MOX–Report No. 30/2012

Convergence of quasi-optimal Stochastic Galerkin Methods for a class of PDES with random coefficients

Beck, J.; Nobile, F.; Tamellini, L.; Tempone, R.;

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Convergence of quasi-optimal Stochastic Galerkin Methods for a class of PDES with random coefficients*

Joakim Beck$^\flat$, Fabio Nobile$^{\sharp,\dagger}$, Lorenzo Tamellini$^{\sharp,\dagger}$, Raúl Tempone$^\flat$

July 14, 2012

$^\flat$ Applied Mathematics and Computational Science
4700 - King Abdullah University of Science and Technology
Thuwal 23955-6900, Kingdom of Saudi Arabia
`joakim.back.09@ucl.ac.uk`, `raul.tempone@kaust.edu.sa`

$^\sharp$ MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
`fabio.nobile@polimi.it`, `lorenzo.tamellini@mail.polimi.it`

† CSQI - MATHICSE
Ecole Politechnique Fédérale de Lausanne
Station 8, CH 1015, Lausanne, Switzerland

Keywords: Uncertainty Quantification, PDEs with random data, linear elliptic equations, multivariate polynomial approximation, best M -terms polynomial approximation, Stochastic Galerkin method, sub-exponential convergence

AMS Subject Classification: 41A10, 65C20, 65N12, 65N30, 65N35

Abstract

*The authors would like to recognize the support of the PECOS center at ICES, University of Texas at Austin (Project Number 024550, Center for Predictive Computational Science). Support from the VR project "Effektiva numeriska metoder för stokastiska differentialekvationer med tillämpningar" and King Abdullah University of Science and Technology (KAUST) AEA project "Predictability and Uncertainty Quantification for Models of Porous Media" is also acknowledged. The second and third authors have been supported by the Italian grant FIRB-IDEAS (Project n. RBID08223Z) "Advanced numerical techniques for uncertainty quantification in engineering and life science problems". The fourth author is a member of the KAUST SRI Center for Uncertainty Quantification in Computational Science and Engineering.

In this work we consider quasi-optimal versions of the Stochastic Galerkin Method for solving linear elliptic PDEs with stochastic coefficients. In particular, we consider the case of a finite number N of random inputs and an analytic dependence of the solution of the PDE with respect to the parameters in a polydisc of the complex plane $\mathbb{C}^N$. We show that a quasi-optimal approximation is given by a Galerkin projection on a weighted (anisotropic) total degree space and prove a (sub)exponential convergence rate. As a specific application we consider a thermal conduction problem with non-overlapping inclusions of random conductivity. Numerical results show the sharpness of our estimates.

1 Introduction

Partial differential equations with stochastic coefficients have been the subject of growing interest in the scientific community, as they conveniently describe situations in which the coefficients of the PDE are calibrated from noisy and limited measurements and a probabilistic uncertainty model is associated to them. In this context, one may be interested in computing statistics like mean or correlation of the solution of the PDE or possibly statistics of some observables of it, usually called "quantities of interest".

Sampling strategies are widely used to this end, ranging from plain Monte Carlo method to more sophisticated sampling techniques. However, in some cases it is possible to show that the solution is very smooth with respect to the random coefficients, and thus it may be reasonable to use polynomial approximations. In this work, we focus on linear elliptic equations with random diffusion coefficients. These equations exhibit an analytic dependence of the solution on the random input parameters, see e.g. [1, 3, 6, 7, 10, 9].

Two relevant polynomial approximation strategies that can be conveniently applied to the problem at hand are the Stochastic Galerkin [1, 16, 18, 19, 25] and the Stochastic Collocation methods [3, 14, 22, 26], which are a projection technique and an interpolation technique, respectively. In this work, we reconsider the quasi-optimal Stochastic Galerkin method proposed in the previous work [6], and provide rigorous convergence results in the special case in which the analyticity region contains a polydisc in the complex plane $\mathbb{C}^N$. Observe that in this context "quasi-optimal" means that the proposed methods are optimal with respect to upper bounds of the error, that we believe to be quite sharp.

In particular, we will derive, under the aformentioned assumptions, the decay of the coefficients of the polynomial expansion of the solution, following the proof in [11] (see also [7]). Next, following the construction of the quasi-optimal polynomial space proposed in [6] (and to some extent also in [7]) we will show that the well-known total degree polynomial space is a quasi-optimal choice for the Stochastic Galerkin method for the class of problems we are considering. We will then derive the corresponding convergence estimates with two different approaches. The first one is based on Taylor expansion and suitable for isotropic problems; the second one is based on the summability properties of the estimates of the Legendre coefficients of the solution and can be used in an anisotropic setting.

The class of problems considered here includes the example of a thermal conduction problem with non-overlapping inclusions of random conductivity, originally proposed in [4]. Hence, we will be able to reinterpret the numerical results there obtained in view of the estimates shown here. In particular, it will clearly appear that the theoretical estimates we propose capture correctly the behaviour that we observe numerically for the Legendre coefficients and the more than algebraic convergence rate of the global Galerkin error. However, they overestimate considerably the constants in the estimates. Nevertheless, they can be used as the correct ansatz to be fitted by numerical data, resulting in mixed "a priori"/"a posteriori" methods.

It is worth noting that a more general, yet less sharp, convergence estimate for the quasi-optimal Stochastic Galerkin method is provided in [10, 9], where however no explicit construction of the corresponding polynomial space is given. A possible a-priori formula to this end is given in [6], while [8, 17] propose construction of quasi-optimal polynomial spaces with adaptive strategies. In particular, [8] considers Taylor expansions, while [17] is essentially a perturbation method restricted to a small-noise hypothesis. Although very attractive, the main drawback of fully adaptive methods is the cost of exploration of the space of polynomials, that may not be negligible in high dimensions and can be avoided if the correct space of polynomials is prescribed by combining "a priori" information with "a posteriori" estimates.

The rest of this work is organized as follows: after having detailed in Section 2 the problem at hand and stated Assumption A3 on the analyticity requirements for the solutions considered, we will briefly review the Stochastic Galerkin methodology in Section 3. Section 4 presents then the convergence result for quasi-optimal Stochastic Galerkin method, while Section 5 shows that the solution of a generic "inclusions problem" satisfies the analyticity assumptions. Section 6 will recall the details of the inclusions test presented in [4] and show some numerical results that confirm the sharpness of the proposed estimates. Finally, Section 7 will draw some conclusions and perspectives.

2 Problem setting

2.1 A linear elliptic PDE with stochastic coefficients

Let D be a convex polygonal domain in $\mathbb{R}^d$, and let $(\Omega, \mathcal{F}, \mu)$ be a complete probability space, Ω being the set of outcomes, $\mathcal{F} \subset 2^\Omega$ the σ-algebra of events and $\mu : \mathcal{F} \to [0, 1]$ a probability measure. In this work we focus on the stochastic elliptic problem

Problem 1 (Strong formulation). *Find a random field $u : \overline{D} \times \Omega \to \mathbb{R}$, such* *that μ-almost surely there holds:*

$$\begin{cases} -\operatorname{div}(a(\mathbf{x},\omega)\nabla u(\mathbf{x},\omega)) = f(\mathbf{x}) & \mathbf{x} \in D, \\ u(\mathbf{x},\omega) = 0 & \mathbf{x} \in \partial D, \end{cases} \tag{1}$$

where the operators div *and* ∇ *imply differentiation with respect to the physical coordinate only.*

We will work under the following assumptions on the random field $a(\mathbf{x},\omega)$:

Assumption A1 (Continuity and coercivity)**.** *The coefficient $a(\cdot,\omega)$ is a strictly positive and bounded function over D for each random event $\omega \in \Omega$, i.e. there exist two positive constants $\infty > a_{max} > a_{min} > 0$ such that $a_{min} \leq a(\mathbf{x},\omega) \leq a_{max}$ μ-almost surely $\forall\, \mathbf{x} \in D$.*

Assumption A2 ("Finite Dimensional Noise Assumption")**.** *The diffusion coefficient $a(\mathbf{x},\omega)$ can be parametrized using a vector of N real-valued random variables, namely*

$$a(\mathbf{x},\omega) = a(\mathbf{x}, y_1(\omega), y_2(\omega), \ldots, y_N(\omega)\,).$$

Such random variables are independent and uniformely distributed, $\mathbf{y}(\omega) = (y_1(\omega), \ldots, y_N(\omega))^T : \Omega \to \Gamma \subset \mathbb{R}^N, \Gamma = \Gamma_1 \times \Gamma_2 \times \ldots \times \Gamma_N$. Without loss of generality, we further assume $\Gamma_i = [-1,1]$, so that the joint probability density function of $\mathbf{y}$, $\varrho : \Gamma \to \mathbb{R}_+$, factorizes as $\varrho(\mathbf{y}) = \prod_{n=1}^N \varrho_n(y_n)$, with $\varrho_n = \frac{1}{2}$.

Assumptions A1 and A2 deserve some comments. First, as an immediate consequence of Assumption A1 and the Lax–Milgram's Lemma we have well-posedness of problem (1) for μ-almost every $\omega \in \Omega$.

Next, under Assumption A2 the solution u of (1) depends on the single realization $\omega \in \Omega$ only through the value taken by the random vector $\mathbf{y}$. We can therefore replace the probability space $(\Omega, \mathcal{F}, \mu)$ with $(\Gamma, B(\Gamma), \varrho(\mathbf{y})d\mathbf{y})$, where $B(\Gamma)$ denotes the Borel σ-algebra on Γ and $\varrho(\mathbf{y})d\mathbf{y}$ is the measure of the vector $\mathbf{y}$.

Finally, we observe that more general problems can be addressed within this setting. In particular, problems depending on a set of N non-uniform random variables $z_1, \ldots, z_N$ may be included in this setting by introducing a non-linear map $y_i = \Theta(z_i)$ that transforms each of them into uniform random variables, following the well known theory on copulas, see [20]. In the case a mapping Θ is not available, one could still reduce the problem to the uniform case, by introducing an auxiliary density $\hat{\varrho} = \frac{1}{2^N}$ as suggested in [2]. This will lead to analogous error estimates as those derived in this work, up to a multiplicative constant factor proportional to $\|\varrho/\hat{\varrho}\|_{L^\infty(\Omega)}$. Even the assumption of independence of the random variables, although very convenient for the development of the tensorized techniques proposed below, is not essential and could be removed whenever the density ϱ does not factorize, again by introducing an auxiliary density $\hat{\varrho} = \frac{1}{2^N}$.

Observe however that this framework does not immediately include problems where $a(\mathbf{x}, \omega)$ is not bounded away from zero, like the important case where $a(\mathbf{x}, \omega)$ is a lognormal random field, i.e. $a(\mathbf{x}, \omega) = e^{\gamma(\mathbf{x}, \omega)}$, with $\gamma(\mathbf{x}, \omega)$ being a Gaussian random field.

