2, \ldots)$ to be considered are all in the set $U = [-1,1]^{\mathbb{N}}$, i.e. the unit ball of the sequence space $\ell^\infty(\mathbb{N})$ (with $\mathbb{N}$ replaced by $\{1, \ldots, K\}$ in the case (1.7) that the number of parameters is finite),

iii) for each $a(x,y)$ to be considered, we have for every $x \in D$ and every $y \in U$

$$a(x,y) = \bar{a}(x) + \sum_{j\geq 1} y_j \psi_j(x). \tag{1.8}$$

Under these assumptions, we consider the map $y \mapsto u(y)$ from U to V, where $u(y)$ is the solution of (1.1) with coefficient given by (1.8). We shall work under the assumption that the ellipticity condition (1.3) holds uniformly for $y \in U$.

Uniform Ellipticity Assumption: *there exist* $0 < r \leq R < \infty$ *such that for all* $x \in D$ *and for all* $y \in U$

$$0 < r \leq a(x,y) \leq R < \infty. \tag{1.9}$$

We refer to assumption (1.9) as $\mathbf{UEA}(r,R)$ in the following. In particular, $\mathbf{UEA}(r,R)$ implies $r \leq \bar{a}(x) \leq R$ for all $x \in D$, since we can choose $y_j = 0$ for all $j \in \mathbb{N}$. Also observe that the validity of the lower and upper inequality in (1.9) for all $y \in U$ are respectively equivalent to the conditions that

$$\sum_{j\geq 1} |\psi_j(x)| \leq \bar{a}(x) - r, \quad x \in D, \tag{1.10}$$

and

$$\sum_{j\geq 1} |\psi_j(x)| \leq R - \bar{a}(x), \quad x \in D. \tag{1.11}$$

1.2 Previous results

In the paper [9], we have established several results concerning the smoothness and approximation of the function $y \mapsto u(y)$ by multivariate polynomial in y with coefficients in V. To describe these, we introduce the following standard multivariate notation. We denote the countable set of "finitely supported" sequences of nonnegative integers by

$$\mathcal{F} := \{\nu = (\nu_1, \nu_2, \ldots) : \nu_j \in \mathbb{N}, \text{and } \nu_j \neq 0 \text{ for only a finite number of } j\}. \tag{1.12}$$

So

$$|\nu| := \sum_{j \geq 1} |\nu_j| \tag{1.13}$$

is finite if and only if $\nu \in \mathcal{F}$. For $\nu \in \mathcal{F}$ supported in $\{1, \cdots, J\}$, we define the partial derivative

$$\partial^\nu u = \frac{\partial^{|\nu|} u}{\partial^{\nu_1} y_1 \cdots \partial^{\nu_J} y_J},$$

and the multi-factorial

$$\nu! := \prod_{j \geq 1} \nu_j! \text{ where } 0! := 0.$$

If $\alpha = (\alpha_j)_{j \geq 1}$ is a sequence of complex numbers, we define for all $\nu \in \mathcal{F}$

$$\alpha^\nu := \prod_{j \geq 1} \alpha_j^{\nu_j},$$

where throughout the paper we use the convention $0^0 := 1$. We are interested in the convergence towards $u(y)$ of the power series

$$\sum_{\nu \in \mathcal{F}} t_\nu y^\nu, \qquad y \in U, \tag{1.14}$$

where the Taylor coefficients $t_\nu \in V$ are defined as

$$t_\nu := \frac{1}{\nu!} \partial^\nu u(0), \qquad \nu \in \mathcal{F}.$$

We define the sequence

$$b := (b_j)_{j \geq 1}, \qquad b_j := \frac{\|\psi_j\|_{L^\infty(D)}}{\bar{a}_{\min}}, \tag{1.15}$$

where $\bar{a}_{\min} := \inf_{x \in D} \bar{a}(x)$. A first set of results of [9] shows that under $\mathbf{UEA}(r, R)$, all partial derivatives of u with respect to y are well defined for all $y \in U$, and satisfy the estimate

$$\|\partial^\nu u(y)\|_V \leq \frac{\|f\|_{V^*}}{r} |\nu|! \prod_{j \geq 1} \Big(\frac{\|\psi_j\|_{L^\infty(D)}}{a_{\min}(y)} \Big)^{\nu_j}, \tag{1.16}$$

where $a_{\min}(y) := \inf_{x \in D} a(x, y)$. Since $a_{\min}(0) = \bar{a}_{\min}$, this estimate implies in particular

$$\|t_\nu\|_V \leq \frac{\|f\|_{V^*}}{r} \frac{|\nu|!}{\nu!} b^\nu \tag{1.17}$$

A second set of results deals with the summability properties of sequences of the type $\Big(\frac{|\nu|!}{\nu!} b^\nu \Big)_{\nu \in \mathcal{F}}$ appearing in the above estimates. These summability properties are closely linked with those of the sequence b as expressed by the following result of [9].

Theorem 1.1 *For $0 < p < 1$, $\left(\frac{|\nu|!}{\nu!}b^\nu\right)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$ if and only if* (i) $\sum_{j\geq 1} b_j < 1$, *and* (ii) $(b_j) \in \ell^p(\mathbb{N})$.

Combining this result with the estimate (1.17), we find as an immediate corollary that if the sequence b is such that $\sum_{j\geq 1} b_j < 1$ and $(b_j) \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ belongs to $\ell^p(\mathcal{F})$. In other words, under the condition that $\sum_{j\geq 1} b_j < 1$, the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ *inherits* the summability (or sparsity) of the sequence $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1}$ (since $(b_j) \in \ell^p(\mathbb{N})$ if and only if $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$). In particular, since it is assumed $p < 1$ we have $(\|t_\nu\|)_{\nu\in\mathcal{F}} \in \ell^1(\mathcal{F})$ which implies that the power series (1.14) is summable in V uniformly in $y \in U$ and that the sum of this series is indeed $u(y)$.

Let us make some remarks about these results which will help to better understand the motivation of the present paper. First of all, the good news in this theorem is that it shows that $u(x, y)$ can be uniformly recovered for all $y \in U$ by retaining a controlled number of terms of the expansion (1.14). Namely, $(\|t_\nu\|_V) \in \ell^p(\mathcal{F})$ implies that for each $N > 0$, there is a finite set $\Lambda_N \subset \mathcal{F}$ with $\#(\Lambda_N) = N$ (corresponding to indices of the largest $\|t_\nu\|_V$) that satisfies

$$\sup_{y\in U} \|u(y) - \sum_{\nu\in\Lambda_N} t_\nu y^\nu\|_V \leq \sum_{\nu\notin\Lambda_N} \|t_\nu\|_V \leq \|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})} N^{-s}, \quad s := \frac{1}{p} - 1. \tag{1.18}$$

The proof of the second inequality, originally due to Stechkin, is recalled in Section 3.3 of this paper. In other words, all of the solutions to the family of parametric problems can be simultaneously approximated by a polynomial in y with coefficients in V with a control on the number of terms in the polynomial expansion. We have obtained an algebraic convergence rate even when the number of involved parameters is infinite, showing that our approximation is not prone to the curse of dimensionality.

Another important point is that the assumption $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$ is quite reasonable. It would be implied by mild regularity assumptions on the coefficients $a(x, y)$ with respect to the x variable. Consider, for example, univariate problems on $D =]-1, 1[$ and the case of a Fourier expansion,

$$a(x, y) = \bar{a}(x) + \sum_{k\geq 0} y_{2k+1}\alpha_k \cos(2\pi kx) + \sum_{k\geq 1} y_{2k}\beta_k \sin(2\pi kx)$$

where $(\alpha_k)_{k\geq 0}$ and $(\beta_k)_{k\geq 1}$ are certain normalizing sequences. It is known that if the function $a(\cdot, y) - \bar{a}$ is in $\mathrm{Lip}(s, L^1)$ for some $s > 1$, then its Fourier coefficients satisfy the decay estimate

$$|y_{2k+1}\alpha_k| + |y_{2k}\beta_k| \leq C|k|^{-s}, \qquad k \geq 1, \quad y \in U$$

with C depending on the $\mathrm{Lip}(s, L^1)$-norm of $a(\cdot, y) - \bar{a}$. Assuming that this norm is bounded independently of y which is arbitrary in U, this is equivalent to the decay estimate

$$|\alpha_k| + |\beta_k| \leq C|k|^{-s}, \qquad k \geq 1,$$

and therefore

$$\|\psi_j\|_{L^\infty(D)} \leq Cj^{-s}, \qquad j \geq 1, \tag{1.19}$$

after suitably reindexing the Fourier basis elements. Therefore, the $\ell^p(\mathbb{N})$ summability of the sequence $(b_j)_{j\geq 1}$ is ensured when $s > \frac{1}{p}$. Results like this persist for higher space dimension d, general domains and other bases such Karhúnen-Loève (e.g. [21, 20]) and wavelet expansions.

On the negative side, the assumption that $\sum_{j\geq 1} b_j < 1$, which is equivalent to $\sum_{j\geq 1} \|\psi_j\|_{L^\infty(D)} < \bar{a}_{\min}$ is quite strong. Indeed, it implies $\mathbf{UEA}(r, R)$ with $r := \bar{a}_{\min} - \sum_{j\geq 1} \|\psi_j\|_{L^\infty(D)} > 0$ and $R = \bar{a}_{\max} + \sum_{j\geq 1} \|\psi_j\|_{L^\infty(D)} < +\infty$, but it could be quite stronger than $\mathbf{UEA}(r, R)$. This is in particular the case when the supports of the ψ_j have some disjointness, such as in the wavelet case where only a few of the wavelets overlap at a given scale, or in the case of characteristic functions of disjoint sets.

1.3 Objective

This brings to the forefront the question of whether the property that the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ belongs to $\ell^p(\mathcal{F})$ might hold under the weaker and more natural $\mathbf{UEA}(r,R)$. Closely related is the question of approximating $u(y)$ in (1.14) by partial sums: for finite sets $\Lambda\subset\mathcal{F}$ and $y\in U$, we define

$$S_\Lambda u(y) := \sum_{\nu\in\Lambda} t_\nu y^\nu \in V. \tag{1.20}$$

We say that *a sequence $(\Lambda_N)_{N\geq 1}\subset\mathcal{F}$ of finite sets exhausts $\mathcal{F}$* if any finite $\Lambda\subset\mathcal{F}$ is contained in all Λ_N for $N\geq N_0$ with N_0 sufficiently large. One purpose of the present paper is to prove the following theorem.

Theorem 1.2 *If $a(x,y)$ satisfies $\mathbf{UEA}(r,R)$ and if $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1}\in\ell^p(\mathbb{N})$ for some $p<1$, then*

$$u(y)=\sum_{\nu\in\mathcal{F}} t_\nu y^\nu,\quad y\in U, \tag{1.21}$$

where the functions $t_\nu\in V$ are as in (1.14) and $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}\in\ell^p(\mathcal{F})$ for the same value of p. The convergence in (1.21) *is to be understood in the following unconditional sense: if $(\Lambda_N)_{N\geq 1}\subset\mathcal{F}$ is any sequence of finite sets which exhausts $\mathcal{F}$, then the partial sums $S_{\Lambda_N}u(y)=\sum_{\nu\in\Lambda_N}t_\nu y^\nu$ satisfy*

$$\lim_{N\to+\infty}\sup_{y\in U}\|u(y)-S_{\Lambda_N}u(y)\|_V=0. \tag{1.22}$$

If Λ_N is the set of $\nu\in\mathcal{F}$ corresponding to indices of the largest $\|t_\nu\|_V$, we have in addition the convergence rate estimate

$$\sup_{y\in U}\|u(y)-S_{\Lambda_N}u(y)\|_V\leq\|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})}N^{-s},\quad s:=\frac{1}{p}-1. \tag{1.23}$$

Similar to the results in [9], this theorem reveals that the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ inherits the summability properties of the sequence $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1}$, however now under the weaker assumption $\mathbf{UEA}(r,R)$. The proof of Theorem 1.2 requires much finer estimates on the $\|t_\nu\|_V$ than (1.17) which was used in [9]. Our key ingredients for getting such estimates rely on complex analysis. More precisely, we extend the definition of $u(y)$ to $u(z)$ for the complex variable $z=(z_j)_{j\geq 1}$ (by using the z_j instead of y_j in the definition of a by (1.8)) where each z_j has modulus less than 1. Therefore z belongs to the polydisc

$$\mathcal{U}:=\otimes_{j\geq 1}\{z_j\in\mathbb{C}:|z_j|\leq 1\}. \tag{1.24}$$

Using (1.10) and (1.11), it is readily seen that when the functions $\bar a$ and ψ_j are real valued, then $\mathbf{UEA}(r,R)$ implies that for all $x\in D$ and $z\in\mathcal{U}$,

$$0<r\leq\Re(a(x,z))\leq|a(x,z)|\leq 2R, \tag{1.25}$$

and therefore the corresponding solution $u(z)$ is well defined in V for all $z\in\mathcal{U}$ according to the complex valued version of Lax-Milgram theorem. More generally, we may as well consider an expansion of the form,

$$a(x,z)=\bar a+\sum_{j\geq 1}z_j\psi_j$$

where $\bar a$ and ψ_j are complex valued function and replace $\mathbf{UEA}(r,R)$ by its complex valued counterpart.

Uniform Ellipticity Assumption in $\mathbb{C}$: *there exist $0<r\leq R<\infty$ such that for all $x\in D$ and all $z\in\mathcal{U}$*

$$0<r\leq\Re(a(x,z))\leq|a(x,z)|\leq R<\infty. \tag{1.26}$$

We refer to (1.26) as $\mathbf{UEAC}(r,R)$. We will prove the following complex valued version of Theorem 1.2, which clearly implies it as a particular case.

Theorem 1.3 *If $a(x,z)$ satisfies* **UEAC**(r,R) *for some $0 < r \leq R < \infty$, and if $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then for all $z \in \mathcal{U}$*

$$u(z) = \sum_{\nu\in\mathcal{F}} t_\nu z^\nu \quad \text{in} \quad V, \tag{1.27}$$

where $t_\nu \in V$ and $(\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$ for the same value of p. The convergence in (1.27) *is to be understood in the following unconditional sense: if $(\Lambda_N)_{N\geq 1}$ is any sequence of finite sets which exhausts $\mathcal{F}$, the partial sums $S_{\Lambda_N} u(z) = \sum_{\nu\in\Lambda_N} t_\nu z^\nu$ satisfy*

$$\lim_{N\to+\infty} \sup_{z\in\mathcal{U}} \|u(z) - S_{\Lambda_N} u(z)\|_V = 0. \tag{1.28}$$

If in addition Λ_N is the set of $\nu \in \mathcal{F}$ corresponding to indices of the largest $\|t_\nu\|_V$, we have the convergence estimate

$$\sup_{z\in\mathcal{U}} \|u(z) - S_{\Lambda_N} u(z)\|_V \leq \|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})} N^{-s}, \quad s := \frac{1}{p} - 1. \tag{1.29}$$

1.4 Outline of the paper

The analyticity of u with respect to the variable z is discussed in §2, in which we use Cauchy's integral formula to derive a general estimate of the form

$$\|t_\nu\|_V \leq \frac{\|f\|_{V^*}}{\delta} \prod_{j\geq 1} \rho_j^{\nu_j},$$

where $\delta > 0$, and where $(\rho_j)_{j\geq 1}$ is any sequence of positive numbers such that

$$\sum_{j\geq 1} \rho_j |\psi_j(x)| \leq \Re(\bar{a}(x)) - \delta, \quad x \in D.$$

In order to prove the main Theorem 1.3, we introduce in §3 a particular choice of $\rho = (\rho_j)_{j\geq 1}$ that depends on ν and satisfies the above constraint with $\delta = r/2$, leading to a corresponding estimate for $\|t_\nu\|_V$. We then use this estimate in order to prove the ℓ^p summability of the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$.

There is also interest in the validity of expansions like (1.21) when the monomial basis z^ν is replaced by other polynomial bases. For example in the analysis of stochastic elliptic problems, certain polynomial bases (depending on the underlying probability measure describing the stochasticity) may lead to improved results. This is discussed in [9] where for example it is shown that a tensor product Legendre polynomial basis gives improved results if the underlying probability measure is Lebesgue measure and if we agree to measure distortion in a least squares sense. The improvement comes in the form that the assumptions on the summability of $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1}$ for a given convergence rate N^{-s} are weaker. Motivated by this, we show in §4 that a version of Theorem 1.3 holds with respect to tensor product Legendre expansions.

The approximation of u by its partial sum $S_{\Lambda_N} u(y)$ is described by the data of $N = \#(\Lambda_N)$ functions $t_\nu \in V$. In practical numerical computations, these functions are themselves approximated by space discretization in the x variable, for instance using the Finite Element method. We discuss in §5 the additional error resulting from such discretizations. For this purpose we introduce a smoothness space $W \subset V$ that governs the rate of convergence of Finite Element methods, and we study the ℓ^p summability of the W norms of the Taylor and Legendre coefficients. Based on this analysis, we introduce a specific choice of Λ_N and of Finite Element spaces V_ν that depend on $\nu \in \Lambda_N$, and we study the approximation of u by a truncated Taylor series $\sum_{\nu\in\Lambda_N} \tilde{t}_\nu y^\nu$ with coefficients $\tilde{t}_\nu$ picked from the V_ν spaces, as well as by truncated Legendre series with

coefficients picked in the V_ν spaces. We show that the convergence rate in terms of the total number of degrees of freedom $N_{dof} = \sum_{\nu\in\Lambda_N} \dim(V_\nu)$ is either the same as when approximating a *single instance* $u(y)$ in the V norm, or as when approximating u *without space discretization*.

In the present work, we only consider the model problem (1.1) and the situation where the coefficients depend on the parameters in an affine manner. We expect that our general approach based on analyticity can be useful for more general problems with smooth yet not necessarily affine parameter dependance.

Let us finally stress that the main objective of this paper is to establish approximation results, *not* to propose a specific algorithm for computing such approximations. Our results should therefore be considered as a benchmark for the convergence analysis of numerical methods for the approximation of parametric and stochastic PDE's in the x and y variables. The two most commonly used numerical methods are Galerkin projection [4, 22] and collocation [3, 18, 19]. In order to retrieve the same convergence rates which are proved in the present paper, such methods need to be developed within an adaptive framework, with the goal of selecting proper sets Λ_N and finite element spaces V_ν throughout the numerical computation. This will be the object of future work.

2 Analyticity of u and estimates of Taylor coefficients

2.1 Domains of holomorphy

Our main vehicle in proving Theorem 1.3 is to exploit the fact if **UEAC**(r, R) holds, then the map $z \mapsto u(z)$ is a V valued and bounded analytic function in certain domains which are larger than $\mathcal{U}$. For $0 < \delta \leq 2R < \infty$ we define the set

$$\mathcal{A}_\delta = \{z \in \mathbb{C}^{\mathbb{N}} : \delta \leq \Re(a(x,z)) \leq |a(x,z)| \leq 2R \quad \text{for every} \quad x \in D.\} \tag{2.1}$$

Clearly, if **UEAC**(r, R) holds and if $0 < \delta < r$, then $\mathcal{A}_\delta$ contains $\mathcal{U}$. According to the Lax-Milgram theorem, we also find that for all $z \in \mathcal{A}_\delta$, there exists a unique solution $u(z) \in V$ which satisfies

$$\|u(z)\|_V \leq \frac{\|f\|_{V^*}}{\delta}. \tag{2.2}$$

Our first interest is to establish holomorphy of the map $z \mapsto u(z)$ with respect to the countably many variables z_j. This fact stems from the observation that the function $u(z)$ is the solution to the operator equation $A(z)u(z) = f$, where the operator $A(z) \in \mathcal{L}(V, V^*)$ depends in an affine manner on each variable z_j. To make our presentation more self-contained we give a direct proof. We start from a stability result, which is also used further in this section.

Lemma 2.1 *If u and $\tilde{u}$ are solutions of* (1.2) *with the same right hand side f and with coefficients α and $\tilde{\alpha}$, respectively, and if these coefficients both satisfy the assumption* (1.5)*, then*

$$\|u - \tilde{u}\|_V \leq \frac{\|f\|_{V^*}}{r^2} \|\alpha - \tilde{\alpha}\|_{L^\infty(D)}. \tag{2.3}$$

Proof: Substracting the variational formulations (1.2) for u and $\tilde{u}$, we find that for all $v \in V$,

$$0 = \int_D \alpha \nabla u \cdot \nabla v - \int_D \tilde{\alpha} \nabla \tilde{u} \cdot \nabla v = \int_D \alpha(\nabla u - \nabla \tilde{u}) \cdot \nabla v + \int_D (\alpha - \tilde{\alpha}) \nabla \tilde{u} \cdot \nabla v. \tag{2.4}$$

Therefore $w = u - \tilde{u}$ is the solution of $\int_D \alpha \nabla w \cdot \nabla v = L(v)$ where $L(v) := \int_D (\alpha - \tilde{\alpha}) \nabla \tilde{u} \cdot \nabla v$. Hence

$$\|w\|_V \leq \frac{\|L\|_{V^*}}{r},$$

and we obtain (2.3) since

$$\|L\|_{V^*} = \max_{\|v\|_V = 1} |L(v)| \leq \|\alpha - \tilde{\alpha}\|_{L^\infty(D)} \|\tilde{u}\|_V \leq \|\alpha - \tilde{\alpha}\|_{L^\infty(D)} \frac{\|f\|_{V^*}}{r}.$$

□

Lemma 2.2 *At any $z \in \mathcal{A}_\delta$, the function $z \mapsto u(z)$ admits a complex derivative $\partial_{z_j} u(z) \in V$ with respect to each variable z_j. This derivative is the weak solution of the problem: for $z \in \mathcal{A}_\delta$, find $\partial_{z_j} u(z) \in V$ such that*

$$\int_D a(x,z) \nabla \partial_{z_j} u(z) \cdot \nabla v = L_0(v) := -\int_D \psi_j \nabla u(z) \cdot \nabla v, \qquad \text{for all } v \in V. \tag{2.5}$$

Proof: We fix $j \geq 1$ and $z \in \mathcal{A}_\delta$. We denote by e_j the Kronecker sequence with 1 at index j and 0 at other indices. For $h \in \mathbb{C} \setminus \{0\}$ we consider the difference quotient

$$w_h(z) = \frac{u(z + he_j) - u(z)}{h}. \tag{2.6}$$

We notice that this quotient is well defined for h small enough: if $|h| \|\psi_j\|_{L^\infty(D)} \leq \frac{\delta}{2}$, we clearly have

$$\frac{\delta}{2} \leq \Re(a(x, z + he_j)) \leq |a(x, z + he_j)| \leq 2R + \frac{\delta}{2}, \quad x \in D,$$

and therefore $u(z + he_j)$ is well defined as an element of V. For such a small enough h, we have for all $v \in V$,

$$\begin{aligned} 0 &= \int_D a(x, z + he_j) \nabla u(z + he_j) \cdot \nabla v - \int_D a(x,z) \nabla u(z) \cdot \nabla v \\ &= h \int_D a(x,z) \nabla w_h(z) \cdot \nabla v + \int_D (a(x, z + he_j) - a(x,z)) \nabla u(z + he_j) \cdot \nabla v \\ &= h \int_D a(x,z) \nabla w_h(z) \cdot \nabla v + h \int_D \psi_j \nabla u(z + he_j) \cdot \nabla v \end{aligned}$$

and therefore w_h is the unique solution to

$$\int_D a(x,z) \nabla w_h(z) \cdot \nabla v = L_h(v), \quad \text{for all } v \in V,$$

where $L_h : v \to L_h(v) := -\int_D \psi_j \nabla u(z + he_j) \cdot \nabla v$ is a continuous, linear functional on V. The linear functional $L_h(\cdot)$ varies continuously in V^* with h as h tends to 0: indeed, we have for all $v \in V$,

$$|L_h(v) - L_0(v)| = |\int_D \psi_j (\nabla u(z + he_j) - \nabla u(z)) \cdot \nabla v| \leq \|\psi_j\|_{L^\infty(D)} \|u(z + he_j) - u(z)\|_V \|v\|_V,$$

and since the stability estimate (2.3) implies

$$\|u(z + he_j) - u(z)\|_V = \|\nabla u(z + he_j) - \nabla u(z)\|_{L^2(D)} \leq |h| \|\psi_j\|_{L^\infty(D)} \frac{4\|f\|_{V^*}}{\delta^2},$$

it follows that L_h converges towards L_0 in V^* as $h \to 0$. Therefore w_h converges in V towards w_0, which is the solution to

$$\int_D a(x,z)\nabla w_0(z) \cdot \nabla v = L_0(v), \quad \text{for all } v \in V.$$

Hence $\partial_{z_j} u(z) = w_0$ exists in V and is the unique solution of the variational problem (2.5). □

We further note that $\mathcal{A}_\delta$ also contains certain *polydiscs*. Let $\rho := (\rho_j)_{j\geq 1}$ be a sequence of positive numbers and define

$$\mathcal{U}_\rho = \otimes_{j\geq 1}\{z_j \in \mathbb{C} : |z_j| \leq \rho_j\} = \{z_j \in \mathbb{C} : z = (z_j)_{j\geq 1} \ ; \ |z_j| \leq \rho_j\}. \tag{2.7}$$

We say that a sequence $\rho = (\rho_j)_{j\geq 1}$ is *δ-admissible* if and only if for every $x \in D$

$$\sum_{j\geq 1} \rho_j |\psi_j(x)| \leq \Re(\bar{a}(x)) - \delta. \tag{2.8}$$

If the sequence ρ is δ-admissible, then the polydisc $\mathcal{U}_\rho$ is contained in $\mathcal{A}_\delta$. Indeed, we have on the one hand for all $z \in \mathcal{U}_\rho$ and for almost every $x \in D$

$$\Re(\bar{a}(x,z)) \geq \Re(\bar{a}(x)) - \sum_{j\geq 1} |z_j \psi_j(x)| \geq \Re(\bar{a}(x)) - \sum_{j\geq 1} \rho_j |\psi_j(x)| \geq \delta,$$

and on the other hand, for every $x \in D$

$$|a(x,z)| \leq |\bar{a}(x)| + \sum_{j\geq 1} |z_j \psi_j(x)| \leq |\bar{a}(x)| + \Re(\bar{a}(x)) - \delta \leq 2|\bar{a}(x)| \leq 2R,$$

where we have used the bound $|\bar{a}(x)| \leq R$ which is a trivial consequence of $\mathbf{UEAC}(r,R)$.