Finally, we denote by $L^2_\varrho(\Gamma)$ the space of square integrable functions on Γ with respect to the measure $\frac{1}{2^N} d\mathbf{y}$, and by $V = H^1_0(D)$ the space of square integrable functions in D with square integrable distributional derivatives and with zero trace on the boundary, equipped with the gradient norm $\|v\|_V = \|\nabla v\|_{L^2(D)}$, $\forall v \in V$. Its dual space will be denoted by V'. Moreover, since V and $L^2_\varrho(\Gamma)$ are Hilbert spaces, we can define the tensor space $V \otimes L^2_\varrho(\Gamma)$ as the completion of formal sums $v(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{k'} v_{D,k}(\mathbf{x}) v_{\Gamma,k}(\mathbf{y})$, $\{v_{D,k}\} \subset V$, $\{v_{\Gamma,k}\} \subset L^2_\varrho(\Gamma)$ with respect to the inner product

$$(v, \widehat{v})_{V \otimes L^2_\varrho(\Gamma)} = \sum_{k,\ell} (v_{D,k}, \widehat{v}_{D,\ell})_V, (v_{\Gamma,k}, \widehat{v}_{\Gamma,\ell})_{L^2_\varrho(\Gamma)}.$$

We are now in the position to write a weak formulation of (1),

Problem 2 (Weak formulation)**.** *Find $u \in V \otimes L^2_\varrho(\Gamma)$ such that $\forall\, v \in V \otimes L^2_\varrho(\Gamma)$*

$$\int_\Gamma \int_D a(\mathbf{x}, \mathbf{y}) \nabla u(\mathbf{x}, \mathbf{y}) \cdot \nabla v(\mathbf{x}, \mathbf{y})\, \varrho(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} = \int_\Gamma \int_D f(\mathbf{x}) v(\mathbf{x}, \mathbf{y})\, \varrho(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}. \quad (2)$$

Thanks again to Assumption A1 and the Lax-Milgram lemma, there exists a unique solution to problem (2) for any $f \in V'$, with $\|u\|_{V \otimes L^2_\varrho(\Gamma)} \leq \frac{\|f\|_{V'}}{a_{min}}$. We remark that u can be understood either as a function in the tensor space $H^1_0(D) \otimes L^2_\varrho(\Gamma)$ or as a $H^1_0(D)$-valued square-integrable function of $\mathbf{y} \in \Gamma$, i.e. $u \in L^2_\varrho(\Gamma; H^1_0(D))$; we will use either notation depending on the situation.

2.2 Regularity of u with respect to the random parameters

Concerning the regularity of the solution u with respect to the input $\mathbf{y}$, it is well-known that, under reasonable assumptions on the regularity of the coefficient a, u is analytic in every $\mathbf{y} \in \Gamma$. We refer e.g. to [10, 9] for a proof in the case of linear dependence of the diffusion coefficient a on the parameters y_i, and to [6] for the more general case in which $a(\mathbf{x}, \mathbf{y})$ is infinitely many times differentiable with respect to $\mathbf{y}$ and $\exists\, r_1, \ldots, r_N \in \mathbb{R}_+$ s.t.

$$\left\| \frac{1}{a} \cdot \frac{\partial^{i_1 + \ldots + i_N} a}{\partial y_1^{i_1} \cdots \partial y_N^{i_N}} \right\|_{L^\infty(D)} \leq \prod_{n=1}^N r_n^{i_n} \quad \forall \mathbf{y} \in \Gamma, \quad \forall\, i_1, \ldots, i_N \in \mathbb{N}. \quad (3)$$

In this work, we will restrict our focus to the case in which u obeys the following Assumption:

Assumption A3 ("Polydisc Analyticity"). *The complex continuation of u, denoted by $u^* : \mathbb{C}^N \to H_0^1(D)$ is a $H_0^1(D)$-valued holomorphic function in the polydisc*

$$E_{S_1,\dots,S_N} = \prod_{n=1}^{N} E_{n,S_n}, \quad E_{n,S_n} = \{z_n \in \mathbb{C} : |z_n| \leq S_n\}$$

for each $1 < S_n < S_n^*$, with $\sup_{\mathbf{z}\in E_{S_1,\dots,S_N}} \|u^*(\mathbf{z})\|_{H_0^1(D)} \leq B_u$, and $B_u = B_u(S_1, S_2, \dots, S_N) \to \infty$ as $S_n \to S_n^*$, $n = 1, \dots, N$.

Remark 3. *We will see that this class of functions includes e.g. the solution of the inclusions tests already investigated in [4], as well as other elliptic problems that depend on few coefficients that can be varied independently one to another in given intervals. An example is given by elasticity problems with uncertain Young modulus and Poisson ratio. On the other hand, this is not the correct framework for diffusion coefficients that have the form $a(\mathbf{x}, \omega) = \sum_{n=1}^N b_n(\mathbf{x})y_n(\omega)$ with functions b_n with overlapping supports, which will be typically the case for a truncated Karhunen-Loève expansion of a correlated random field.*

Problem (2) can be discretized in space by introducing e.g. a finite element discretization with piecewise continuous polynomials over a triangulation $\mathcal{T}_h$ of the physical domain D, $V_h(D) \subset H_0^1(D)$. Such semi-discrete solution will thus belong to $V_h(D) \otimes L_\varrho^2(\Gamma)$, and will feature the same regularity properties of the continuous solution u.

3 Galerkin polynomial approximation in the stochastic dimension

As anticipated in the introduction, since u is a smooth function of y_i, $i = 1, \dots, N$, it is sound to approximate it with global polynomials. Thus, we introduce a polynomial subspace of $L_\varrho^2(\Gamma)$, which we denote by $\mathbb{P}(\Gamma)$, and look for a fully discrete solution $u_{h,\mathrm{w}} \in V_h(D) \otimes \mathbb{P}(\Gamma)$ solving

Problem 4 (Fully discrete weak formulation). *Find $u_{h,\mathrm{w}} \in V_h(D) \otimes \mathbb{P}(\Gamma)$ such that $\forall\, v \in V_h(D) \otimes \mathbb{P}(\Gamma)$*

$$\int_\Gamma\int_D a_N(\mathbf{x},\mathbf{y})\nabla u_{h,\mathrm{w}}(\mathbf{x},\mathbf{y}) \cdot \nabla v(\mathbf{x},\mathbf{y})\, \varrho(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} = \int_\Gamma\int_D f(\mathbf{x})v(\mathbf{x},\mathbf{y})\, \varrho(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}, \tag{4}$$

with the understanding that the polynomial space $\mathbb{P}(\Gamma)$ should be designed to have good approximation properties while having a number of degrees of freedom as low as possible. This is the well known *Stochastic Galerkin formulation* (see e.g. [1, 16, 18, 19, 25]). In this respect a Tensor Product polynomial space that contains all the N-variate polynomials with maximum degree in each variable lower than a given $\mathrm{w} \in \mathbb{N}$ is not a good choice. Indeed, its dimension grows exponentially fast with the number of random variables N, i.e. $\dim \mathbb{P}(\Gamma) = (1 + \mathrm{w})^N$. A valid alternative choice that has been widely used in literature (see e.g. [16, 24, 27]) is given by the Total Degree polynomial space, that includes those polynomials whose total degree is lower than or equal to w: such space contains indeed only $\binom{N+\mathrm{w}}{N}$ polynomials, which is much lower than $(1 + \mathrm{w})^N$, and still has good approximation properties. A number of possible polynomial spaces has been listed and analyzed e.g. in [4]. One could also introduce anisotropy in the approximation, with the aim to enrich the polynomial space only in those directions of the stochastic space which contribute the most to the total variability of the solution.

To solve Problem 4 in practice, it is convenient to endow $\mathbb{P}(\Gamma)$ with a $\varrho(\mathbf{y})d\mathbf{y}$-orthonormal basis: to this end we take advantage of the tensor structure of $L^2_\varrho(\Gamma)$ and build the elements of such basis as products of $\varrho_n(y_n)dy_n$-orthonormal polynomials on Γ_n, which we denote as $\{\Psi_{q_n}\}_{q_n \in \mathbb{N}}$:

$$\boldsymbol{\Psi}_{\mathbf{q}}(\mathbf{y}) = \prod_{n=1}^{N} \Psi q_n(y_n) \quad \mathbf{q} = (q_1,\, q_2,\, \ldots, q_n),\ \mathbf{q} \in \mathbb{N}^N. \tag{5}$$

Families of $\varrho_n(y_n)dy_n$-orthonormal polynomials exist for many probability distribution: we recall Legendre polynomials for uniform measures and Hermite polynomials for Gaussian measures (see [27] for the general Askey scheme), for which explicit formulae and computing algorithms are available, see e.g. [15]. As a word of caution, we can note that the work [13] showed that there exist probability measures, such as the lognormal one, which admit a family of orthonormal polynomials that however does not form a basis for $L^2_\varrho(\Gamma)$, i.e. there exist functions in $L^2_\varrho(\Gamma)$ that cannot be approximated with arbitrary precision by linear combinations of such orthonormal polynomials.

To construct general polynomial spaces we introduce a sequence of increasing index sets $\Lambda(\mathrm{w})$, $\mathrm{w} \in \mathbb{N}$, such that

$$\Lambda(0) = \{(0, \ldots, 0)\}, \quad \Lambda(\mathrm{w}) \subseteq \Lambda(\mathrm{w}+1) \subset \mathbb{N}^N \text{ for } \mathrm{w} \geq 0, \quad \mathbb{N}^N = \bigcup_{\mathrm{w} \in \mathbb{N}} \Lambda(\mathrm{w}),$$

each with cardinality M, and consider the corresponding polynomial spaces

$$\mathbb{P}_{\Lambda(\mathrm{w})}(\Gamma) = \mathit{span}\, \{\boldsymbol{\Psi}_{\mathbf{q}}(\mathbf{y}),\ \mathbf{q} \in \Lambda(\mathrm{w})\}$$

for the approximation of $u_{h,\mathrm{w}}$ with the Stochastic Galerkin method. In other words, the Stochastic Galerkin method will compute the coefficients $u_{\mathbf{q}} \in V_h(D)$ of the expansion

$$u_{h,\mathrm{w}}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{q} \in \Lambda(\mathrm{w})} u_{\mathbf{q}}(\mathbf{x}) \boldsymbol{\Psi}_{\mathbf{q}}(\mathbf{y}). \tag{6}$$

Such expansion is usually known as generalized Polynomial Chaos Expansion (gPCE). Having the gPCE expansion of $u_{h,\mathrm{w}}$ (6) allows us to compute easily the mean and variance of $u_{h,\mathrm{w}}$ as

$$\mathbb{E}\left[u_{h,\mathrm{w}}(\mathbf{x},\cdot)\right] = u_{\mathbf{0}}(\mathbf{x}), \qquad \mathbb{V}\mathrm{ar}\left[u_{h,\mathrm{w}}(\mathbf{x},\cdot)\right] = \sum_{\mathbf{q}\in\Lambda(\mathrm{w})} u_{\mathbf{q}}^2(\mathbf{x}) - u_{\mathbf{0}}^2(\mathbf{x}).$$

Finally, using (6) in the weak formulation (4) and choosing as test function $v_h(\mathbf{x})\boldsymbol{\Psi}_{\boldsymbol{\kappa}}(\mathbf{y})$, v_h being a finite element basis function, we obtain a set of M linear systems for the modes $u_{\mathbf{q}}(\mathbf{x})$, that will be usually coupled due to the presence in (4) of non-zero terms like $\int_{\Gamma_n} a(\mathbf{x},\mathbf{y})\Psi_{q_n}(y_{q_n})\Psi_{\kappa_n}(y_{\kappa_n})\varrho_n(y_n)dy_n$; see e.g. [4, 23] and references therein for more details on the discrete problem.