Similar to (1.10), we notice that the validity of the lower inequality in (1.26) for all $z \in \mathcal{U}$ is equivalent to the condition that

$$\sum_{j\geq 1} |\psi_j(x)| \leq \Re(\bar{a}(x)) - r, \quad x \in D. \tag{2.9}$$

This shows that the constant sequence $\rho_j = 1$ is δ-admissible for all $0 < \delta \leq r$, and that for $\delta < r$ there exist δ-admissible sequences such that $\rho_j > 1$ for all $j \geq 1$, i.e. such that the polydisc $\mathcal{U}_\rho$ is strictly larger than $\mathcal{U}$ in every variable.

2.2 Convergence of the polynomial expansion of $u(z)$

As a first consequence of these observations, we prove convergence of the series $\sum_{\nu\in\mathcal{F}} t_\nu z^\nu$ towards the function $u(z)$ *for a specific summation process*, under the condition that the series defining $a(x,z)$ converges in $L^\infty(D)$ uniformly over $z \in \mathcal{U}$.

Proposition 2.3 *If* $\mathbf{UEAC}(r,R)$ *holds for some* $0 < r \leq R < \infty$ *and if*

$$\sup_{z\in\mathcal{U}} \|a(\cdot,z) - \bar{a}(\cdot) - \sum_{1\leq j\leq J} z_j \psi_j(\cdot)\|_{L^\infty(D)} \to 0 \quad \text{as } J \to \infty \tag{2.10}$$

then there exists a sequence of $(\Lambda_N^*)_{N\geq 1}$ *of finite subsets which exhausts* $\mathcal{F}$ *and such that*

$$\lim_{N\to+\infty} \sup_{z\in\mathcal{U}} \|u(z) - S_{\Lambda_N^*} u(z)\|_V = 0.$$

Proof: For any $N > 0$, there exists $J = J(N)$ large enough such that

$$\sup_{z\in\mathcal{U}} \|a(\cdot,z) - \bar{a}(\cdot) - \sum_{1\le j\le J} z_j\psi_j(\cdot)\|_{L^\infty(D)} \le \frac{1}{2N}\frac{r^2}{\|f\|_{V^*}}. \tag{2.11}$$

For such a J and $z \in \mathcal{U}$, we define the set $E_J = \{1,\cdots,J\}$ and z_{E_J} which is obtained from z by putting to 0 all entries z_j for $j > J$ and leaving them unchanged for $j \le J$. Therefore (2.11) is equivalent to

$$\sup_{z\in\mathcal{U}} \|a(\cdot,z) - a(\cdot,z_{E_J})\|_{L^\infty(D)} \le \frac{1}{2N}\frac{r^2}{\|f\|_{V^*}}. \tag{2.12}$$

We then define

$$u_J(z_1,\cdots,z_J) = u(z_{E_J}) \tag{2.13}$$

Combining (2.12) with the stability estimate of Lemma 2.1, we find that for all $z \in \mathcal{U}$

$$\|u(z) - u_J(z_1,\cdots,z_J)\|_V \le \frac{\|f\|_{V^*}}{r^2}\frac{1}{2N}\frac{r^2}{\|f\|_{V^*}} = \frac{1}{2N}.$$

On the other hand, we have seen that the function u_J is holomorphic in an open neighbourhood of the J-dimensional polydisc $\otimes_{1\le j\le J}\{z_j \in \mathbb{C} : |z_j| \le 1\}$. From standard results on Banach space valued holomorphic functions (see Proposition 3.5 of [12] or Theorem 2.1.2 of [15]), this implies its analyticity and therefore the uniform summability of its Taylor series on this polydisc. Therefore there exists $K = K(N)$ large enough such that if we set

$$\Lambda_N^* := \{\nu \in \mathcal{F}\ ;\ |\nu| \le K \text{ and } \{j\ :\ \nu_j \ne 0\} \subset E_J\},$$

we have

$$\sup_{z\in\mathcal{U}} \|u_J(z_1,\cdots,z_J) - S_{\Lambda_N^*}u(z)\|_V \le \frac{1}{2N},$$

where we use the fact that $S_{\Lambda_N^*}u(z) = S_{\Lambda_N^*}u_J(z_1,\cdots,z_J)$ since $\Lambda_N^* \subset E_J$. Combining both estimates we obtain that

$$\sup_{z\in\mathcal{U}} \|u(z) - S_{\Lambda_N^*}u(z)\|_V \le \frac{1}{N},$$

which proves convergence. The exhaustion property of the Λ_N^* is ensured by assuming that $J(N)$ and $K(N)$ are strictly increasing, which is achievable without loss of generality. □

2.3 Estimates of the Taylor coefficients

In order to prove Theorem 1.3, we need estimates of the $\|t_\nu\|_V$. These are given by the following result.

Lemma 2.4 *If* **UEAC**(r,R) *holds for some* $0 < r \le R < \infty$ *and if* $\rho = (\rho_j)_{j\ge1}$ *is a* δ*-admissible sequence for some* $0 < \delta < r$*, then for any* $\nu \in \mathcal{F}$ *we have the estimate*

$$\|t_\nu\|_V \le \frac{\|f\|_{V^*}}{\delta}\prod_{j\ge1}\rho_j^{-\nu_j} = \frac{\|f\|_{V^*}}{\delta}\rho^{-\nu}, \tag{2.14}$$

where we use the convention that $t^{-0} = 1$ *for any* $t \ge 0$*.*

Proof: Let $\nu = (\nu_j)_{j\ge1} \in \mathcal{F}$ and let $J = \max\{j \in \mathbb{N} : \nu_j \ne 0\}$. Recalling the function u_J defined by (2.13), we therefore have

$$\partial^\nu u(0) = \frac{\partial^{|\nu|}u_J}{\partial z_1^{\nu_1}\cdots\partial z_J^{\nu_J}}(0,\cdots,0).$$

From the assumption that ρ is δ-admissible, we have that

$$\|u_J(z_1,\cdots,z_J)\|_V \leq \frac{\|f\|_{V^*}}{\delta}, \tag{2.15}$$

for all $(z_1,\cdots,z_J)$ in the J-dimensional polydisc

$$\mathcal{U}_{\rho,J} := \otimes_{1\leq j\leq J}\{z_j\in\mathbb{C} : |z_j|\leq \rho_j\}. \tag{2.16}$$

Introducing the sequence $\tilde{\rho}$ defined by

$$\tilde{\rho}_j = \rho_j + \varepsilon \text{ if } j\leq J,\ \ \tilde{\rho}_j = \rho_j \text{ if } j>J,\quad \varepsilon := \frac{\delta}{2\|\sum_{j\leq J}|\psi_j|\|_{L^\infty(D)}},$$

it is easily checked that $\tilde{\rho}$ is $\frac{\delta}{2}$-admissible and therefore $\mathcal{U}_{\tilde{\rho}}\subset\mathcal{A}_{\frac{\delta}{2}}$. We thus infer from Lemma 2.2 that for each $z\in\mathcal{U}_{\tilde{\rho}}$, u is holomorphic in each variable z_j.

It follows that u_J is holomorphic in each variable $z_1,\cdots,z_J$ on the polydisc $\otimes_{1\leq j\leq J}\{|z_j|<\tilde{\rho}_j\}$ which is an open neighbourhood of $\mathcal{U}_{\rho,J}$. We may thus apply the Cauchy formula (Theorem 2.1.2 of [15]) recursively in each variable z_j and write

$$u_J(\tilde{z}_1,\cdots,\tilde{z}_J) = (2\pi i)^{-J}\int\limits_{|z_1|=\rho_1}\cdots\int\limits_{|z_J|=\rho_J}\frac{u_J(z_1,\cdots,z_J)}{(\tilde{z}_1-z_1)\cdots(\tilde{z}_J-z_J)}dz_1\cdots dz_J.$$

By differentiation, this yields

$$\frac{\partial^{|\nu|}}{\partial z_1^{\nu_1}\cdots\partial z_J^{\nu_J}}u_J(0,\cdots,0) = \nu!(2\pi i)^{-J}\int\limits_{|z_1|=\rho_1}\cdots\int\limits_{|z_J|=\rho_J}\frac{u_J(z_1,\cdots,z_J)}{z_1^{\nu_1}\cdots z_J^{\nu_J}}dz_1\cdots dz_J,$$

and therefore, using (2.15), we obtain the estimate

$$\left\|\frac{\partial^{|\nu|}u_J}{\partial z_1^{\nu_1}\cdots\partial z_J^{\nu_J}}(0,\cdots,0)\right\|_V \leq \nu!\frac{\|f\|_{V^*}}{\delta}\prod_{j\leq J}\rho_j^{-\nu_j},$$

which is equivalent to (2.14). □

For a given $\nu\in\mathcal{F}$, we may search for the best bound provided by Lemma 2.4. This bound is given by

$$\|t_\nu\| \leq \inf_{0<\delta\leq r}\left\{\frac{\|f\|_{V^*}}{\delta}\inf\{\rho^{-\nu}\ ;\ \rho \text{ is } \delta\text{-admissible}\}\right\}.$$

The infimum is not easily computable. In particular, for a given δ such that $0<\delta<r$, the minimization problem

$$\inf\{\rho^{-\nu}\ ;\ \rho \text{ is } \delta\text{-admissible}\},$$

does not have a simple explicit solution due to the form of the constraint

$$\sum_{j\geq 1}\rho_j|\psi_j(x)| \leq \Re(\bar{a}(x)) - \delta,\ \ x\in D,$$

that generally couples all values of ρ_j. Only in the particular case (1.7) where the ψ_j have disjoint supports, one may decouple the search of the optimal ρ_j which is then given by

$$\rho_j^* = \min_{x\in D}\frac{\Re(\bar{a}(x))-\delta}{|\psi_j(x)|}.$$

Note that in this case, the optimal sequence ρ^* is independent of ν. In the general case where the supports of ψ_j overlap, the optimal sequence ρ^* should vary with ν.

Since this optimal sequence is not accessible to us, our strategy for proving Theorem 1.3 is to carefully design for each ν a certain sequence $\rho = \rho(\nu)$ which satisfies the constraint of δ-admissibility, and use the resulting estimate in order to prove that $(\|t_\nu\|_V)_{\nu\in\mathcal{F}} \in \ell^p(\mathcal{F})$.

3 Proof of Theorem 1.3

3.1 A choice of $\frac{r}{2}$-admissible sequences

From now on, we work with the particular choice $\delta = \frac{r}{2}$. Therefore, the estimate (2.14) of Lemma 2.4 takes the form

$$\|t_\nu\|_V \leq \frac{2\|f\|_{V^*}}{r}\prod_{j\geq 1}\rho_j^{-\nu_j} = \frac{2\|f\|_{V^*}}{r}\rho^{-\nu}, \tag{3.1}$$

for all $\frac{r}{2}$-admissible sequence ρ.