4 Quasi-optimal Stochastic Galerkin method for analytic functions in polydiscs

We consider here the basis for $\mathbb{P}_{\Lambda(\mathrm{w})}(\Gamma)$ given by multivariate Legendre polynomials. From the optimality if the Galerkin procedure, the Galerkin error is equivalent to the best approximation error measured in the $V\otimes L^2_\varrho(\Gamma)$ norm. The optimal M-dimensional polynomial space for the Stochastic Galerkin method is therefore the one spanning the Legendre polynomials corresponding to the M largest coefficients in the generalized Polynomial Chaos expansion (6). This choice indeed minimizes the energy of the projection error

$$\|u - \sum_{\mathbf{q}\in\Lambda(\mathrm{w})} u_{\mathbf{q}}\boldsymbol{\Psi}_{\mathbf{q}}\|^2_{V\otimes L^2_\varrho(\Gamma)} = \sum_{\mathbf{q}\notin\Lambda(\mathrm{w})} \|u_{\mathbf{q}}\|^2_V, \tag{7}$$

over all the possible choices of $\Lambda(\mathrm{w})$ with fixed cardinality M.

A possible strategy to assess the convergence rate of the resulting approximation of u is to order the Legendre coefficients $\|u_{\mathbf{q}}\|^2_V$ in decreasing order according to a suitable a-priori estimate and study the summability properties of the sequence thus obtained. This idea has been investigated e.g. in [10, 9] for the case when the diffusion coefficient can be written as $a(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{\infty} y_i b_i(\mathbf{x})$, with y_i uniform random variables over $[-1,1]$ and $\{\|b_i\|_\infty\}_{i\in\mathbb{N}} \in \ell^p$ for some $p<1$. It is then possible to prove an algebraic convergence of the L^2_ϱ error with rate $1/p - 1/2$. The proof is however not constructive, i.e. no algorithm is presented to build a sequence of polynomial approximations with such convergence rate. Uniform convergence results are given in [10], as well as in [8, 17].

In this work we will restrict our focus to the case in which the solution u obeys Assumption A3. In this case we are able to prove a subexponential rate of convergence and to give explicit formulae for the construction of the corresponding sequence of polynomial approximations.

4.1 Construction of the quasi-optimal polynomial space

We start by proving a result on the decay of the coefficients of the Legendre expansion for u satisfying Assumption A3. To this end, we first need the following simple Lemma, whose proof is straightforward.

Definition 5. *Let $\mathcal{E}_{\delta_1,\dots,\delta_N}$ be the family of Benrstein polyellipses $\mathcal{E}_{\delta_1,\dots,\delta_N} = \prod_{n=1}^N \mathcal{E}_{n,\delta_n}$ with*

$$\mathcal{E}_{n,\delta_n} = \Big\{ z_n \in \mathbb{C} : \mathfrak{Re}\,(z) = \frac{\delta_n + \delta_n^{-1}}{2} \cos\phi, \\ \mathfrak{Im}\,(z) = \frac{\delta_n - \delta_n^{-1}}{2} \sin\phi,\ \phi \in [0, 2\pi) \Big\}, \quad \delta_n > 1.$$

Lemma 6. *Let $\delta_n(S_n) = S_n + \sqrt{S_n^2 - 1}$, with S_n as in Assumption A3. The polyellipse $\mathcal{E}_{\delta_1(S_1),\dots,\delta_N(S_N)}$ is the largest polyellipse of the family $\mathcal{E}_{\delta_1,\dots,\delta_N}$ included in the polydisc $E_{S_1,\dots,S_N}$ in Assumption A3.*

Next, we also need to introduce the monodimensional $L^\infty(\Gamma)$-normalized Legendre polynomials $\Psi_j^\infty(t)$, $j = 0, 1, \dots$ for which the following properties hold:

- $\Psi_j^\infty(1) = 1$;
- $\Psi_j(t) = \sqrt{2j+1}\Psi_j^\infty(t)$ with $\Psi_j(t)$ as in (5).

We are now in position to prove the following estimate on the Legendre coefficients.

Proposition 7. *If the solution u fulfills Assumption A3, the coefficients of the Legendre expansion* (6) *decay as*

$$\|u_{\mathbf{q}}\|_V \le C_{Leg}\, e^{-\sum_{n=1}^N g_n q_n} \prod_{n=1}^N \sqrt{2q_n + 1}, \tag{8}$$

with $g_n = \log(\delta_n(S_n))$ and $C_{Leg}(S_1, \dots, S_N) = B_u(S_1, \dots, S_N) \prod_{n=1}^N \frac{l(\mathcal{E}_{n,\delta_n(S_n)})}{4(\delta_n(S_n) - 1)}$, for all $S_n < S_n^*$.

Here $l(\mathcal{E}_{n,\delta_n(S_n)})$ denotes the length of the ellipse $\mathcal{E}_{n,\delta_n(S_n)}$ in Lemma 6, $\delta_n(S_n)$ is as in Lemma 6, and $B_u(S_1, \dots, S_N)$ is as in Assumption A3.

Proof. The proof follows closely the argument in [11, Section 12.4]. Once fixed the radii $S_n < S_n^*$ in Assumption A3, from Lemma 6 we have that u is analytic in and on $\mathcal{E}_{\delta_1(S_1),\dots,\delta_N(S_N)}$, and hence we can exploit the Cauchy's formula to rewrite the $\mathbf{q}$-th Legendre coefficient as

$$\begin{aligned} u_{\mathbf{q}} &= \int_\Gamma u(\mathbf{x}, \mathbf{y}) \mathbf{\Psi}_{\mathbf{q}}(\mathbf{y}) \varrho(\mathbf{y}) d\mathbf{y} \\ &= \int_\Gamma \mathbf{\Psi}_{\mathbf{q}}(\mathbf{y}) \varrho(\mathbf{y}) \oint_{\mathcal{E}_{\delta_1(S_1),\dots,\delta_N(S_N)}} \frac{u^*(\mathbf{x}, \mathbf{z})}{\prod_n 2\pi i (z_n - y_n)} d\mathbf{z} d\mathbf{y} \\ &= \oint_{\mathcal{E}_{\delta_1(S_1),\dots,\delta_N(S_N)}} u^*(\mathbf{x}, \mathbf{z}) \prod_{n=1}^N \frac{1}{2} \int_{\Gamma_n} \frac{\Psi_{q_n}(y_n)}{2\pi i (z_n - y_n)} dy_n d\mathbf{z}. \end{aligned}$$

Next, let

$$\mathbb{I}_{q_n}(z_n) = \int_{\Gamma_n} \frac{\Psi^\infty_{q_n}(y_n)}{(z_n - y_n)} dy_n \,.$$

From [11, Lemma 12.4.6] it follows that for all $z_n \in \mathcal{E}_{n,\delta_n(S_n)}$ we have

$$|\,\mathbb{I}_{q_n}(z_n)| \leq \pi \frac{(1/\delta_n(S_n))^{q_n}}{\delta_n(S_n) - 1}.$$

Then we can estimate the $\mathbf{q}$-th Legendre coefficient of u by

$$\begin{aligned}
\|u_{\mathbf{q}}\|_V &\leq \sup_{\mathcal{E}_{\delta_1(S_1),\ldots,\delta_N(S_N)}} \|u^*\|_V \prod_{n=1}^N \frac{\sqrt{2q_n+1}}{4\pi} \oint_{\mathcal{E}_{n,\delta_n}} |\,\mathbb{I}_{q_n}(z_n)| dz_n \\
&\leq \sup_{\mathcal{E}_{\delta_1(S_1),\ldots,\delta_N(S_N)}} \|u^*\|_V \prod_{n=1}^N \frac{\sqrt{2q_n+1}}{4\pi} \pi \frac{(1/\delta_n(S_n))^{q_n}}{\delta_n(S_n)-1} \oint_{\mathcal{E}_{n,\delta_n}} dz_n \\
&\leq \sup_{\mathcal{E}_{\delta_1(S_1),\ldots,\delta_N(S_N)}} \|u^*\|_V \prod_{n=1}^N \frac{\sqrt{2q_n+1}\, l(\mathcal{E}_{n,\delta_n})}{4(\delta_n(S_n)-1)} e^{-q_n \log(\delta_n(S_n))}.
\end{aligned}$$

Finally observe that

$$\sup_{\mathcal{E}_{\delta_1(S_1),\ldots,\delta_N(S_N)}} \|u^*\|_V \leq \sup_{E_{S_1,\ldots,S_N}} \|u^*\|_V \leq B_u(S_1,\ldots,S_N).$$

□

Observe that the square root factor in (8) is asymptotically negligible compared to the exponentially decreasing term $e^{-\sum_n g_n q_n}$. Motivated by this fact, we introduce the following Corollary, that will be crucial in the following of the paper.

Corollary 8 (Exponential decay of the Legendre coefficients)**.** *The Legendre coefficients of u satisfying Assumption A3 can be accurately estimated as*

$$\|u_{\mathbf{q}}\|_V \leq \widehat{C}_{Leg} \prod_{n=1}^N e^{-\widehat{g}_n q_n}. \tag{9}$$

for some $\widehat{g}_n < g_n$ and $\widehat{C}_{Leg} > C_{Leg}$. For instance, for all $0 < \epsilon < 1$, one could take $\widehat{g}_n = g_n(1-\epsilon)$ and $\widehat{C}_{Leg} = C_{Leg}\prod_n \left(e^{\epsilon g_n/2}/\sqrt{\epsilon g_n e}\right)$.

Given the estimate for the decay of the Legendre coefficients of u in equation (9), the family of (anisotropic) Total Degree (TD) sets $\mathbb{P}_{TD(\mathrm{w},\widehat{\mathbf{g}})}(\Gamma)$, with

$$TD(\mathrm{w},\widehat{\mathbf{g}}) = \{\mathbf{q} \in \mathbb{N}^N : \sum_{n=1}^N \widehat{g}_n q_n \leq \mathrm{w}\},$$

is a sharp estimate of the optimal polynomial space for the Stochastic Galerkin method, provided that estimate (9) is in turn sharp. Indeed, following the procedure proposed in [6], one can define the quasi-optimal index set Λ by selecting all multi-indices $\mathbf{q}$ for which the *estimated decay* of the corresponding Legendre coefficient is above a fixed threshold $\epsilon \in \mathbb{R}_+$,

$$\Lambda_\epsilon = \left\{ \mathbf{q} \in \mathbb{N}^N : \widehat{C}_{Leg} \prod_{n=1}^{N} e^{-\widehat{g}_n q_n} \geq \epsilon \right\},$$

or equivalently

$$\Lambda(\mathrm{w}) = \left\{ \mathbf{q} \in \mathbb{N}^N : \sum_{n=1}^{N} \widehat{g}_n q_n \leq \mathrm{w},\ \mathrm{w} = \lceil -\log \epsilon / \widehat{C}_{Leg} \rceil \right\} = TD(\mathrm{w}, \widehat{\mathbf{g}}).$$

In the following, $u_{TD(\mathrm{w},\widehat{\mathbf{g}})} = \sum_{\mathbf{q}\in TD(\mathrm{w},\widehat{\mathbf{g}})} u_{\mathbf{q}} \mathbf{\Psi}_{\mathbf{q}} \in V \otimes \mathbb{P}_{TD(\mathrm{w},\widehat{\mathbf{g}})}(\Gamma)$ will denote the TD Stochastic Galerkin approximation of u.