Given $\nu \in \mathcal{F}$, we have at our disposal the ability to choose the sequence ρ tailored to ν as long as it is $\frac{r}{2}$-admissible. To begin this choice, we first choose J_0 large enough such that

$$\sum_{j>J_0}\|\psi_j\|_{L^\infty(D)} \leq \frac{r}{12}, \tag{3.2}$$

Such a J_0 exists under the assumptions of Theorem 1.3 because $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p \subset \ell^1$. Note that we may always assume (and will do so in what follows) that up to some reindexing of the basis elements ψ_j the sequence $(\|\psi_j\|_{L^\infty})_{j\geq 1}$ is non-increasing. We first split $\mathbb{N}$ into the two sets $E := \{1 \leq j \leq J_0\}$ and $F := \mathbb{N}\backslash E$. Next we choose $\kappa > 1$ such that

$$(\kappa - 1)\sum_{j\leq J_0}\|\psi_j\|_{L^\infty(D)} \leq \frac{r}{4}. \tag{3.3}$$

For our given ν we shall use the sequence ρ defined by

$$\rho_j := \kappa, \quad j \in E; \quad \rho_j := \max\{1, \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}\}, \quad j \in F \tag{3.4}$$

where we use the notation ν_E for the restriction of ν to a set E and as before $|\cdot|$ denotes the ℓ^1 norm so that $|\nu_F| := \sum_{j>J_0}\nu_j$. We also make the convention that $\frac{\nu_j}{|\nu_F|} = 0$ when $|\nu_F| = 0$. Let us verify that the sequence ρ defined in (3.4) is $\frac{r}{2}$-admissible. To do so, we estimate for every $x \in D$

$$\begin{aligned}
\sum_{j\geq 1}\rho_j|\psi_j(x)| &\leq \kappa\sum_{j\leq J_0}|\psi_j(x)| + \sum_{j>J_0}\max\{1, \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}\}|\psi_j(x)| \\
&\leq (\kappa-1)\sum_{j\leq J_0}|\psi_j(x)| + \sum_{j\leq J_0}|\psi_j(x)| + \sum_{j>J_0}|\psi_j(x)| + \frac{r}{4} \\
&\leq \frac{r}{4} + \sum_{j\geq 1}|\psi_j(x)| + \frac{r}{4} \\
&\leq \Re(\bar{a}(x)) - \tfrac{r}{2},
\end{aligned}$$

where the last inequality uses (2.9) which follows from **UEAC**(r, R). The bound (3.1) is therefore valid for this particular sequence ρ, and it can be rewritten as

$$\|t_\nu\|_V \le \frac{2\|f\|_{V^*}}{r}\Big(\prod_{j\in E}\eta^{\nu_j}\Big)\Big(\prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j}\Big). \tag{3.5}$$

where $\eta := \frac{1}{\kappa} < 1$ and

$$d_j := \frac{4\|\psi_j\|_{L^\infty}}{r}.$$

In the second product on the right hand side of (3.5), we use the convention that a factor equals 1 if $\nu_j = 0$. Observe that from (3.2), we have

$$\|d\|_{\ell^1} = \sum_{j>J_0} d_j \le \frac{1}{3}. \tag{3.6}$$

3.2 Proof of ℓ^p-summability

The estimate (3.5) has the general form

$$\|t_\nu\|_V \le C_r\alpha(\nu_E)\,\beta(\nu_F). \tag{3.7}$$

Let $\mathcal{F}_E$ (respectively $\mathcal{F}_F$) be the collection of $\nu \in \mathcal{F}$ that are supported on E (respectively on F). Then, for any $0 < p < \infty$, we have

$$\sum_{\nu\in\mathcal{F}}\|t_\nu\|_V^p \le C_r^p\sum_{\nu\in\mathcal{F}}\alpha(\nu_E)^p\beta(\nu_F)^p = C_r^p\Big(\sum_{\nu\in\mathcal{F}_E}\alpha(\nu)^p\Big)\Big(\sum_{\nu\in\mathcal{F}_F}\beta(\nu)^p\Big) =: C_r^p\,A_E\,A_F. \tag{3.8}$$

In our particular setting, the first factor A_E is easily estimated by factorization:

$$A_E = \sum_{\nu\in\mathcal{F}_E}\alpha(\nu)^p = \sum_{\nu\in\mathcal{F}_E}\prod_{j\in E}\eta^{p\nu_j} = \prod_{j\in E}(\sum_{n\ge 0}\eta^{np}) = \Big(\frac{1}{1-\eta^p}\Big)^{J_0} < \infty. \tag{3.9}$$

So we are left with showing that A_F is finite. In our particular setting, we have

$$\beta(\nu) := \prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j} \le \frac{|\nu_F|^{|\nu_F|}}{\prod_{j\in F}\nu_j^{\nu_j}}d^{\nu_F}, \quad \nu\in\mathcal{F}_F, \tag{3.10}$$

where we have used the notation $d^{\nu_F} = \prod_{j\in F} d_j^{\nu_j}$ and our convention that $0^0 = 1$. We first transform the quantities of the form n^n into $n!$ by using Stirling type estimates: for all $n \ge 1$, we have

$$\frac{n!e^n}{e\sqrt{n}} \le n^n \le \frac{n!e^n}{\sqrt{2\pi}\sqrt{n}}. \tag{3.11}$$

Using the right inequality in (3.11) without even using the factor $\sqrt{2\pi}\sqrt{n}$, we obtain

$$|\nu_F|^{|\nu_F|} \le |\nu_F|!e^{|\nu_F|}.$$

On the other hand, using the left inequality in (3.11), we obtain

$$\prod_{j\in F}\nu_j^{\nu_j} \ge \frac{\nu_F!e^{|\nu_F|}}{\prod_{j\in F}\max\{1, e\sqrt{\nu_j}\}}.$$

Injecting these estimates into (3.10) gives

$$\beta(\nu) \le \frac{|\nu_F|!}{\nu_F!} d^{\nu_F} \prod_{j\in F} \max\{1, e\sqrt{\nu_j}\} \le \frac{|\nu_F|!}{\nu_F!} \bar{d}^{\nu_F}, \tag{3.12}$$

where

$$\bar{d}_j := e d_j, \quad j \in F.$$

Here we have used the crude bound $e\sqrt{n} \le e^n$ for $n \ge 1$ to replace $\max\{1, e\sqrt{\nu_j}\}$ by e^{ν_j}. We notice that

$$\|\bar{d}\|_{\ell^1} = e\|d\|_{\ell^1} \le \frac{e}{3} < 1.$$

Since $\bar{d}$ is also ℓ^p summable, we may apply Theorem 1.1 with $b = \bar{d}$, and conclude that A_F is finite.

3.3 Summability and convergence rate

We have just shown that the sequence $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ is in $\ell^p(\mathcal{F})$ and therefore in $\ell^1(\mathcal{F})$. We prove unconditional summability by the following standard argument.

Let $(\Lambda_N)_{N\ge1}$ be a sequence of finite sets which exhausts $\mathcal{F}$. We recall the particular sequence $(\Lambda_N^*)_{N\ge1}$ of Proposition 2.3. According to this result, we know that for all $\varepsilon > 0$ there exists $M = M(\varepsilon)$ such that

$$\|u(z) - S_{\Lambda_M^*} u(z)\|_V \le \frac{\varepsilon}{2},$$

for all $z \in \mathcal{U}$. Since $(\Lambda_N^*)_{N\ge1}$ exhausts $\mathcal{F}$, we may also assume that

$$\sum_{\nu\notin\Lambda_M^*} \|t_\nu\|_V \le \frac{\varepsilon}{2}.$$

Since $(\Lambda_N)_{N\ge1}$ exhausts $\mathcal{F}$, there exists N_0 such that $\Lambda_M^* \subset \Lambda_N$ for all $N \ge N_0$. It follows that for all $N \ge N_0$, and every $z \in \mathcal{U}$,

$$\|u(z) - S_{\Lambda_N} u(z)\|_V \le \|u(z) - S_{\Lambda_M^*} u(z)\|_V + \|S_{\Lambda_N\setminus\Lambda_M^*} u(z)\|_V \le \frac{\varepsilon}{2} + \sum_{\nu\notin\Lambda_M^*} \|t_\nu\|_V \le \varepsilon.$$

The convergence rate estimate (1.29) is obtained by writing for any $z \in \mathcal{U}$

$$\|u(z) - \sum_{\nu\in\Lambda_N} t_\nu z^\nu\|_V \le \sum_{\nu\notin\Lambda_N} \|t_\nu\|_V \le \sum_{n>N} \gamma_n,$$

where Λ_N is the set of $\nu \in \mathcal{F}$ corresponding to indices of the largest $\|t_\nu\|_V$ and where $(\gamma_n)_{n\ge1}$ is the decreasing rearrangement of the $\|t_\nu\|_V$. We then use the observation due to Stechkin that for any $0 < p \le q \le \infty$ and any $N \in \mathbb{N}$

$$\Big(\sum_{n>N} \gamma_n^q\Big)^{\frac{1}{q}} \le N^{\frac{1}{q}-\frac{1}{p}} \Big(\sum_{n\ge1} \gamma_n^p\Big)^{\frac{1}{p}}. \tag{3.13}$$

For $q < \infty$ this is easily proved by combining the two estimates

$$\sum_{n>N} \gamma_n^q \le \gamma_N^{q-p} \sum_{n>N} \gamma_n^p \le \gamma_N^{q-p} \sum_{n\ge1} \gamma_n^p \quad \text{and} \quad N\gamma_N^p \le \sum_{n\le N} \gamma_n^p \le \sum_{n\ge1} \gamma_n^p.$$

The case $q = \infty$ is a straightforward adaptation. Using (3.13) with $q = 1$ and $0 < p < 1$ we obtain

$$\sup_{z\in\mathcal{U}} \|u(z) - S_{\Lambda_N} u(z)\|_V \le N^{-s} \|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})}, \quad s := \frac{1}{p} - 1.$$

We close this section with some remarks on the nature of Theorem 1.3. This theorem states that whenever **UEAC**(r, R) is satisfied and if $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then $(\|t_\nu\|_V)_\nu \in \ell^p(\mathcal{F})$. Its norm, $\|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})}$ is the constant in the error bound of best N-term approximation. It depends on r and on the sequence $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1}$. However, inspection of the proof of Theorem 1.1 of [9], shows that this error bound depends on the sequence $(\psi_j)_{j\geq 1}$ not only through its norm $\|(\|\psi_j\|_{L^\infty(D)})\|_{\ell^p(\mathbb{N})}$ as one might expect, but also through other, rather implicit quantities such as the smallest integer J_1 such that $\sum_{j>J_1} \|\psi_j\|^p_{L^\infty(D)} < \varepsilon$ where ε in turn depends on r. A more explicit bound only in terms of r and $\|(\|\psi_j\|_{L^\infty(D)})\|_{\ell^p(\mathbb{N})}$ can be obtained if, for example, $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^q(\mathbb{N})$ for some $q < p$ (see Remark 7.7 in [9]).

4 Tensorized Legendre expansions

4.1 Statement of the result

In this section, we study another analytic expansion of u by changing the monomial basis to a Legendre basis. Our motivation for doing this is that if we agree to measure goodness of fit in a least squares sense, the Legendre expansions result in better decay estimates for N term approximation than monomial expansions (see Theorem 4.1 below).

We will consider two types of Legendre expansions which differ only in their normalization of the basis. In one variable, the Legendre basis $(P_n)_{n\geq 0}$ is usually defined with the L^∞ normalization

$$\|P_n\|_{L^\infty([-1,1])} = P_n(1) = 1. \tag{4.1}$$

We also consider the L^2 normalized sequence $L_n(t) = \sqrt{2n+1}P_n(t)$, which satisifies

$$\int_{-1}^{1} |L_n(t)|^2 \frac{dt}{2} = 1.$$

It is important to note that $L_0 = P_0 = 1$. We continue with our multivariate notation from the previous sections. For $\nu \in \mathcal{F}$, we consider the tensorized version of these polynomials

$$P_\nu(y) := \prod_{j\geq 1} P_{\nu_j}(y_j) \text{ and } L_\nu(y) := \prod_{j\geq 1} L_{\nu_j}(y_j). \tag{4.2}$$

Note that $(L_\nu)_{\nu\in\mathcal{F}}$ is an orthonormal basis of $L^2(U, d\mu)$ where $d\mu$ denotes the tensor product of the (probability) measures $\frac{dy_j}{2}$ on $[-1, 1]$ and is therefore a probability measure on $U = [-1, 1]^{\mathbb{N}}$. The probability measure $d\mu$ on U induces for $0 < p \leq \infty$ the measurable spaces $L^p(U, d\mu)$, and the Bochner spaces $L^p(U, V, d\mu)$ of μ-measurable mappings from U to V which are p-summable. We use these spaces for $p = 2$ and for $p = \infty$.

In contrast to the monomial expansion, we restrict our study to the approximation of $u(y)$ by tensorized Legendre series in U. We shall however make use of the analytic dependence of $u(z)$ on z in the estimation of the Legendre coefficients. Since $u \in L^\infty(U, V, d\mu) \subset L^2(U, V, d\mu)$, it admits unique expansions

$$u(y) = \sum_{\nu\in\mathcal{F}} u_\nu P_\nu(y) = \sum_{\nu\in\mathcal{F}} v_\nu L_\nu(y), \tag{4.3}$$

that converge in $L^2(U, V, d\mu)$, where the coefficients $u_\nu, v_\nu \in V$ are defined by

$$v_\nu := \int_U u(y)L_\nu(y)d\mu(y) \text{ and } u_\nu := \Big(\prod_{j\geq 1}(1 + 2\nu_j)\Big)^{1/2} v_\nu. \tag{4.4}$$

The following theorem is the analog to Theorem 1.3 for Legendre expansions.

Theorem 4.1 *If $a(x,z)$ satisfies* **UEAC**(r,R) *for some $0<r\le R<\infty$ and if $(\|\psi_j\|_{L^\infty})_{j\ge 1}\in \ell^p(\mathbb{N})$ for some $p<1$, then the sequences $(\|u_\nu\|_V)_{\nu\in\mathcal{F}}$ and $(\|v_\nu\|_V)_{\nu\in\mathcal{F}}$ belong to $\ell^p(\mathcal{F})$ for the same value of p. The Legendre expansions* (4.3) *converge in $L^\infty(U,V)$ in the following sense: if $(\Lambda_N)_{N\ge 1}$ is any sequence of finite sets which exhausts $\mathcal{F}$ then the partial sums $S_{\Lambda_N}u(y):=\sum_{\nu\in\Lambda_N}u_\nu(x)P_\nu(y)=\sum_{\nu\in\Lambda_N}v_\nu(x)L_\nu(y)$ satisfy*

$$\lim_{N\to+\infty}\sup_{y\in U}\|u(y)-S_{\Lambda_N}u(y)\|_V=0. \tag{4.5}$$

If Λ_N is the set of $\nu\in\mathcal{F}$ corresponding to indices of the largest $\|u_\nu\|_V$, we have in addition the convergence estimate

$$\sup_{y\in\mathcal{U}}\|u(y)-S_{\Lambda_N}u(y)\|_V\le\|(\|u_\nu\|_V)\|_{\ell^p(\mathcal{F})}N^{-s},\quad s:=\frac{1}{p}-1. \tag{4.6}$$

If Λ_N is the set of $\nu\in\mathcal{F}$ corresponding to indices of the largest $\|v_\nu\|_V$, we have the convergence estimate

$$\|u-S_{\Lambda_N}u\|_{L^2(U,V,d\mu)}\le\|(\|v_\nu\|_V)\|_{\ell^p(\mathcal{F})}N^{-s},\quad s:=\frac{1}{p}-\frac{1}{2}. \tag{4.7}$$

4.2 Estimates of the Legendre coefficients

In order to prove their $\ell^p(\mathcal{F})$ summability we estimate the quantities $\|u_\nu\|_V$ and $\|v_\nu\|_V$. By (4.4),

$$\|u_\nu\|_V=\Big(\prod_{j\ge 1}(1+2\nu_j)\Big)^{\frac{1}{2}}\|v_\nu\|_V,\qquad \nu\in\mathcal{F}. \tag{4.8}$$

Therefore the $\|v_\nu\|_V\le\|u_\nu\|_V$ and it will be sufficient to prove the ℓ^p summability of $(\|u_\nu\|_V)_{\nu\in\mathcal{F}}$.

Lemma 4.2 *Assume that* **UEAC**(r,R) *holds for some $0<r\le R<\infty$. Let $\rho=(\rho_j)_{j\ge 1}$ be a δ-admissible sequence for some $0<\delta<r$ that satisfies $\rho_j>1$ for all j such that $\nu_j\neq 0$. Then for any $\nu\in\mathcal{F}$ we have the estimate*

$$\|u_\nu\|_V\le\frac{\|f\|_{V^*}}{\delta}\prod_{j\ge 1,\nu_j\neq 0}\phi(\rho_j)(2\nu_j+1)\rho_j^{-\nu_j}, \tag{4.9}$$

where $\phi(t):=\frac{\pi t}{2(t-1)}$ for $t>1$.

Proof: Let $\nu=(\nu_j)_{j\ge 1}\in\mathcal{F}$ be fixed. The function valued coefficient u_ν is given by

$$u_\nu=\prod_{j=1}^{J}(2\nu_j+1)\int_U u(y)P_\nu(y)d\mu(y). \tag{4.10}$$

In the case $\nu=0$, the estimate (4.9) is immediate since $\mu(U)=1$ implies

$$\|u_0\|_V=\|\int_U u(y)d\mu(y)\|_V\le\sup_{y\in U}\|u(y)\|_V\le\frac{\|f\|_{V^*}}{r}\le\frac{\|f\|_{V^*}}{\delta}.$$

We now assume that $\nu\neq 0$. For notational simplicity we assume that $\nu_j\neq 0$ for $j\le J$ and $\nu_j=0$ for $j>J$ for some integer $J\ge 1$. This can always be achieved by reordering the basis $(\psi_j)_{j\ge 1}$. Partitioning the variable y into

$$y=(y_1,\cdots,y_J,y'),\quad y':=(y_{J+1},y_{J+2},\cdots)\in[-1,1]^{\mathbb{N}}=U,$$

we may rewrite (4.10) as

$$u_\nu=\prod_{j=1}^{J}(2\nu_j+1)\int_U w_\nu(y')d\mu(y'), \tag{4.11}$$

where

$$w_\nu(y') := \int\limits_{[-1,1]^J} u(y_1, \cdots, y_J, y') \left(\prod_{j=1}^{J} P_{\nu_j}(y_j) \right) \frac{dy_1}{2} \cdots \frac{dy_J}{2}.$$

For a fixed $y' \in U$, we now use the holomorphy properties of $(z_1, \cdots, z_J) \mapsto u(z_1, \cdots, z_J, y')$ in order to evaluate $\|w_\nu(y')\|_V$. For this purpose, we introduce for any $n \in \mathbb{N}$ the following function of a single complex variable w:

$$Q_n(w) := \int\limits_{-1}^{1} \frac{P_n(s)}{w - s} ds, \quad |w| > 1, \ n = 1, 2, \ldots, \tag{4.12}$$

and the multivariate functions

$$Q_\nu(z_1, \ldots, z_J) := \prod_{j=1}^{J} Q_{\nu_j}(z_j), \tag{4.13}$$

which are well defined as long as $|z_j| > 1$, whenever $\nu_j \neq 0$. Following [10] (page 19), we introduce for any $s > 1$ the ellipse in the complex plane

$$\mathcal{E}_s := \{ \frac{w + w^{-1}}{2} \ ; \ \ |w| = s \},$$

which has semi-axes of length $\frac{s+s^{-1}}{2}$ and $\frac{s-s^{-1}}{2}$. For our given ρ and ν, let

$$\mathcal{E}_{\rho,J} := \otimes_{1 \leq j \leq J} \, \mathcal{E}_{\rho_j}$$

be the tensor product of these ellipses in the variables $z_1, \ldots, z_J$.

Note that the ellipse $\mathcal{E}_s$ and its interior are contained in the closed disc of radius s, and therefore the polyellipse $\mathcal{E}_{\rho,J}$ and its interior are contained in the interior of the polydisc $\mathcal{U}_{\rho,J}$ defined in (2.16). Moreover $z = (z_1, \cdots, z_J, y')$ is contained in $\mathcal{U}_\rho$ for any $(z_1, \cdots, z_J)$ in $\mathcal{E}_{\rho,J}$ or its interior and any $y' \in U$. Since $[-1,1]^J$ is contained in the interior of $\mathcal{E}_{\rho,J}$, we may thus invoke the same holomorphy argument as in the proof of Lemma 2.4 and recursively apply Cauchy's integral formula in each ellipse $\mathcal{E}_{\rho_j}$ in the variables z_j, $j = 1, ..., J$ to obtain for any $(y_1, \cdots, y_J) \in [-1,1]^J$ and any $y' \in U$,

$$u(y_1, \cdots, y_J, y') = \frac{1}{(2\pi i)^J} \int\limits_{\mathcal{E}_\rho} \frac{u(z_1, \cdots, z_J, y')}{(y_1 - z_1) \cdots (y_J - z_J)} dz_1 \cdots dz_J. \tag{4.14}$$

Multiplying by $\prod_{j=1}^{J} P_{\nu_j}(y_j)$ and integrating over $[-1,1]^J$ with respect to $\frac{dy_1}{2} \cdots \frac{dy_J}{2}$, we therefore obtain

$$w_\nu(y') = 2^{-J} \int\limits_{\mathcal{E}_\rho} u(z_1, \cdots, z_J, y') Q_\nu(z_1, \ldots, z_J) dz_1 \cdots dz_J.$$

Since $\mathcal{E}_{\rho,J}$ is contained in $\mathcal{U}_{\rho,J}$, we find that for all

$$(z_1, \cdots, z_J) \in \mathcal{E}_{\rho,J} \text{ and } y' \in U \Rightarrow (z_1, \cdots, z_J, y') \in \mathcal{U}_\rho \Rightarrow \|u(z_1, \cdots, z_J, y')\|_V \leq \frac{\|f\|_{V^*}}{\delta}.$$

Injecting this bound in the above integral we thus obtain

$$\|w_\nu(y')\|_V \leq \frac{\|f\|_{V^*}}{\delta} \Big(\prod_{j=1}^{J} \frac{\rho_j}{2} \Big) \max_{(z_1, \ldots, z_J) \in \mathcal{E}_{\rho,J}} |Q_\nu(z_1, \ldots, z_J)|,$$

where we have used the fact that each of the ellipses $\mathcal{E}_{\rho_j}$ has perimeter of length $\leq 2\pi\rho_j$. We now use the following estimate established at the bottom of page 313 in [10]

$$\max_{z\in\mathcal{E}_t}|Q_n(z)| \leq \frac{\pi\ t^{-n}}{t-1},$$

which yields

$$\max_{(z_1,\ldots,z_J)\in\mathcal{E}_{\rho,J}}|Q_\nu(z_1,\ldots,z_J)| \leq \prod_{j=1}^{J}\frac{\pi\ \rho_j^{-\nu_j}}{\rho_j-1},$$

and therefore

$$\|w_\nu(y')\|_V \leq \frac{\|f\|_{V^*}}{\delta}\prod_{j=1}^{J}\phi(\rho_j)\rho_j^{-\nu_j}.$$

Combining this estimate with (4.11), we obtain the lemma. □

4.3 A choice of $\frac{r}{2}$-admissible sequences

We again work with the particular choice $\delta = \frac{r}{2}$ and the estimate (4.9) thus becomes

$$\|u_\nu\|_V \leq \frac{2\|f\|_{V^*}}{r}\prod_{j\geq 1,\nu_j\neq 0}\phi(\rho_j)(2\nu_j+1)\rho_j^{-\nu_j}, \tag{4.15}$$

for all $\frac{r}{2}$-admissible sequence such that $\rho_j > 1$ for all j such that $\nu_j \neq 0$ This estimate is slightly more pessimistic than the estimate (3.1) for the monomial series coefficients due to the presence of the factor $\phi(\rho_j)(2\nu_j+1)$, but we shall see that this does not affect the final result on ℓ^p summability. For this purpose we slightly modify the choice of the sequence ρ introduced in §3.1.