4.2 Convergence analysis for the isotropic case

We begin the convergence analysis for the TD approximation of u from the isotropic setting, following closely the argument in [21]. Therefore, we further assume that Assumption A3 holds with $S_n = S$, for $n = 1, \ldots, N$. As a consequence, the parameters δ_n describing the polyellipses in Lemma 6 are all equal, as well as the coefficients $\widehat{g}_n$ driving the decay of the Legendre coefficients in Proposition 7 and Corollary 8. Thus the optimal polynomial space is indeed the isotropic TD, $TD(\mathrm{w}, \mathbf{1}) = \{\mathbf{q} \in \mathbb{N}^N : \sum_{n=1}^N q_n \leq \mathrm{w}\}$. For simplicity, we will denote this set simply as $TD(\mathrm{w})$, and the corresponding solution as $u_{TD(\mathrm{w})}$. Moreover, we will denote the polydiscs in Assumption A3 as E_S, the constant in Assumption A3 as $B_u(S)$ and the polyellipses in Lemma 6 and Proposition 7 as $\mathcal{E}_{\delta(S)}$.

We first recall the following optimality result for the Stochastic Galerkin approximation, whose proof can be found e.g. in [21].

Theorem 9. *Under Assumption A1, we have that the Stochastic Galerkin solution $u_{TD(\mathrm{w})}$ corresponding to $\mathbb{P}_{TD(\mathrm{w})}(\Gamma)$ satisfies*

$$\left\| u - u_{TD(\mathrm{w})} \right\|_{L^2_\varrho(\Gamma;V)} \leq C_{opt} \inf_{v \in V \otimes \mathbb{P}_{TD(\mathrm{w})}(\Gamma)} \| u - v \|_{L^2_\varrho(\Gamma;V)},$$

where C_{opt} is a constant depending on a_{min}, a_{max}.

Note that indeed such Theorem does not require the isotropic assumption. Next, we shall need the following Lemma (see [5] for a proof), which conversely relies on the hypothesis that all the radii of the analyticity polydiscs are equal.

Lemma 10. *Suppose that u satisfies Assumptions A3 with $S_n = S$ for $n = 1,\dots,N$, and let $\mathcal{M}_{u,\mathrm{w}}$ be the Maclaurin polynomial of u on the complex domain,*

$$\mathcal{M}_{u,\mathrm{w}}(\mathbf{z}) = \sum_{\mathbf{q}\in TD(\mathrm{w})} \alpha_{\mathbf{q}} \prod_{n=1}^{N} z_n^{q_n},$$

with $\alpha_{\mathbf{q}} \in V$, $\alpha_{\mathbf{q}}(\mathbf{x}) = \dfrac{1}{\prod_{n=1}^{N} q_n!} \dfrac{\partial^{q_1+\ldots+q_n}}{\partial y_1^{q_1}\cdots \partial y_N^{q_n}} u(\mathbf{x},\mathbf{y})_{|\mathbf{y}=\mathbf{0}}$.
Then, for any $0 < R < S$, we have the estimate

$$\sup_{\mathbf{z}\in E_R} \|u^*(\mathbf{z}) - \mathcal{M}_{u,\mathrm{w}}(\mathbf{z})\|_V \leq \frac{B_u(S)}{S/R-1} e^{-h\mathrm{w}},$$

with $B_u(S)$ as in Assumption A3 and $h = \log \dfrac{S}{R}$.

The convergence rate for the isotropic TD approximation can then be estimated combining Theorem 9 and Lemma 10.

Theorem 11. *Suppose that u satisfies Assumptions A3 with $S_n = S$ for $n = 1,\dots,N$. Then, the Stochastic Galerkin solution $u_{TD(\mathrm{w})}$ satisfies*

$$\left\|u - u_{TD(\mathrm{w})}\right\|_{L^2_\varrho(\Gamma;V)} \leq C_{opt} \frac{B_u(S)}{S-1} e^{-h\mathrm{w}},$$

with $B_u(S)$ as in Lemma 10, $h = \log S/R$, $R = 1$, and C_{opt} as in Theorem 9.

Proof. We use Lemma 10 with $R = 1$ (note that the intersection of E_1 with the real axis is Γ). Then we have

$$\begin{aligned}
\left\|u - u_{TD(\mathrm{w})}\right\|_{L^2_\varrho(\Gamma;V)} &\leq C_{opt} \inf_{v\in V\otimes\mathbb{P}_{TD(\mathrm{w})}(\Gamma)} \|u - v\|_{L^2_\varrho(\Gamma;V)} \\
&\leq C_{opt} \|u - \mathcal{M}_{\mathrm{w},u}\|_{L^2_\varrho(\Gamma;V)} \\
&\leq C_{opt} \|u - \mathcal{M}_{\mathrm{w},u}\|_{L^\infty(\Gamma;V)} \leq C_{opt} \frac{B_u(S)}{S-1} e^{-h\mathrm{w}}.
\end{aligned}$$

□

Theorem 11 states an exponential convergence of the error with respect to the total degree of the polynomial approximation. In practice however one is more concerned with the convergence of $u_{TD(\mathrm{w})}$ with respect to the number of degrees of freedom, i.e. the dimension M of the space $TD(\mathrm{w})$. Hence, we are lead to the problem of finding an estimate for the function $\mathrm{w} = \mathrm{w}(M)$.

Note that the inverse of such function, $M = M(\mathrm{w})$, is known analytically, $M = \binom{N+\mathrm{w}}{N}$. The function $\mathrm{w}(M)$ could thus be easily computed numerically: it is of course increasing in M and decreasing in N, i.e. the level w needed to have M terms in the set is lower for higher N, see Figure 1. In general, we can state the following proposition.

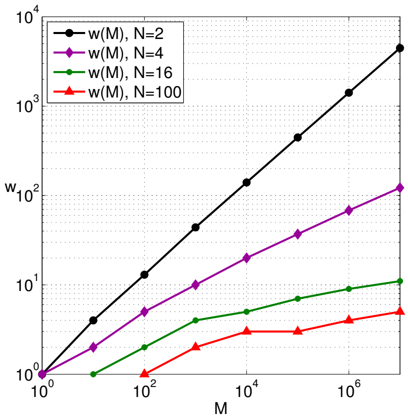

Figure 1: $\mathrm{w}(M)$ for different values of N.

Proposition 12. *For every $M > 0$, there holds*

$$\left\| u - u_{TD(\mathrm{w})} \right\|_{L^2_\varrho(\Gamma;V)} \leq C_{opt} \frac{B_u(S)}{S-1} M^{-h/(1+\log N)}, \tag{10}$$

with $B_u(S)$ as in Lemma 10, $h = \log S/R$, $R = 1$, and C_{opt} as in Theorem 9. Furthermore, in the asymptotic limit $\mathrm{w} \geq N$, that holds for instance if $M > 4^N$, there holds

$$\left\| u - u_{TD(\mathrm{w})} \right\|_{L^2_\varrho(\Gamma;V)} \leq C_{opt} \frac{B_u(S)}{S-1} e^{-\frac{hN}{2e} \sqrt[N]{M}}. \tag{11}$$

Proof. Equation (10) can be proved (see also [21, eq. 25]) by observing that

$$M = \prod_{i=1}^{N} \left(1 + \frac{\mathrm{w}}{i}\right) = \exp\left(\sum_{i=1}^{N} \log\left(1 + \frac{\mathrm{w}}{i}\right) \right) \leq \exp\left(\sum_{i=1}^{N} \frac{\mathrm{w}}{i} \right)$$
$$= \exp\left(\mathrm{w} \sum_{i=1}^{N} \frac{1}{i} \right) \leq e^{\mathrm{w}(\log(N)+1)}.$$

Therefore $\log M \leq \mathrm{w}\,(\log(N) + 1)$, hence $\mathrm{w} \geq \frac{\log M}{1+\log N}$ and $e^{-\mathrm{w}h} \leq M^{-h/(1+\log N)}$. In the asymptotic limit $\mathrm{w} \geq N$ we have instead

$$M = \prod_{i=1}^{N} \left(1 + \frac{\mathrm{w}}{i}\right) \leq \frac{2^N \mathrm{w}^N}{N!} \Rightarrow \mathrm{w} \geq \left(N! 2^{-N} M\right)^{1/N} \geq \frac{N}{2e} M^{1/N}.$$

Finally, using the well-known Stirling approximation of $N!$ we have that $\binom{2N}{N} \leq 4^N$ for all $N > 0$ so that $M > 4^N$ implies $\mathrm{w} \geq N$. □

4.3 Convergence analysis for the anisotropic case

In this Section we remove the isotropic assumption, and we derive a convergence estimate with an argument substantially different from the previous Section. We start with two technical Lemmas that we will need in the following.

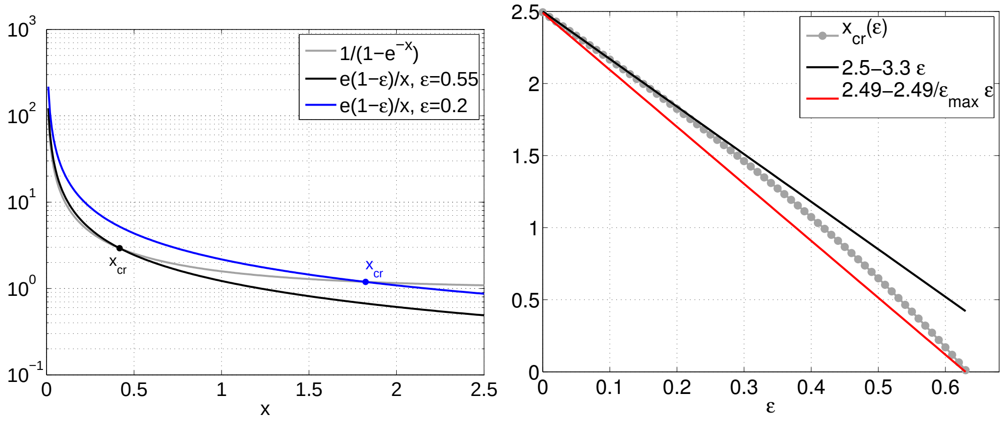

Figure 2: Left: Graphical representation of inequality (12) for $\epsilon = 0.2$ and $\epsilon = 0.55$. Right: value of x_{cr} and bound in equation (13).

Lemma 13. *For $0 < \epsilon < \frac{e-1}{e} = \epsilon_{max} \approx 0.63$, there holds*

$$\frac{1}{1-e^{-x}} \leq \frac{(1-\epsilon)e}{x}, \quad 0 < x \leq x_{cr}(\epsilon). \tag{12}$$

Moreover, the function $x_{cr}(\epsilon)$ can be bounded as

$$\alpha_L - \beta_L \epsilon \leq x_{cr(\epsilon)} \leq \alpha_U - \beta_U \epsilon, \tag{13}$$

with $\alpha_L = 2.49, \beta_L = (2.49/\epsilon_{max}), \alpha_U = 2.5, \beta_U = 3.3$.

Proof. For $x > 0$ and $\epsilon < 1$, (12) is actually equivalent to

$$e^{-x} \leq 1 - \frac{1}{(1-\epsilon)e} x$$

that can hold for $0 < x < x_{cr}(\epsilon)$ only if $-\frac{1}{(1-\epsilon)e} > -1$, hence $\epsilon < \frac{e-1}{e}$. Finally, equation (13) can be verified numerically. □

Lemma 14. *Given any $C_{log,M} \in (0, 1/e]$, there holds*

$$M \leq e^{C_{log,M} N \sqrt[N]{M}}, \tag{14}$$

for a sufficiently large M, $M > M_{log}$. In particular, for $C_{log,M} = 1/e$ the bound holds for any $M > 0$.

Proof. From the trivial observation that given any $C_{log,M}$ there holds $\log t \leq C_{log,M}\, t$ for sufficiently large t, we have immediately

$$\frac{1}{N} \log(M) = \log(\sqrt[N]{M}) \leq C_{log,M} \sqrt[N]{M} \Rightarrow \log(M) = C_{log,M} N \sqrt[N]{M},$$

hence the thesis of the Lemma. In particular, $\log t$ and $C_{log,M}\, t$ are tangent in $t = e$, with $C_{log,M} = 1/e$. □

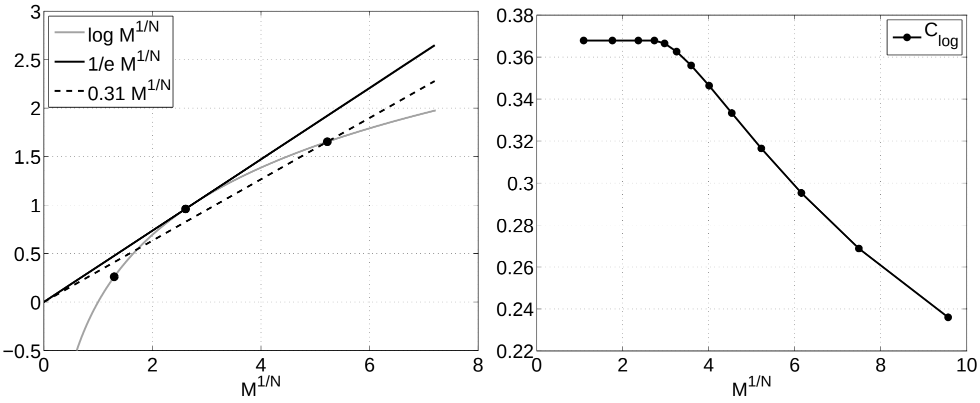

Figure 3: Graphical representation of Lemma 14. Left: graphical visualization of the bound. The dots represent $M^{1/N} = 0.1, 1/e, 0.5$ respectively. Note that the first point is less than $1/e$, therefore the same bound as $M^{1/N} = 1/e$ applies. Right: visualization of $M^{1/N}$ vs. C_{log} for $1 \leq M^{1/N} \leq 10$, numerically assessed.

Figure 2-left shows the effectiveness of (12), while Figure 2-right shows the function $x_{cr}(\epsilon)$, as well as the bounds in equation (13). Similarly, Figure 3-left show some instances of estimate (14), while Figure 3-right shows the value of C_{log} corresponding to a range of values of $\sqrt[N]{M}$.

Next, we consider again the expression for the error

$$\|u - \sum_{\mathbf{q}\in\Lambda(\mathrm{w})} u_{\mathbf{q}}\mathbf{\Psi}_{\mathbf{q}}\|^2_{V\otimes L^2_\varrho(\Gamma)} = \sum_{\mathbf{q}\notin\Lambda(\mathrm{w})} \|u_{\mathbf{q}}\|^2_V.$$

where $\Lambda(\mathrm{w})$ is the set of multiindices corresponding to the best M-terms approximation. Having estimated such optimal set with the total degree set $TD(\mathrm{w}, \widehat{\mathbf{g}})$ and the decay of the Legendre coefficients as exponential in each variable, according to Corollary 8, we have that

$$\begin{aligned}\|u - \sum_{\mathbf{q}\in\Lambda(\mathrm{w})} u_{\mathbf{q}}\mathbf{\Psi}_{\mathbf{q}}\|^2_{V\otimes L^2_\varrho(\Gamma)} &\leq \|u - \sum_{\mathbf{q}\in TD(\mathrm{w},\widehat{\mathbf{g}})} u_{\mathbf{q}}\mathbf{\Psi}_{\mathbf{q}}\|^2_{V\otimes L^2_\varrho(\Gamma)} \\ &= \sum_{\mathbf{q}\notin TD(\mathrm{w},\widehat{\mathbf{g}})} \|u_{\mathbf{q}}\|^2_V \leq \widehat{C}^2_{Leg} \sum_{\mathbf{q}\in\mathbb{N}^N, \mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}},\end{aligned}$$

and we will concentrate on bounding the last term of this inequality, $\sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}}$. To this end, we will need the so-called Stechkin Lemma, see e.g. [12].

Lemma 15 (Stechkin). *Let $0 \leq p \leq q$, and let $\{a_j\}_{j\in\mathbb{N}}$ be a positive decreasing sequence. Then*

$$\left(\sum_{j>M}(a_j)^q\right)^{1/q} \leq M^{-\frac{1}{p}+\frac{1}{q}} \left(\sum_{j\in\mathbb{N}}(a_j)^p\right)^{1/p}.$$

We are now ready to state the main result of this Section.

Theorem 16. *Suppose the Legendre coefficients of u can be bounded as in Corollary 8. Let g_h be the harmonic mean of the rates of the decay of the Legendre coefficients, $g_h = \sqrt[N]{\prod_{n=1}^N \widehat{g}_n}$. Consider the level* w *anisotropic TD approximation of u with rates $\widehat{\mathbf{g}}$, and denote by M its cardinality. Finally, let*

$$\mathbb{S}^\tau = \sum_{\mathbf{q}\in\mathbb{N}^N} e^{-2\tau\mathbf{q}\cdot\widehat{\mathbf{g}}} = \prod_{n=1}^N \frac{1}{1-e^{-2\tau\widehat{g}_n}} < \infty$$

for every $\tau > 0$. Then there holds

$$\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V\otimes L^2_\varrho(\Gamma)} \leq \widehat{C}^2_{Leg} \exp\left(N\sqrt[N]{M}\left(C_{log,M} - \frac{2g_h\delta}{e}\right)\right), \tag{15}$$

for $0 < \delta < \epsilon_{max}$, $C_{log,M}$ as in Lemma 14 and

$$M > \left(\frac{\widehat{g}_n e}{g_h(\alpha_L - \delta\beta_L)}\right)^N. \tag{16}$$

Proof. Using the estimate on the Legendre coefficients in Corollary 8 and Lemma 15 with $q = 1, p = \tau$, we have

$$\frac{1}{\widehat{C}^2_{Leg}}\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V\otimes L^2_\varrho(\Gamma)} = \sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}} \leq M^{1-\frac{1}{\tau}}\mathbb{S}. \tag{17}$$

Now, since (17) holds for every $\tau > 0$ we would like to compute τ^* minimizing $\frac{\mathbb{S}}{\sqrt[\tau]{M}}$,

$$\tau^* = \arg\min_{\tau\in\mathbb{R}_+} \frac{\mathbb{S}}{\sqrt[\tau]{M}} = \arg\min_{\tau\in\mathbb{R}_+} \left(\frac{1}{M\prod_{n=1}^N(1-e^{-2\tau\widehat{g}_n})}\right)^{1/\tau}$$

We do not solve exactly this problem and just discuss the approximated value $\tau^* = e/(2g_h\sqrt[N]{M})$. This choice is motivated in the case $\tau\widehat{g}_n$ small $\forall n = 1,\dots,N$, so that $1 - e^{-2\tau\widehat{g}_n} \approx 2\widehat{g}_n\tau$, as τ^* is the exact optimum solution of the approximated problem

$$\tau^* = \arg\min_{\tau\in\mathbb{R}} \left(\frac{1}{M\tau^N 2^N\prod_{n=1}^N \widehat{g}_n}\right)^{1/\tau}.$$

Plugging $\tau^* = e/(2g_h\sqrt[N]{M})$ in (17) we obtain

$$\sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}} \leq M\left(\frac{1}{M\prod_{n=1}^N(1-e^{-\widehat{g}_n e/(g_h\sqrt[N]{M})})}\right)^{2g_h\sqrt[N]{M}/e}. \tag{18}$$

Next we apply Lemma 13 to bound $1/\left(1-e^{-\widehat{g}_n e/(g_h\sqrt[N]{M})}\right)$, obtaining

$$\frac{1}{1-e^{-\widehat{g}_n e/(g_h\sqrt[N]{M})}} \leq \frac{(1-\epsilon_{M,n})g_h\sqrt[N]{M}}{\widehat{g}_n}, \quad \text{for } \frac{\widehat{g}_n e}{g_h\sqrt[N]{M}} \leq x_{cr}(\epsilon_{M,n}),$$

so that equation (18) simplifies to

$$\sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}} \leq M\left(\prod_{n=1}^{N}(1-\epsilon_{M,n})\right)^{2g_h\sqrt[N]{M}/e}. \tag{19}$$

From (19) one can then see that we need to apply Lemma 13 with $(1-\epsilon_{M,n})$ as small as possible and in particular $(1-\epsilon_{M,n})<1$ to ensure convergence of the estimate. Equivalently, we need to pick the largest ϵ possible in the range $\delta<\epsilon<\epsilon_{max}$, with $\delta>0$. Thus, for each $n=1,\ldots,N$ we choose $\epsilon=\epsilon_{M,n}$ according to the lower bound in (13), i.e.

$$\widehat{g}_n e/(g_h\sqrt[N]{M}) = \alpha_L-\beta_L\epsilon_{M,n} \Rightarrow \epsilon_{M,n} = \left(\alpha_L-\frac{\widehat{g}_n e}{g_h\sqrt[N]{M}}\right)\frac{1}{\beta_L}.$$

Note that the condition $\delta<\epsilon$ enforces a constraint on the minimum value of M,

$$\delta<\left(\alpha_L-\frac{\widehat{g}_n e}{g_h\sqrt[N]{M}}\right)\frac{1}{\beta_L} \Rightarrow \frac{\delta\beta_L-\alpha_L}{\widehat{g}_n e}g_h < -\frac{1}{\sqrt[N]{M}} \Rightarrow M > \left(\frac{\widehat{g}_n e}{g_h(\alpha_L-\delta\beta_L)}\right)^N.$$

Note that the rates are supposed to be ordered increasingly, so that this condition has to be checked for $n=N$ only. With this choice of $\epsilon_{M,n}$, equation (19) futher simplifies to

$$\begin{aligned}
\sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}} &\leq M\left(\prod_{n=1}^{N}(1-\epsilon_{M,n})\right)^{2g_h\sqrt[N]{M}/e} \\
&= M\exp\left(2g_h\sqrt[N]{M}/e\sum_{i=1}^{N}\log(1-\epsilon_{M,n})\right) \\
&\leq M\exp\left(-2g_h\sqrt[N]{M}/e\sum_{i=1}^{N}\epsilon_{M,n}\right) \\
&\leq M\exp\left(-\frac{2g_hN\sqrt[N]{M}\delta}{e}\right).
\end{aligned} \tag{20}$$

Finally, we apply Lemma 14, to obtain the final estimate

$$\sum_{\mathbf{q}\cdot\widehat{\mathbf{g}}>\mathrm{w}} e^{-2\mathbf{q}\cdot\widehat{\mathbf{g}}} \leq \exp\left(N\sqrt[N]{M}\left(C_{log,M}-\frac{2g_h\delta}{e}\right)\right).$$

□

Remark 17 (The role of δ). *Here we neglect the influence of $C_{log,M}$ in estimate* (15) *and further investigate the link between M and δ.*

On the one hand, choosing a small δ will reduce the minimum cardinality M for the estimate to hold, cf. equation (16)*; in the limit $\delta \to 0$, we have $M \geq \left(\frac{\widehat{g}_n e}{g_h \alpha_L}\right)^N$. In the isotropic case, $\widehat{g}_n = g_h$, estimate* (15) *is of the same form of estimate* (11) *in Proposition 12, however under the much milder condition $M \geq (e/\alpha_L)^N \approx 1.09^N$; in a problem with $N = 10$ random variables this would correspond to $M > 3$. On the other hand, $\delta = 0$ in* (15) *would imply no convergence rate. Conversely, the highest convergence would be obtained setting $\delta = \epsilon_{max}$ but would be realized only in the limit $M \to \infty$.*

Remark 18 (Recovering the isotropic result). *We can also compare this result with the isotropic estimate* (11) *in Proposition 12. In that case for $M > 4^N$ we had a rate of $hN/(2e)$, which one would obtain with* (15) *by choosing $\delta = \frac{h}{2g_h}$. Considering e.g. the isotropic problem detailed in next section one could estimate numerically $h \approx 1.5$, $g_h \approx 2$, that would imply*

$$M > \left(\frac{e}{\alpha_L - \frac{h}{2g_h}\beta_L}\right)^N \approx (2.7)^N \approx 2800,$$

or assess h, g theoretically in terms of the radii of the Bernstein ellipses and analyticity regions in Proposition 7 and Theorem 11, resulting in $h \approx 0.025$, $g_h \approx 0.22$ and then $M > 1.2^N$.