Namely, we now first fix $1 < \kappa \leq 2$ such that

$$(\kappa-1)\sum_{j\geq 1}\|\psi_j\|_{L^\infty(D)} \leq \frac{r}{8}. \tag{4.16}$$

With this κ fixed, we now take J_0 as the smallest integer such that

$$\sum_{j>J_0}\|\psi_j\|_{L^\infty(D)} \leq \frac{r(\kappa-1)}{18\pi\kappa}, \tag{4.17}$$

and define E and F for this new choice of J_0. For $j \in E$ we again take $\rho_j = \kappa$ but we now define

$$\rho_j := \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}} + 2, \quad \forall j \in F \text{ such that } \nu_j \neq 0, \tag{4.18}$$

and $\rho_j = 1$ if $j \in F$ and $\nu_j = 0$. Let us show that such a ρ is $\frac{r}{2}$ admissible. To this end, we estimate for every $x \in D$

$$\begin{aligned}
\sum_{j\geq 1}\rho_j|\psi_j(x)| &\leq \kappa\sum_{1\leq j\leq J_0}|\psi_j(x)| + \sum_{j\in F}\frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}|\psi_j(x)| + 2\sum_{j>J_0}|\psi_j(x)| \\
&\leq (\kappa-1)\sum_{1\leq j\leq J_0}|\psi_j(x)| + \sum_{1\leq j\leq J_0}|\psi_j(x)| + \frac{r}{4} + 2\sum_{j>J_0}|\psi_j(x)| \\
&\leq \frac{3r}{8} + \sum_{j\geq 1}|\psi_j(x)| + \sum_{j>J_0}|\psi_j(x)| \\
&\leq \frac{r}{2} + \sum_{j\geq 1}|\psi_j(x)|,
\end{aligned}$$

where for the last inequality we have used

$$\sum_{j>J_0} |\psi_j(x)| \le \sum_{j>J_0} \|\psi_j\|_{L^\infty(D)} \le \frac{r(\kappa-1)}{18\pi\kappa} < \frac{r}{8}.$$

Using (2.9) which follows from **UEAC**(r,R), we thus obtain for every $x \in D$

$$\sum_{j\ge 1} \rho_j |\psi_j(x)| \le \Re(\bar{a}(x)) - \frac{r}{2},$$

and therefore ρ is $\frac{r}{2}$ admissible. Since we also have that $\rho_j > 1$ for all j such that $\nu_j \neq 0$, the bound (4.15) is thus valid. With our particular choice of ρ, the estimate (4.15) reads

$$\|u_\nu\|_V \le \frac{2\|f\|_{V^*}}{r}\Big(\prod_{j\in E,\nu_j\neq 0} \phi(\rho_j)(2\nu_j+1)\eta^{\nu_j}\Big)\Big(\prod_{j\in F,\nu_j\neq 0} \phi(\rho_j)(2\nu_j+1)\big(\frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}+2\big)^{-\nu_j}\Big), \quad (4.19)$$

where $\eta := \frac{1}{\kappa} < 1$. Our next observation is that for all j such that $\nu_j \neq 0$ we have $\rho_j > \kappa$ and therefore

$$\phi(\rho_j) \le C_\kappa := \frac{\pi\kappa}{2(\kappa-1)}.$$

Using the crude estimate $C_\kappa(2n+1) \le (3C_\kappa)^n$ for $n \ge 1$, we find that (4.19) implies the estimate

$$\|u_\nu\|_V \le \frac{2\|f\|_V}{r}\Big(\prod_{j\in E,\nu_j\neq 0} C_\kappa(2\nu_j+1)\eta^{\nu_j}\Big)\Big(\prod_{j\in F}\Big(\frac{|\nu_F|\tilde{d}_j}{\nu_j}\Big)^{\nu_j}\Big), \quad (4.20)$$

where

$$\tilde{d}_j := \frac{12C_\kappa\|\psi_j\|_{L^\infty(D)}}{r}.$$

Using (4.17), we observe that

$$\|\tilde{d}\|_{\ell^1} = \sum_{j>J_0} d_j \le \frac{12C_\kappa}{r}\sum_{j>J_0}\|\psi_j\|_{L^\infty(D)} \le \frac{12C_\kappa}{r}\frac{r(\kappa-1)}{18\pi\kappa} \le \frac{1}{3}.$$

4.4 Proof of Theorem 4.1

Based on the estimate (4.20), the proof of the ℓ^p summability of $(\|u_\nu\|_V)_{\nu\in\mathcal{F}}$ is now very similar to the proof of the ℓ^p summability of $(\|t_\nu\|_V)_{\nu\in\mathcal{F}}$ in Theorem 1.3: we estimate

$$\sum_{\nu\in\mathcal{F}} \|u_\nu\|_V^p \le C_r^p \tilde{A}_E \tilde{A}_F,$$

as in (3.8), where $\tilde{A}_E$ and $\tilde{A}_F$ are slightly modified versions of the factors A_E and A_F. The first factor is given by

$$\tilde{A}_E = \sum_{\nu\in\mathcal{F}_E}\prod_{j\in E,\ \nu_j\neq 0}[C_\kappa(2\nu_j+1)]^p\eta^{p\nu_j} = S^{J_0},$$

where

$$S := 1 + C_\kappa^p\sum_{n\ge 1}(2n+1)^p\eta^{np} < \infty.$$

This gives $\tilde{A}_E < +\infty$.

The factor $\tilde{A}_F$ is exactly of the same form as A_F with the sequence $(d_j)_{j>J_0}$ replaced by $(\tilde{d}_j)_{j\geq 1}$ which has similar properties. Thus the same argument we used to estimate A_F shows that $\tilde{A}_F$ is finite. We have thus proved the ℓ^p summability of $(\|u_\nu\|_V)_{\nu\in\mathcal{F}}$ and in turn of $(\|v_\nu\|_V)_{\nu\in\mathcal{F}}$.

We already know that the series $\sum_{\nu\in\mathcal{F}} u_\nu P_\nu$ converges in $L^2(U,V,d\mu)$ towards u, and since

$$\sum_{\nu\in\mathcal{F}} \|u_\nu P_\nu\|_{L^\infty(U,V)} \leq \sum_{\nu\in\mathcal{F}} \|u_\nu\|_V < \infty,$$

this implies the unconditional convergence of this series in $L^\infty(U,V)$ in the sense expressed in the statement of the Theorem. The same of course holds for the series $\sum_{\nu\in\mathcal{F}} v_\nu L_\nu$ since it has the same partial sums.

The convergence rate estimate (4.6) is obtained by writing for any $y\in U$

$$\|u(y) - \sum_{\nu\in\Lambda_N} u_\nu P_\nu(y)\|_V \leq \sum_{\nu\notin\Lambda_N} \|u_\nu\|_V \leq \sum_{n>N} \gamma_n,$$

where $(\gamma_n)_{n\geq 1}$ denotes the decreasing rearrangement of the $\|u_\nu\|_V$, and then applying (3.13) with $q=1$.

The convergence rate estimate (4.7) is obtained by writing

$$\|u - \sum_{\nu\in\Lambda_N} u_\nu P_\nu\|^2_{L^2(U,V,d\mu)} \leq \sum_{\nu\notin\Lambda_N} \|v_\nu\|^2_V \leq \sum_{n>N} \gamma_n^2,$$

where now $(\gamma_n)_{n\geq 1}$ denotes the decreasing rearrangement of the $\|v_\nu\|_V$, and then applying (3.13) with $q=2$.

5 Spatial regularity and Finite Element discretization

The results of the previous sections allow us to understand how the functions $u(y)$ may be jointly approximated for all $y\in U$ with a prescribed accuracy by a finite linear combination $\sum_{\nu\in\Lambda_N} t_\nu y^\nu$, $\sum_{\nu\in\Lambda_N} u_\nu P_\nu(y)$ or $\sum_{\nu\in\Lambda_N} v_\nu L_\nu(y)$. The numerical realization of such a linear combination would itself involve the approximation of the $t_\nu, u_\nu, v_\nu \in V$ through discretization in D, such as, for example, by the Finite Element method. Specifically, we consider approximation of $u(y)$ in, for example, a bounded Lipschitz polyhedron D by a one parameter, affine family of continuous, piecewise linear Finite Element spaces $(V_h)_{h>0}$ on a shape regular family of simplicial triangulations of meshwidth $h>0$ in the sense of [8] (higher order, isoparametric Finite Element families in curved domains could equally be considered; we confine our analysis to affine, piecewise linear Finite Element families for ease of exposition only). Convergence rates of such Finite Element approximations are determined by the regularity of u in D. For this, further regularity assumptions on f are required. Again for ease of exposition, we shall assume $f\in L^2(D)\subset V^*$. Then

$$\|f\|_{V^*} \leq C_P \|f\|_{L^2(D)}, \tag{5.1}$$

where C_P is the Poincaré constant of D (i.e. $C_P = 1/\sqrt{\lambda_1}$ with λ_1 being the smallest eigenvalue of the Dirichlet Laplacian in D). Then the smoothness space $W\subset V$ is the space of all solutions to the Dirichlet problem

$$-\Delta u = f \quad \text{in} \quad D, \qquad u|_{\partial D} = 0, \tag{5.2}$$

with $f\in L^2(D)$

$$W = \{v\in V : \Delta v \in L^2(D)\}. \tag{5.3}$$

We define the W-(semi) norm and the W-norm by

$$|v|_W = \|\Delta v\|_{L^2(D)}, \qquad \|v\|_W := \|v\|_V + |v|_W. \tag{5.4}$$

It is well-known $W = H^2(D) \cap V$ for convex $D \subset \mathbb{R}^d$. Then any $w \in W$ may be approximated in V with convergence rate $\mathcal{O}(h)$ by continuous, piecewise linear Finite Element approximations on regular quasi-uniform simplicial partitions of D of meshwidth h (cf. e.g. [8, 6]). Therefore, denoting by $M = \dim(V_h) \sim h^{-d}$ the dimension of the Finite Element space, we have for all $w \in W$ the convergence rate

$$\inf_{v_h \in V_h} \|w - v_h\|_V \leq C M^{-\frac{1}{d}} |w|_W. \tag{5.5}$$

More generally, for non-convex polyhedra, the space W is not contained in $H^2(D)$, and the convergence rate as $M = \dim(V_h) \to \infty$ is reduced to

$$\inf_{v_h \in V_h} \|w - v_h\|_V \leq C_t M^{-t} |w|_W. \tag{5.6}$$

with some $0 < t < \frac{1}{d}$. Nevertheless, in the case where $D \subset \mathbb{R}^2$ is a non-convex polygonal domain, the optimal approximation rate (5.5) may be retained by suitable isotropic mesh refinement towards the reentrant corners of ∂D (see [2]). Similar results are available for non-convex Lipschitz polyhedra in $\mathbb{R}^3$ with plane faces for suitable anisotropic mesh refinement towards the reentrant corners and edges of ∂D (see [1]).

We shall prove the following result concerning W norms of the t_ν, u_ν and v_ν and their summability properties.

Theorem 5.1 *Assume that $f \in L^2(D)$ and that $a(x,z)$ satisfies* **UEAC**(r,R) *for some $0 < r \leq R < \infty$. If both $(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$ and $(\|\nabla\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then $(\|t_\nu\|_W)_{\nu\in\mathcal{F}}$, $(\|u_\nu\|_W)_{\nu\in\mathcal{F}}$ and $(\|v_\nu\|_W)_{\nu\in\mathcal{F}}$ belong to $\ell^p(\mathcal{F})$.*

Note that the result for the sequence $(\|v_\nu\|_W)_{\nu\in\mathcal{F}}$ is implied by that for $(\|u_\nu\|_W)_{\nu\in\mathcal{F}}$ due to (4.8). We shall use Theorem 5.1 for the analysis of convergence of Finite Element discretizations. In order to prove this theorem, we first study the analyticity properties of $z \mapsto u(z)$ as a map with values in W (we refer to this as W-analyticity). Accordingly, this leads us to estimates on the $\|\cdot\|_W$ norms of the coefficients t_ν and u_ν and to their ℓ^p summability. Throughout this section, we append assumptions i) - iii) with the additional assumption: iv) the gradients of the functions $\bar{a}$ and ψ_j, for $j \geq 1$, are defined for every $x \in D$ and belong to $L^\infty(D)$.

5.1 W-analyticity of u

We start our proof of W-analyticity with the observation that if $f \in L^2(D)$ and α satisfies the ellipticity assumption (1.5) and is such $\nabla\alpha \in L^\infty(D)$, then the solution u to (1.1) belongs to W. Indeed, from the identity

$$-\Delta u = \frac{1}{\alpha}(f + \nabla\alpha \cdot \nabla u), \tag{5.7}$$

we obtain the estimate

$$|u|_W = \|\Delta u\|_{L^2(D)} \leq \frac{1}{r}(\|f\|_{L^2(D)} + \|\nabla\alpha\|_{L^\infty(D)}\|u\|_V) \leq \frac{1}{r}(\|f\|_{L^2(D)} + \|\nabla\alpha\|_{L^\infty(D)}\frac{\|f\|_{V^*}}{r}), \tag{5.8}$$

and therefore,

$$\|u\|_W \leq \frac{1}{r}\Big(1 + C_P\Big(1 + \frac{\|\nabla\alpha\|_{L^\infty(D)}}{r}\Big)\Big)\|f\|_{L^2(D)}. \tag{5.9}$$

For $0 < \delta \leq 2R$ and $B > 0$, we introduce the complex domain

$$\mathcal{A}_{\delta,B} := \{z \in \mathbb{C}^{\mathbb{N}} \,;\, \delta \leq \Re(a(x,z)) \leq |a(x,z)| \leq 2R \text{ and } |\nabla a(x,z)| \leq B \text{ for all } x \in D\}, \tag{5.10}$$

where the gradient of a is taken with respect to the x variable. It is readily checked that under the assumptions of Theorem 5.1, for $0 < \delta < r$ and sufficiently large B the set $\mathcal{A}_{\delta,B}$ is non-empty. From (5.9), it is clear that $u(z) \in W$ for all $z \in \mathcal{A}_{\delta,B}$ and

$$\sup_{z\in\mathcal{A}_{\delta,B}} \|u(z)\|_W \leq C_{\delta,B} := \frac{1}{\delta}\Big(1 + C_P\Big(1 + \frac{B}{\delta}\Big)\Big)\|f\|_{L^2(D)}. \tag{5.11}$$

We will prove W-analyticity of u on $\mathcal{A}_{\delta,B}$ by using once more a difference quotient argument. The bounds on the difference quotient require the following perturbation result.