The main drawback of (15) *is that for anisotropic problems condition* (16) *on M is dominated by the largest rate, $\widehat{g}_N$. However, for problems with large variations of $\widehat{g}_n$ the random variables corresponding to high values of $\widehat{g}_n$ will not be added to approximations of u with small cardinality M: therefore, one may think of devising an "adaptive" estimate in which the constraint on M and the convergence rate depend on the active variables only.*

Remark 19 (The interplay between C_{log} and δ). *We now also investigate through some numerical computations the effect of C_{log} on estimate* (15)*. To this end, let us denote $C_\delta = \frac{g_h \delta}{e}$, so that estimate* (15) *can be written as*

$$\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V \otimes L^2_\varrho(\Gamma)} \leq \widehat{C}^2_{Leg} \exp\left(N \sqrt[N]{M}\,(C_{log,M} - 2C_\delta)\right).$$

For simplicity, we will work in an isotropic setting, $g_h = \widehat{g}_n$ for $n = 1, \ldots, N$. We consider a uniform sampling of the admissible values of δ, $0 < \delta < \epsilon_{max}$: for each of these values we compute the corresponding values of C_δ and of $\sqrt[N]{M}$ according to equation (16)*, i.e. $\frac{\widehat{g}_n e}{g_h(\alpha_L - \delta\beta_L)}$ (note that in the isotropic case $\widehat{g}_n$ and g_h cancel), and finally we compute numerically C_{log} corresponding to such $\sqrt[N]{M}$. By comparing the values of C_{log} and C_δ thus obtained we can see (cf. Table 1) that $C_{log,M}$ plays a non-negligible role, preventing the estimate to go to zero as $M \to \infty$ for small values of δ. This phenomenon is however alleviated if g_h is higher.*

			$g_h = 1$		$g_h = 2$	
δ	$\sqrt[N]{M}$	C_{log}	$2C_\delta$	rate	$2C_\delta$	rate
0	1.09	0.368	0	0.368	0	0.368
0.05	1.19	0.368	0.0368	0.331	0.0736	0.294
0.1	1.3	0.368	0.0736	0.294	0.147	0.221
0.15	1.43	0.368	0.11	0.258	0.221	0.147
0.2	1.6	0.368	0.147	0.221	0.294	0.0736
0.25	1.81	0.368	0.184	0.184	0.368	0
0.3	2.08	0.368	0.221	0.147	0.441	-0.073
0.35	2.45	0.368	0.258	0.11	0.515	-0.147
0.4	2.97	0.366	0.294	0.0722	0.589	-0.222
0.45	3.79	0.352	0.331	0.0205	0.662	-0.311
0.5	5.22	0.316	0.368	-0.0514	0.736	-0.419
0.55	8.4	0.253	0.405	-0.151	0.809	-0.556
0.6	21.5	0.143	0.441	-0.299	0.883	-0.74

Table 1: Numerical values for C_{log} and C_δ.

We finally close this section with an alternative estimate, presented here for the isotropic case only. Towards this end, we now present a couple of auxiliary results.

Lemma 20. *Let $\widehat{\mathbf{g}} = (\widehat{g}_1, \ldots, \widehat{g}_N)$ be a vector of positive entries. For every $\tau > 0$, define*

$$\mathbb{S} = \left(\sum_{\mathbf{q} \in \mathbb{N}^N} e^{-\tau \mathbf{q} \cdot \widehat{\mathbf{g}}} \right)^{1/\tau} = \left(\prod_{n=1}^{N} \frac{1}{1 - e^{-\tau \widehat{g}_n}} \right)^{1/\tau} < \infty, \tag{21}$$

and let $M = \sum_{\mathbf{q} \cdot \widehat{\mathbf{g}} \leq \mathrm{w}} 1$. Then

$$e^{-\mathrm{w}} \leq \frac{\mathbb{S}}{\sqrt[\tau]{M}}\,. \tag{22}$$

Proof. We have immediately that $M e^{-\tau \mathrm{w}} \leq \sum_{\mathbf{q} \cdot \widehat{\mathbf{g}} \leq \mathrm{w}} e^{-\tau \mathbf{q} \cdot \widehat{\mathbf{g}}} \leq \mathbb{S}^\tau$. □

Lemma 21. *Consider two non negative sequences, $\{a_j\}_{j \in \mathbb{N}}$ monotone decreasing and $\{f_j\}_{j \in \mathbb{N}}$ monotone increasing. Then, for a given $\lambda \in (0,1)$ and $M > 0$ we have*

$$\sum_{j > M} a_j^2 \leq \frac{1}{f_M} \sup_{j > M} \left\{ a_j^{2\lambda} f_j \right\} \sum_{j > M} a_j^{2(1-\lambda)}.$$

Theorem 22 (Alternative Isotropic Estimate)**.** *Under the same conditions of Theorem 16 assume that we have, letting $\epsilon_{\max} \approx 0.63$ as in Lemma 13 and a* *sufficiently large M, namely that $1.09^N < M$. In addition, assume that $g_n = g$, for $n = 1, \dots, N$. Then the estimates*

$$\begin{aligned}\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V\otimes L^2_\varrho(\Gamma)} &\le \widehat{C}^2_{Leg}(1-\exp(-g))^{-N}\exp\left(-\frac{gN}{e}\log((1-\epsilon(M))^{-1})\sqrt[N]{M}\right)\\ &\le C(g)^N\, M^{-g/e\,\log((1-\epsilon(M))^{-1})\,(1+1/2\,\log(M)/N)}\end{aligned} \tag{23}$$

hold, with $C(g) = \widehat{C}^{2/N}_{Leg}\,\frac{\exp\left(-g/e\,\log((1-\epsilon(M))^{-1})\right)}{1-\exp(-g)}$, and

$$\epsilon(M) = \epsilon_{max}\left(1 - \frac{1.09}{\sqrt[N]{M}}\right). \tag{24}$$

Proof. Let $\mathbb{S}$ be as in (21). Using the estimate on the norm of the Legendre coefficients in Corollary 8, here denoted by the sequence $\{a_j\}_{j\in\mathbb{N}}$, and combining Lemma 21 with $\lambda = 1/2$ and Lemma 20 with the choice $f_j = \frac{\sqrt[\tau]{j}}{\mathbb{S}}$ yields $\sup_{j>M}\{a_j f_j\} \le 1$. Therefore, we can estimate

$$\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V\otimes L^2_\varrho(\Gamma)} = \sum_{j>M} a_j^2 \le \widehat{C}^2_{Leg}(1-\exp(-g))^{-N}\min_{\tau>0}\frac{\mathbb{S}}{\sqrt[\tau]{M}}.$$

Consider as before, for a given value of $\tau > 0$, the approximate minimization of $\frac{\mathbb{S}}{\sqrt[\tau]{M}} = \left(\frac{1}{M\,(1-e^{-\tau g})^N}\right)^{1/\tau}$. Taking $\tau = \frac{e}{g\sqrt[N]{M}}$ yields

$$\begin{aligned}\sum_{j>M} a_j^2 \le& \widehat{C}^2_{Leg}(1-e^{-g})^{-N}\left(\frac{1}{M(1-e^{-e/\sqrt[N]{M}})^N}\right)^{\frac{g\sqrt[N]{M}}{e}}\\ \le& \widehat{C}^2_{Leg}(1-e^{-g})^{-N}\exp\left(-\frac{gN}{e}\log((1-\epsilon)^{-1})\sqrt[N]{M}\right)\end{aligned} \tag{25}$$

which holds as long as (cf. Lemma 13) $M \ge \left(\frac{e}{\alpha_L - \epsilon\beta_L}\right)^N = \left(\frac{e}{\alpha_L(1-\epsilon\beta_L/\alpha_L)}\right)^N = \left(\frac{e}{\alpha_L(1-\epsilon/\epsilon_{max})}\right)^N \approx \left(\frac{1.09}{1-\epsilon/0.63}\right)^N > 1.09^N$. Observe now that the choice (24) is optimal for the bound (25). Finally, the last inequality in (23) follows from (25) recalling the inequality $M^{1/N} \ge 1 + \frac{\log(M)}{N} + \frac{1}{2}\left(\frac{\log(M)}{N}\right)^2$. □

5 The inclusions problem

We now consider a generic "inclusions problem" in which the diffusion coefficient in (1) is given by

$$a(\mathbf{x},\mathbf{y}) = a_0 + \sum_{n=1}^{N}\gamma_n\chi_n(\mathbf{x})y_n, \tag{26}$$

where $\chi_n(\mathbf{x})$ are the indicator functions of the disjoint subdomains $D_n \subset D = [0,1]^2, D_n \cap D_m = \varnothing$ for $n \neq m$, and y_n are independent random variables uniformly distributed in $[y_{min}, y_{max}]$ with $y_{min} > -a_0$, so that Assumptions A1 and A2 are satisfied, as well as condition (3) ensuring the analyticity of u. Finally, γ_n are real coefficients, $0 < \gamma_n \leq 1$, whose values determines the possible anisotropy of the problem.

We will first prove that we can apply Corollary 8, and therefore that the TD sets are quasi-optimal sets for such problem. Then, we will apply Theorems 11 and 16 and show that the numerical results obtained for such problem are in agreement with the predicted convergence rates.

We shall begin by reparametrizing the diffusion coefficient in terms of new random variables distributed over $[-1,1]$, so that we can apply the discussion of the previous Section. For the sake of notation, we will still denote the new variables as y_i, i.e. $y_i \sim \mathcal{U}(-1,1)$. The new diffusion coefficient will be therefore

$$a(\mathbf{x},\mathbf{y}) = a_0 + \sum_{n=1}^{N} \gamma_n \chi_n(\mathbf{x}) \left(\frac{y_n+1}{2}(y_{max}-y_{min}) + y_{min} \right). \tag{27}$$

We can now prove the following lemma on the analyticity region of u, that we denote by Σ.

Lemma 23. *The solution u to* (4) *corresponding to a diffusion coefficient* (27) *is analytic in the region*

$$\Sigma = \prod_{n=1}^{N} \Sigma_n, \quad \Sigma_n = \{ z_n \in \mathbb{C} : \mathfrak{Re}\,(z_n) > T_n \},$$

with $-1 > T_n > T_n^* = \dfrac{2a_0 + \gamma_n(y_{max}+y_{min})}{\gamma_n(y_{min}-y_{max})}$. *Moreover,* $\sup_{\mathbf{z}\in\Sigma} \|u*\|_{H_0^1(D)} \leq B_u(T)$, *with*

$$B_u(T) = \frac{\|f\|_{V'}}{a_0 + \sum_{n=1}^{N} \gamma_n \left(\frac{1-|T_n|}{2}(y_{min}-y_{max}) + y_{min} \right)}.$$

Proof. As already pointed out, since u satisfies condition (3) then it is analytic in each direction y_n. In particular, having fixed the values of all the random variables but the n-th, let us write $a_n^*(\mathbf{x}, z_n) = a(\mathbf{x}, y_1, y_2, \ldots, y_{n-1}, z_n, y_{n+1}, \ldots, y_N)$ and $u_n^*(\mathbf{x}, z_n) = u(\mathbf{x}, y_1, y_2, \ldots, y_{n-1}, z_n, y_{n+1}, \ldots, y_N)$. Such u_n can be extended in $\Sigma_n = \{z_n \in \mathbb{C} : \mathfrak{Re}\,(z_n) > T_n\}$ for every T_n with $-1 > T_n > T_n^*$, where T_n^* is computed as the value such that

$$\exists\, \mathbf{x} \in D : a_n(\mathbf{x}, T_n^*) = a(\mathbf{x}, y_1, y_2, \ldots, y_{n-1}, T_n^*, y_{n+1}, \ldots, y_N) = 0.$$

This amounts to impose

$$a_0 + \gamma_n \left(\frac{T_n^*+1}{2}(y_{max}-y_{min}) + y_{min} \right) = 0,$$

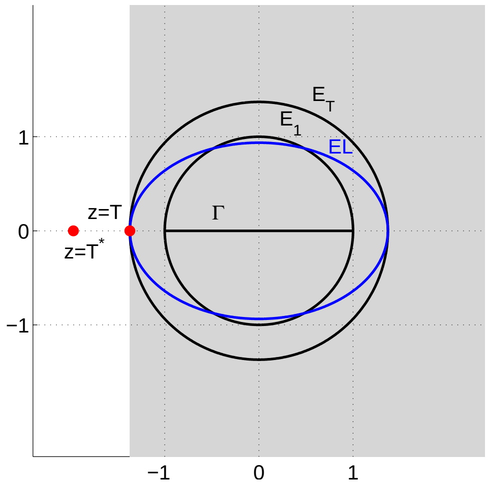

Figure 4: Regions of the complex plane along the n-th direction for the inclusions problem. For simplicity we drop here the subscript n in the plot. The gray area denotes the analyticity region Σ_n considered. $z_n = T_n^*$ is the singularity up to which it is possible to extend u^* along y_n. EL is the ellipse used to estimate the decay of the Legendre coefficients (Proposition 7/Corollary 8), while E_1 and E_T are the circles used to prove the convergence of TD estimates in the case of an isotropic setting $\gamma_1, \gamma_2, \ldots, \gamma_N = \gamma$ (Theorem 11).

whose solution is $T_n^* = (2a_0 + \gamma_n(y_{max} + y_{min}))/\big(\gamma_n(y_{min} - y_{max})\big)$. Indeed, since the subdomains D_n do not overlap, $a_n(\mathbf{x}, T_n^*) = 0$ in D_n only, i.e. T_n^* does not depend on the value of any of the other random variable y_i. Thus, the analyticity region of u is the cartesian product of the analyticity regions Σ_n, and the bound for $B_u(T)$ follows immediately. □

5.1 Convergence results

Theorems 11 and 16 apply immediately: in particular, see Figure 4 for Theorem 11. We summarize the results for the inclusions problem in the following proposition.

Proposition 24.

1. *The Legendre coefficients of the solution of the inclusions problem decay as*

$$\|u_{\mathbf{q}}\|_V \leq C(\epsilon) e^{-(1-\epsilon)\mathbf{q}\cdot\mathbf{g}}, \tag{28}$$

with

$$g_n = \log(|T_n| + \sqrt{T_n^2 - 1}), \quad -1 > T_n > T_n^*, \tag{29}$$

T_n^* *as in Lemma 23 and* ϵ *an in Corollary 8.*

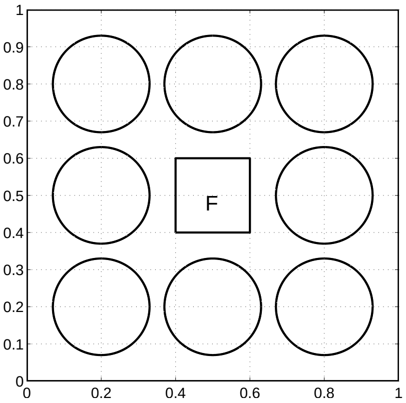

Figure 5: Physical domain for the isotropic inclusions problem, The inclusions are labelled anti-clock-wise, starting from the bottom-left corner.

2. *The polynomial space* $\mathbb{P}_{TD(\mathrm{w},\mathbf{g})}(\Gamma)$ *is the quasi-optimal space for the Stochastic Galerkin method when solving the inclusion problem.*

3. *The convergence rate of such quasi-optimal approximation is stated in Theorem 16, equation* (15)*.*

4. *Moreover, in the isotropic setting where* $\gamma_1, \gamma_2, \ldots, \gamma_N = \gamma$*, there holds* $T_1^* = T_2^* = \ldots = T_N^* = T^*, g_1 = g_2 = \ldots = g_N = g$ *and we also have an exponential decay of the error with respect to* w *with rate* $h = \log |T|$*, as stated in Theorem 11, equation* (11)*.*

In the forthcoming Section we will verify the quality of this analysis, both in an isotropic and an anisotropic setting. However, instead of (15) we will actually consider a simplified ansatz, cf. (23), i.e.

$$\|u - u_{TD(\mathrm{w},\widehat{\mathbf{g}})}\|^2_{V \otimes L^2_\varrho(\Gamma)} \leq C \exp\left(-\frac{2g_h}{e} N \sqrt[N]{M}\right) \tag{30}$$

and verify that if provides a sharp bound of the error for all $M > 0$.

6 Numerical results

6.1 Isotropic problem

We now consider the inclusions problem analyzed in [4]. In the first setting considered the subdomains in equation (26) are $N = 8$ disjoint circular subdomains as in Figure 5, $\gamma_n = 1$ for every $n = 1, \ldots, 8$. The random variable y_n are uniformly distributed in $[-0.99, -0.2]$. In addition, we choose $a_0 = 1$ and $f = 100\chi_F$, χ_F being the indicator function of the square located in the middle of the domain, cf. Figure 5. The aim of this Section is to reanalyze the numerical results obtained in [4] in view of the Theorems just proved. In that work, we considered several polynomial approximation spaces, and for each of them we computed the corresponding Stochastic Galerkin approximations, u_{SG}. Then, we introduced the bounded linear functional $\Theta : H_0^1(D) \to \mathbb{R}$,

$$\Theta(u) = \int_F u(\mathbf{x}) d\mathbf{x}$$

and we monitored the convergence of $\Theta(u_{SG})$ with respect to the L^2 norm error for the Stochastic Galerkin approximation,

$$\varepsilon = \sqrt{\mathbb{E}\left[(\Theta(u_{SG}) - \Theta(u_{ref}))^2\right]}. \tag{31}$$

Note that for this problem we do not have an exact solution, therefore the error is computed with respect to a reference solution. To this end, we have considered the Stochastic Galerkin approximation computed for the TD polynomial space at level $\mathrm{w} = 9$, which includes approximately 24000 Legendre polynomials. The L^2 error is calculated via a Monte Carlo approximation, i.e.

$$\varepsilon \simeq \left(\frac{1}{W_{MC}} \sum_{l=1}^{W_{MC}} \left[\Theta(u_{SG}(\mathbf{y}_l)) - \Theta(u_{ref}(\mathbf{y}_l))\right]^2 \right)^{1/2}, \tag{32}$$

where $\mathbf{y}_l$, $l = 1, .., W_{MC}$, are randomly chosen points in Γ. To this end, $W_{MC} = 1000$ points have proven to be enough to recover a smooth convergence curve.

Figure 6 shows the effectiveness of the proposed estimate (28) for the decay of the Legendre coefficients in the gPCE expansion of $\Theta(u)$. Indeed, after having computed the Galerkin solution, we have at disposition the coefficients of the gPCE expansion of u, that we can compare with (28). The rates $\widehat{g}$ have been assessed by fitting the Legendre coefficients computed, but the procedure described in [4, 6] could have been employed as well. Their numerical value is roughly around $1.90 - 1.99$, i.e. there is not a perfect isotropy: this can be explained by the fact that the inclusions are not equally distant from the control area F. Observe that the theoretical rate predicted is at most $\log(|T^*| + \sqrt{T^{*2} - 1}) \approx 0.22$. Thus the estimate we provide in Corollary 8 captures the right behaviour of the decay of the Legendre coefficients (i.e. exponential), but is very conservative. Yet, it can still provide the ansatz for a calibrated estimate, which is what we propose in this work.

Figure 7-left shows the convergence with respect to the level w of the $L^2_\varrho(\Gamma)$ error for the TD approximation of $\Theta(u)$, and shows an optimal agreement between the numerical results and the exponential decay predicted in Theorem 11. Note however that the rate h observed experimentally is $h \approx 1.5$, which is again much larger than the theoretically predicted rate, which amounts to at most $h = \log |T^*| \approx 0.025$.

Figure 7-right shows instead the convergence with respect to M of the error (31) squared for different polynomial approximations, (namely: Total Degree

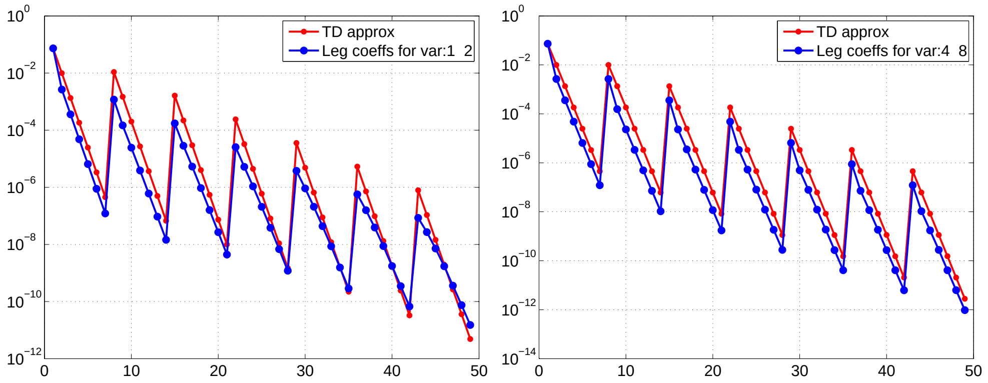

Figure 6: Comparison between some coefficients of gPCE expansion of $\Theta(u)$, computed with a highly accurate Galerkin approximation (TD(9)) and the corresponding bound (28) suitably tuned. The multiindices corresponding to the coefficients shown in the plots are nonzero only in $y_1 - y_2$ (left) and $y_4 - y_8$ (right) and ordered in lexicographic order.