Lemma 5.2 *Let u and $\tilde{u}$ be solutions of (1.2) for the same $f \in L^2(D)$ with coefficients α and $\tilde{\alpha}$ which satisfy* (1.5)*, and which, in addition, are such that $\nabla\alpha$ and $\nabla\tilde{\alpha}$ belong to $L^\infty(D)$. Then there holds the estimate*

$$|u - \tilde{u}|_W \leq \frac{1}{r}\Big(\|\alpha - \tilde{\alpha}\|_{L^\infty(D)}|u|_W + \|\nabla(\alpha - \tilde{\alpha})\|_{L^\infty(D)}\|u\|_V + \|\nabla\tilde{\alpha}\|_{L^\infty(D)}\|u - \tilde{u}\|_V\Big). \tag{5.12}$$

Proof: We know from (5.9) that $u, \tilde{u} \in W$. Moreover, they satisfy $\tilde{\alpha}\Delta\tilde{u} - \alpha\Delta u = \nabla\alpha \cdot \nabla u - \nabla\tilde{\alpha} \cdot \nabla\tilde{u}$. Therefore we get

$$\Delta(u - \tilde{u}) = \frac{1}{\tilde{\alpha}}\Big((\alpha - \tilde{\alpha})\Delta u + \nabla(\alpha - \tilde{\alpha}) \cdot \nabla u + \nabla\tilde{\alpha} \cdot \nabla(u - \tilde{u})\Big).$$

Since $\tilde{\alpha}$ satisfies (1.5), we obtain (5.12) by taking the L^2 norm of this identity. □

Lemma 5.3 *At any $z \in \mathcal{A}_{\delta,B}$, the mapping $z \mapsto u(z)$ admits a complex derivative $\partial_{z_j} u(z) \in W$ with respect to each variable z_j.*

Proof: From the proof of Lemma 2.2, we know that the difference quotient $w_h(z)$ is solution to

$$-\text{ div }(a(x,z)\nabla w_h) = L_h = \nabla\psi_j \cdot \nabla u(z + he_j) + \psi_j \Delta u(z + he_j) \in L^2(D),$$

and therefore

$$-\Delta w_h = \frac{1}{a(x,z)}\Big(\nabla a(x,z) \cdot \nabla w_h + \nabla\psi_j \cdot \nabla u(z + he_j) + \psi_j \Delta u(z + he_j)\Big).$$

We already know from Lemma 2.2 that w_h converges in V towards $w_0 = \partial_{z_j} u(z)$. Similarly w_0 is solution to

$$-\Delta w_0 = \frac{1}{a(x,z)}\Big(\nabla a(x,z) \cdot \nabla w_0 + \nabla\psi_j \cdot \nabla u(z) + \psi_j \Delta u(z)\Big).$$

Both w_h and w_0 are therefore in W, and by subtracting the two equations we find that

$$|w_h - w_0|_W \leq \frac{1}{\delta}\Big(B\|w_h - w_0\|_V + \|\nabla\psi_j\|_{L^\infty(D)}\|u(z + he_j) - u(z)\|_V + \|\psi_j\|_{L^\infty(D)}|u(z + he_j) - u(z)|_W\Big).$$

As $h \to 0$, the three terms on the right hand side tend to 0, by Lemma 2.2, 2.1 and 5.2, respectively. This completes the proof. □

5.2 Proof of Theorem 5.1

Analogous to §2.1, we say that a sequence ρ is (δ, B)-admissible if and only if for all $x \in D$,

$$\sum_{j\geq 1} \rho_j |\psi_j(x)| \leq \Re(\bar{a}(x)) - \delta \text{ and } \sum_{j\geq 1} \rho_j |\nabla\psi_j(x)| \leq B - \|\nabla\bar{a}\|_{L^\infty(D)}. \tag{5.13}$$

It is immediate to check that if ρ is (δ, B)-admissible, then the polydisc $\mathcal{U}_\rho$ is contained in $\mathcal{A}_{\delta,B}$. This allows us to establish the following result for estimating the Taylor and Legendre coefficients.

Lemma 5.4 *If* **UEAC**(r,R) *holds for some* $0<r\leq R<\infty$ *and if* $\rho=(\rho_j)_{j\geq 1}$ *is a* (δ,B)*-admissible sequence for some* $0<\delta<r$ *and a sufficiently large value of* $B>0$*, then for any* $\nu\in\mathcal{F}$ *we have the estimate*

$$\|t_\nu\|_W\leq C_{\delta,B}\rho^{-\nu}, \tag{5.14}$$

where $C_{\delta,B}$ *is as in* (5.11) *and where we use the convention that* $t^{-0}=1$ *for any* $t\geq 0$*. If in addition, the sequence* ρ *satisfies* $\rho_j>1$ *for all* j *such that* $\nu_j\neq 0$*, we have the estimate*

$$\|u_\nu\|_W\leq C_{\delta,B}\prod_{j\geq 1,\nu_j\neq 0}\phi(\rho_j)(2\nu_j+1)\rho_j^{-\nu_j}, \tag{5.15}$$

where $\phi(t):=\frac{\pi t}{2(t-1)}$ *for* $t>1$*.*

Proof: the proof of (5.14) is the same as that of Lemma 2.4 and the proof of (5.15) is the same as that of Lemma 4.2, using analyticity and the global bound (5.11). □

We may now complete the proof of Theorem 5.1. Once again we work with the particular choice $\delta=\frac{r}{2}$. In order to prove the ℓ^p summability properties announced in Theorem 5.1, we need to slightly modify the definition of the $\frac{r}{2}$-admissible sequences ρ which were proposed in §3.1 and §4.3, since they should now also satisfy the second condition in (5.13) for some fixed $B>0$.

Let us explain this modification in the case of the estimates for the $\|t_\nu\|_W$. We first choose J_0 large enough such that

$$\sum_{j>J_0}(\|\psi_j\|_{L^\infty(D)}+\|\nabla\psi_j\|_{L^\infty(D)})\leq\frac{r}{12}, \tag{5.16}$$

We again split $\mathbb{N}$ into the two sets $E:=\{0<j\leq J_0\}$ and $F:=\mathbb{N}\setminus E$, and we choose $\kappa>1$ such that

$$(\kappa-1)\sum_{1\leq j\leq J_0}(\|\psi_j\|_{L^\infty(D)}+\|\nabla\psi_j\|_{L^\infty(D)})\leq\frac{r}{4}, \tag{5.17}$$

For our given ν we use the sequence ρ defined by

$$\rho_j:=\kappa,\quad j\in E;\quad \rho_j:=\max\{1,\frac{r\nu_j}{4|\nu_F|(\|\psi_j\|_{L^\infty(D)}+\|\nabla\psi_j\|_{L^\infty(D)})}\},\quad j\in F. \tag{5.18}$$

By the same considerations as in §3.1, we find that

$$\sum_{j\geq 1}\rho_j|\psi_j(x)|\leq\Re(\bar a(x))-\frac{r}{2},$$

but we also find that for every $x\in D$

$$\sum_{j\geq 1}\rho_j|\nabla\psi_j(x)|+\|\nabla\bar a\|_{L^\infty(D)}\leq B:=\kappa\sum_{1\leq j\leq J_0}\|\nabla\psi_j\|_{L^\infty(D)}+\sum_{j>J_0}\|\nabla\psi_j\|_{L^\infty(D)}+\frac{r}{4}+\|\nabla\bar a\|_{L^\infty(D)}. \tag{5.19}$$

We thus find that $\rho=(\rho_j)_{j\geq 1}$ is $(\frac{r}{2},B)$-admissible with this choice for B. This leads to the estimate

$$\|t_\nu\|_W\leq C_{\frac{r}{2},B}\Big(\prod_{j\in E}\eta^{\nu_j}\Big)\Big(\prod_{j\in F}\Big(\frac{|\nu_F|d_j}{\nu_j}\Big)^{\nu_j}\Big). \tag{5.20}$$

where $\eta:=\frac{1}{\kappa}<1$ and

$$d_j:=4\frac{\|\psi_j\|_{L^\infty(D)}+\|\nabla\psi_j\|_{L^\infty(D)}}{r},$$

satisfies $\|d\|_{\ell^1}=\sum_{j>J_0}d_j\leq\frac{1}{3}$.

From then on, the proof of the ℓ^p-summability of the $\|t_\nu\|_W$ is exactly the same as in §3.2, based on the estimate (5.20). A similar modification of the sequence ρ proposed in §4.3 in the case of the Legendre coefficients leads to a similar conclusion. The proof of Theorem 5.1 is therefore complete. □

5.3 Convergence rates of finite element approximations

We finally discuss the approximation of u by a linear combination of the form $\sum_{\nu\in\Lambda}\tilde{t}_\nu y^\nu$, $\sum_{\nu\in\Lambda}\tilde{u}_\nu P_\nu(y)$, or $\sum_{\nu\in\Lambda}\tilde{v}_\nu L_\nu(y)$ where $\Lambda\subset\mathcal{F}$ is finite and when the coefficients $\tilde{t}_\nu$, $\tilde{u}_\nu$ and $\tilde{v}_\nu$ are Finite Element approximations of t_ν, u_ν and v_ν, respectively, from finite element spaces $(V_\nu)_{\nu\in\Lambda}$ in the one-parameter family $(V_h)_{h>0}$ in (5.6). As indiated in [9, 5], to achieve optimal approximation rates in terms of the overall number of degrees of freedom denoted by N_{dof}, it will be crucial that for given $\nu\in\Lambda\subset\mathcal{F}$, the approximation space V_ν may depend on ν.

Let us consider the Taylor expansion (1.14). Under the assumptions of Theorem 5.1,

$$C_V := \|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})} \text{ and } C_W := \|(|t_\nu|_W)\|_{\ell^p(\mathcal{F})},\ \bar{C}_W := \|(\|t_\nu\|_W)\|_{\ell^p(\mathcal{F})}$$

are finite. We introduce the vector $\mathcal{M}=(M_\nu)_{\nu\in\Lambda}$ of the dimensions $M_\nu=\dim V_\nu$, $\nu\in\Lambda$, of the finite element approximation spaces V_ν used for approximating the t_ν. Note that for a sequence of Finite Element spaces obtained from successive mesh refinements, not all integers M may arise as dimensions. Moreover, when optimizing the choice of M_ν it will be convenient to allow M_ν to take noninteger values in $(0,\infty)$. We assume for now that the error bound (5.6) holds for all such M up to increasing C_t in this bound (we will ultimately remove this assumption below).

Thus, we express the approximation rate in terms of the total number of degrees of freedom involved in this approximation:

$$N_{dof} := \sum_{\nu\in\Lambda} M_\nu. \tag{5.21}$$

The approximation error in $L^\infty(U,V)$ may be estimated by

$$\sup_{y\in U}\|u(y)-\sum_{\nu\in\Lambda}\tilde{t}_\nu y^\nu\|_V \le \sum_{\nu\in\Lambda}\|t_\nu-\tilde{t}_\nu\|_V + \sum_{\nu\notin\Lambda}\|t_\nu\|_V. \tag{5.22}$$

The first term in the right hand side of (5.22) corresponds to the error occuring from the finite element discretization of the t_ν. According to (5.6), we may find $\tilde{t}_\nu\in V_\nu$ such that

$$\|t_\nu-\tilde{t}_\nu\|_V \le C_t M_\nu^{-t}|t_\nu|_W. \tag{5.23}$$

For example, we may take for $\tilde{t}_\nu$ the V-orthogonal projection of t_ν onto V_ν.

The second term in the right hand side corresponds to the error incurred by truncating the Taylor series. By taking $\Lambda:=\Lambda_N$ the set of indices corresponding to the largest $\|t_\nu\|_V$, it is bounded by

$$\sum_{\nu\notin\Lambda}\|t_\nu\|_V \le C_V N^{-s},\ \ s:=\frac{1}{p}-1. \tag{5.24}$$

Therefore, the global error is bounded by

$$\sup_{y\in U}\|u(y)-\sum_{\nu\in\Lambda_N}\tilde{t}_\nu y^\nu\|_V \le C_t\sum_{\nu\in\Lambda_N} M_\nu^{-t}|t_\nu|_W + C_V N^{-s}. \tag{5.25}$$

We now allocate the degrees of freedom M_ν in such a way that the total number is minimized for a fixed contribution $C_t\sum_{\nu\in\Lambda_N}M_\nu^{-t}|t_\nu|_W$ to the error. Since the global error bound (5.22) also contains the term $C_V N^{-s}$ which is independent of this allocation, it is natural to require that both contributions be of the same order. We therefore consider the minimization problem

$$\min\{\sum_{\nu\in\Lambda_N} M_\nu\ ;\ \sum_{\nu\in\Lambda_N} M_\nu^{-t}|t_\nu|_W \le N^{-s}\}. \tag{5.26}$$

To solve this problem, let us first treat the M_ν as continuous variables. Introducing a Lagrange multiplier, we obtain

$$M_\nu = \eta |t_\nu|_W^{\frac{1}{1+t}}, \tag{5.27}$$

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

$$N^{-s} = \sum_{\nu \in \Lambda_N} M_\nu^{-t} |t_\nu|_W = \eta^{-t} \sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}}, \tag{5.28}$$

and therefore

$$\eta = N^{\frac{s}{t}} \Big(\sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}} \Big)^{\frac{1}{t}}. \tag{5.29}$$

Combining this with (5.27) and summing over $\nu \in \Lambda_N$, we find

$$N_{dof} = N^{\frac{s}{t}} \Big(\sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}} \Big)^{\frac{1+t}{t}}. \tag{5.30}$$

We now distinguish between two cases.

(i) $t \leq s$: This also means that $p \leq \frac{1}{t+1}$ and therefore

$$\sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}} \leq C_W^{\frac{1}{1+t}} \tag{5.31}$$

for any set Λ_N. According to (5.30), we thus have

$$N_{dof} \leq C_W^{\frac{1}{t}} N^{\frac{s}{t}}. \tag{5.32}$$

Combining this with the fact that the global error is controlled by $(C_t + C_V)N^{-s}$, we find

$$\sup_{y \in U} \|u(y) - \sum_{\nu \in \Lambda_N} \tilde{t}_\nu y^\nu\|_V \leq C N_{dof}^{-t}, \tag{5.33}$$

where $C := (C_t + C_V)C_W$.

(ii) $s \leq t$: in this case $\sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}}$ may not be uniformly bounded and we estimate it using Hölder's inequality that gives

$$\sum_{\nu \in \Lambda_N} |t_\nu|_W^{\frac{1}{1+t}} \leq C_W^{\frac{1}{1+t}} N^\delta, \;\; \delta := 1 - \frac{1}{p(1+t)} > 0.$$

According to (5.30), we thus have

$$N_{dof} \leq C_W^{\frac{1}{t}} N^{\frac{s+(1+t)\delta}{t}} = C_W^{\frac{1}{t}} N. \tag{5.34}$$

Combining this with the fact that the global error is controlled by $(C_t + C_V)N^{-s}$, we find

$$\sup_{y \in U} \|u(y) - \sum_{\nu \in \Lambda_N} \tilde{t}_\nu y^\nu\|_V \leq C N_{dof}^{-s} \tag{5.35}$$

where $C := (C_t + C_V)C_W$.

We may summarize as follows the above dichotomy:

- In the first case the rate of convergence N_{dof}^{-t} is governed by the space discretization, since it is the same as the rate that would occur when approximating a *single instance* $u(z)$ in the V norm.
- In the second case, the rate of convergence N_{dof}^{-s} is governed by the discretization in the parameter variable, since it is the same as the rate that would occur without any space discretization.

This analysis therefore reveals that when the index set Λ and the finite element spaces V_ν are properly chosen, then the number of degrees of freedom resulting from space discretization and parameter discretization essentially *do not multiply*. This finding corresponds to the error bounds obtained for sparse tensor Finite Element discretizations in [5] under stronger assumptions.

In the above derivation, however, we assumed M_ν to be real-valued. To circumvent this problem, we assign for such a real-valued M_ν the subspace V_ν of dimension $\lfloor M_\nu \rfloor$. Therefore,

$$N_{dof} = \sum_{\nu\in\Lambda_N} \dim V_{M_\nu} \le \sum_{\nu\in\Lambda_N} M_\nu.$$

However, the approximation error estimate (5.23) is no longer valid, but one easily checks that it can be replaced by

$$\|t_\nu - \tilde{t}_\nu\|_V \le \bar{C}_t M_\nu^{-t} \|t_\nu\|_W \tag{5.36}$$

with $\bar{C}_t = \max\{1, 2^t C_t\}$. Solving the modified minimization problem

$$\min\{ \sum_{\nu\in\Lambda_N} M_\nu \; ; \; \sum_{\nu\in\Lambda_N} M_\nu^{-t} \|t_\nu\|_W \le N^{-s}\} \tag{5.37}$$

we end up with the same estimates (5.33) and (5.35), now with $C := (\bar{C}_t + C_V)\bar{C}_W$.

A similar analysis applies to the approximated Legendre expansions $\sum_{\nu\in\Lambda} \tilde{u}_\nu P_\nu(y)$ with the error measured in $L^\infty(U,V)$ and $\sum_{\nu\in\Lambda} \tilde{v}_\nu L_\nu(y)$ with the error measured in $L^2(U,V,d\mu)$. We collect these findings by the following theorem. We denote by $\tilde{t}_\nu$, $\tilde{u}_\nu$ and $\tilde{v}_\nu$ the V-projection of t_ν, u_ν and v_ν, respectively, onto V_ν.

Theorem 5.5 *Assume that the finite element spaces have the approximation property* (5.6)*. Then under the same assumptions as in Theorem 5.1, the following holds:*

(i) *With Λ_N the set of indices corresponding to the largest $\|t_\nu\|_V$, there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that*

$$\sup_{y\in U} \|u(y) - \sum_{\nu\in\Lambda_N} \tilde{t}_\nu y^\nu\|_V \le C N_{dof}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - 1,$$

where $N_{dof} = \sum_{\nu\in\Lambda_N} M_\nu$ and $C = (\bar{C}_t + \|(\|t_\nu\|_V)\|_{\ell^p(\mathcal{F})})\|(\|t_\nu\|_W)\|_{\ell^p(\mathcal{F})}$.

(ii) *With Λ_N the set of indices corresponding to the largest $\|u_\nu\|_V$, there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that*

$$\sup_{y\in U} \|u(y) - \sum_{\nu\in\Lambda_N} \tilde{u}_\nu P_\nu(y)\|_V \le C N_{dof}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - 1,$$

$N_{dof} = \sum_{\nu\in\Lambda_N} M_\nu$ and $C = (\bar{C}_t + \|(\|u_\nu\|_V)\|_{\ell^p(\mathcal{F})})\|(\|u_\nu\|_W)\|_{\ell^p(\mathcal{F})}$.

(iii) *With Λ_N the set of indices corresponding to the largest $\|v_\nu\|_V$, there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that*

$$\|u - \sum_{\nu\in\Lambda_N} \tilde{v}_\nu L_\nu\|_{L^2(U,V,d\mu)} \le C N_{dof}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - \frac{1}{2},$$

where $N_{dof} = \sum_{\nu\in\Lambda_N} M_\nu$ and $C = (\bar{C}_t^2 + \|(\|v_\nu\|_V)\|_{\ell^p(\mathcal{F})}^2)^{\frac{1}{2}}\|(\|v_\nu\|_W)\|_{\ell^p(\mathcal{F})}$.

References

[1] T. Apel, *Anisotropic finite elements: Local estimates and applications* Series "Advances in Numerical Mathematics", Teubner, Stuttgart, 1999.

[2] I. Babuška, R.B. Kellogg and J. Pitkäranta, *Direct and inverse estimates for finite elements with mesh refinement* , Numer. Math. **33** (1979) 447-471.

[3] I. Babuška, F. Nobile and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM J. Num. Anal., **45**(2007), 1005–1034.

[4] I. Babuska, R. Tempone and G. E. Zouraris, *Galerkin finite element approximations of stochastic elliptic partial differential equations*, SIAM J. Numer. Anal. **42**(2004), 800–825.

[5] M. Bieri, R. Andreev and Ch. Schwab, *Sparse Tensor Discretization of Elliptic sPDEs* SIAM J. Sci. Comput. **31**(6)(2009) 4281-4304.

[6] S. Brenner and L.R. Scott, *The mathematical theory of Finite Elements* (2nd Ed.), Springer (2008).

[7] A. Buffa, Y. Mada, A. T. Patera, C. Prudhomme, and G. Turinici, *A priori convergence of the greedy algorithm for the parameterized reduced basis* , preprint (November 2009)

[8] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Elsevier, Amsterdam 1978.

[9] A. Cohen, R. DeVore and C. Schwab, *Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs*, Report 2009-02, Seminar for Applied Mathematics, ETH Zürich.

[10] Philipp J. Davis, *Interpolation and Approximation*, Blaisdell Publishing Company, 1963.

[11] R. DeVore, *Nonlinear Approximation*, Acta Numerica **7**(1998), 51–150.

[12] S. Dineen, *Complex Analysis on Infinite Dimensional Spaces*, Springer Monographs in Mathematics, Springer Verlag, Berlin, 1999.

[13] Ph. Frauenfelder, Ch. Schwab and R.A. Todor: Finite elements for elliptic problems with stochastic coefficients Comp. Meth. Appl. Mech. Engg. **194** (2005) 205-228.

[14] R. Ghanem and P. Spanos, *Spectral techniques for stochastic finite elements*, Arch. Comput. Meth. Eng. **4**(1997), 63–100.

[15] M. Hervé, *Analyticity in infinite dimensional spaces*, De Gruyter, Berlin, 1989.

[16] M. Kleiber and T. D. Hien, *The stochastic finite element methods*, John Wiley & Sons, Chichester, 1992.

[17] R. Milani, A. Quarteroni and G. Rozza, *Reduced basis methods in linear elasticity with many parameters* Comp. Meth. Appl. Mech. Engg. **197** (2008), 4812-4829.

[18] F. Nobile, R. Tempone and C.G. Webster, *A sparse grid stochastic collocation method for elliptic partial differential equations with random input data* , SIAM J. Num. Anal. **46**(2008), 2309–2345.

[19] F. Nobile, R. Tempone and C.G. Webster, *An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data* SIAM J. Num. Anal. **46**(2008), 2411–2442.

[20] R. Todor, *Robust eigenvalue computation for smoothing operators*, SIAM J. Num. Anal. **44**(2006), 865–878.

[21] Ch. Schwab and R.A. Todor, *Karhúnen-Loève Approximation of Random Fields by Generalized Fast Multipole Methods*, Journal of Computational Physics **217** (2006), 100-122.

[22] R.A. Todor and Ch. Schwab, *Convergence Rates of Sparse Chaos Approximations of Elliptic Problems with stochastic coefficients* IMA Journ. Numer. Anal. **44**(2007) 232-261.

Albert Cohen
UPMC Univ Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France
cohen@ann.jussieu.fr

Ronald DeVore
Department of Mathematics, Texas A& M University, College Station, TX 77843, USA
rdevore@math.tamu.edu

Christoph Schwab
Seminar for Applied Mathematics, ETH Zürich, CH 8092 Zürich, Switzerland
schwab@math.ethz.ch

Research Reports

No.	Authors/Title
10-03	*A. Cohen, R. DeVore and C. Schwab* Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs
10-02	*V. Gradinaru, G.A. Hagedorn, A. Joye* Tunneling dynamics and spawning with adaptive semi-classical wave-packets
10-01	*N. Hilber, S. Kehtari, C. Schwab and C. Winter* Wavelet finite element method for option pricing in highdimensional diffusion market models
09-41	*E. Kokiopoulou, D. Kressner, N. Paragios, P. Frossard* Optimal image alignment with random projections of manifolds: algorithm and geometric analysis
09-40	*P. Benner, P. Ezzatti, D. Kressner, E.S. Quintana-Ortí, A. Remón* A mixed-precision algorithm for the solution of Lyapunov equations on hybrid CPU-GPU platforms
09-39	*V. Wheatley, P. Huguenot, H. Kumar* On the role of Riemann solvers in discontinuous Galerkin methods for magnetohydrodynamics
09-38	*E. Kokiopoulou, D. Kressner, N. Paragios, P. Frossard* Globally optimal volume registration using DC programming
09-37	*F.G. Fuchs, A.D. McMurray, S. Mishra, N.H. Risebrom, K. Waagan* Approximate Riemann solvers and stable high-order finite volume schemes for multi-dimensional ideal MHD
09-36	*Ph. LeFloch, S. Mishra* Kinetic functions in magnetohydrodynamics with resistivity and hall effects
09-35	*U.S. Fjordholm, S. Mishra* Vorticity preserving finite volume schemes for the shallow water equations
09-34	*S. Mishra, E. Tadmor* Potential based constraint preserving genuinely multi-dimensional schemes for systems of conservation laws
09-33	*S. Mishra, E. Tadmor* Constraint preserving schemes using potential-based fluxes. III. Genuinely multi-dimensional central schemes for for MHD equations