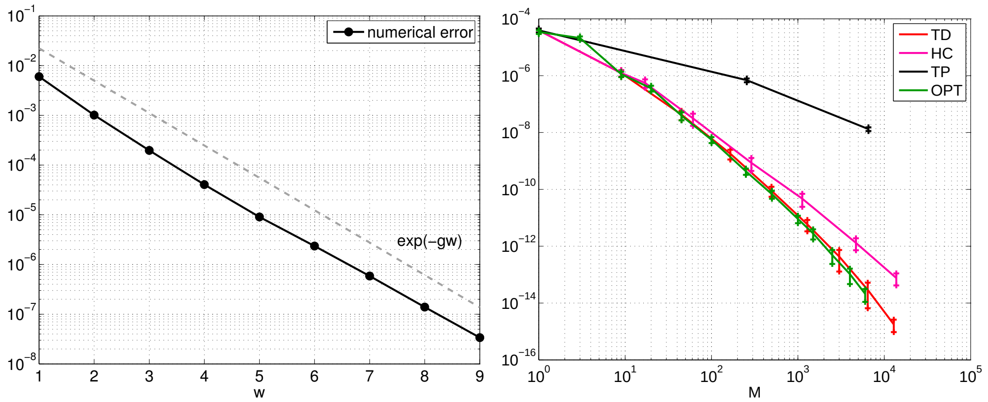

Figure 7: Left: convergence of the error (31) squared with respect to w for the quasi-optimal TD Galerkin approximation. Right: Convergence of the error (31) squared in terms of the dimension of the polynomial space, for the TD approximation, as well as Tensor Product (TP), Hyperbolic cross (HC) and best M-terms (OPT) approximations.

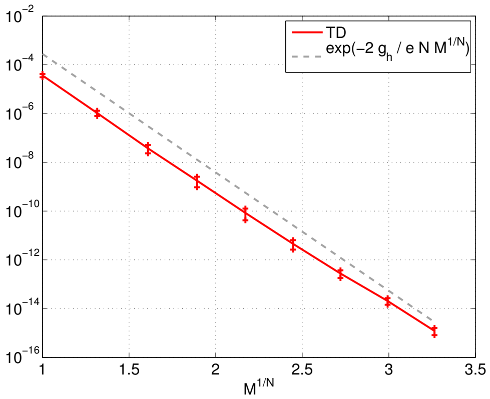

Figure 8: Convergence of the error (31) squared for the TD approximation with respect to $\sqrt[N]{M}$, $N = 8$, compared with the simplified theoretical estimate (30).

(TD), Hyperbolic Cross (HC) and Tensor Product (TP) spaces) as well as an estimate of the optimal convergence for the Galerkin method. The latter has been estimated by rearraging in decreasing order the coefficients of the Galerkin solution $TD(9)$ and using again a Monte Carlo estimate for the L^2 error, as in equation (32). Since the convergences have been estimated using a Monte Carlo sampling, we also provide in the plot uncertainty bars corresponding to ± 3 standard deviations of the Monte Carlo estimator. As already observed in [4], the TD approximation is the most efficient approximation scheme for the problem of interest, and now can be also understood as the quasi-optimal approximation, as indeed its convergence curve is very close to the best M-terms convergence.

Finally, Figure 8 shows that the theoretical convergence estimates for the error of the TD approximation appears to be quite sharp, even in its simplified form (30) and appeearently without any constraint on M. In particular, observe that the value of g_h used here is 1.9, i.e. it has been computed by fitting the Legendre coefficients (and pretending a perfect isotropy) and not by fitting the error convergence itself (as it was done for Figure 7 instead). For large values of M however, such simplified estimate seems to be too optimistic. Yet, one should also consider that the convergence curve may be sligthly miscalculated, due to the Monte Carlo approximation of the L^2 error, and to the fact that the Legendre coefficients computed are not exact, but rather approximated by a “overkilling” Galerkin procedure.

6.2 Anisotropic problem

The second test we consider is an anisotropic problem with 4 random variables uniformly distributed in $[-0.99, 0]$, acting on the inclusions illustrated in Figure

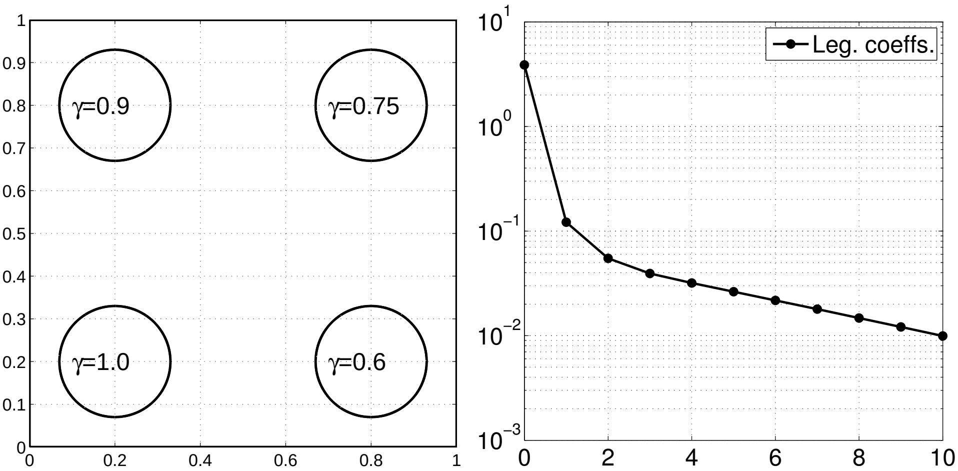

Figure 9: Left: physical domain for the anisotropic inclusions problem. The numbers inside each inclusion are the corresponding values γ_n. Right: decay of the Legendre coefficients for $\mathbf{q} = [q\,0\,0\,0]$, $0 \leq q \leq 10$ in semilog scale. A preasymptotic non-exponential regime is clearly present for $q \leq 2$.

9, located at the corners of the domain. The anisotropy is given by the coefficients γ_n in the expression (26) of the diffusion coefficient, that have been chosen as detailed in Figure 9.

In contrast with the isotropic setting just analyzed, here the forcing term and the quantity of interest $\Theta(u)$ are now defined over the whole domain rather than on the smaller area F. Finally, the reference solution is now an isotropic TD Stochastic Galerkin approximation at level $\mathrm{w} = 22$, and the L^2 approximation error is computed with $M = 3000$ Monte Carlo samples.

Compared to the previous case, in this setting the exponential bound on the decay of the Legendre coefficients is not sharp, as a slower preasymptotic regime appears, see Figure 9-right: in turn this implies that the anisotropic TD sets will not be a tight estimate of the best M-terms approximation, see Figure 10-left. However, using the numerical procedure described in [4, 6] is it possible to compute some "effective" exponential rates that yield to anisotropic TD sets with good convergence properties, cf. again Figure 10-left.

The numerical value of such effective rates is approximately $\widehat{\mathbf{g}} = (0.4,\ 1.37,\ 2.27,\ 3.17)$. Observe that we could also have determined $\widehat{\mathbf{g}}$ by formula (29) in Proposition 24. This would have resulted in $\widehat{\mathbf{g}} \approx (0.20,\ 0.68,\ 1.12,\ 1.51)$, that is roughly half the numerically assessed rates. This is a further confirmation that the theoretical estimates, although not sharp, give a good ansatz to the qualitative features of the problem. Incidentally, note that for the purpose of building a sequence of TD sets what really matters is not the absolute value of $\widehat{\mathbf{g}}$, rather the ratio between the rates, the absolute value being important only in the estimate of

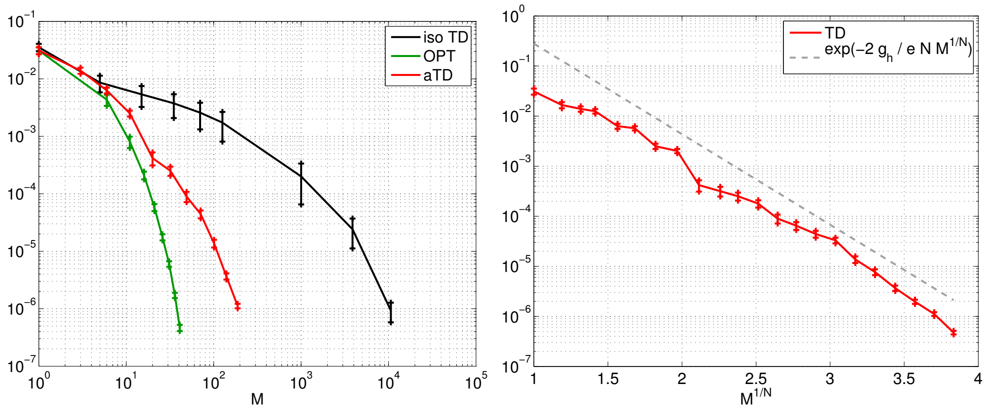

Figure 10: Left: convergence of isotropic and anisotropic TD sets, compared to the convergence of the best M-terms approximation error. Right: Convergence of the L^2_ϱ error of the anisotropic TD approximation with respect to $\sqrt[N]{M}$, compared with the simplified theoretical estimate (30).

the convergence rate.

Finally, figure 10-right shows that also in this case the simplified estimate (30) on the convergence of the anisotropic TD set seems to be quite sharp and to hold without restrictions on the cardinality M of the approximation.

7 Conclusions

In this work we have analyzed the approximability of solution of linear elliptic PDEs with stochastic coefficients that are analytic in a polydisc in the complex domain. Although somehow restrictive, this hypothesis is satistified by a number of problems that arise in various engineering fields, as briefly illustrated in Remark 3. This setting has allowed us to use in a very natural way Bernstein ellipses to estimate of the decay of the Legendre coefficients, as recalled in Proposition 7, and consequently to prove that total degree polynomial spaces represent a quasi-optimal approximation of the best M-terms polynomial approximation. We have then proved with two different arguments the subexponential convergence of the Galerkin approximation of u in such polynomial spaces, see Theorems 11 and 16.

We have verified both the estimate of the decay of the Legendre coefficients and that of the error convergence on two numerical tests, re-examining the results we had obtained in the previous work [4]. The results obtained allow us to claim that the theoretical estimates provided in this work are in essence correct, in the sense that they provide valid ansatzs to be fitted with numerical “a posteriori” information, i.e. with a view to a combined “a-priori”/“a-posteriori” approach, as already explored in [4, 6].

References

[1] I. Babuška, R. Tempone, and G. E. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM J. Numer. Anal.*, 42(2):800–825, 2004.

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

[3] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Review*, 52(2):317–355, June 2010.

[4] J. Bäck, F. Nobile, L. Tamellini, and R. Tempone. Stochastic spectral Galerkin and collocation methods for PDEs with random coefficients: a numerical comparison. In J.S. Hesthaven and E.M. Ronquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*, pages 43–62. Springer, 2011. Selected papers from the ICOSAHOM '09 conference, June 22-26, Trondheim, Norway.

[5] T. Bagby, L. Bos, and N. Levenberg. Multivariate simultaneous approximation. *Constr. Approx.*, 18(4):569–577, 2002.

[6] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Mod. Methods Appl. Sci. (M3AS)*, 22(9), 2012.

[7] M. Bieri, R. Andreev, and C. Schwab. Sparse tensor discretization of elliptic SPDEs. *SIAM J. Sci. Comput.*, 31(6):4281–4304, 2009/10.

[8] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab. Sparse adaptive taylor approximation algorithms for parametric and stochastic elliptic pdes. SAM-Report 2011-44, Seminar für Angewandte Mathematik, ETH, Zurich, 2011.

[9] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best n-term Galerkin approximations for a class of elliptic sPDEs. *Foundations of Computational Mathematics*, 10:615–646, 2010. 10.1007/s10208-010-9072-2.

[10] A. Cohen, R. Devore, and C. Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[11] P.J. Davis. *Interpolation and approximation.* Dover Publications Inc., New York, 1975. Republication, with minor corrections, of the 1963 original, with a new preface and bibliography.

[12] Ronald A. DeVore. Nonlinear approximation. In *Acta numerica, 1998*, volume 7 of *Acta Numer.*, pages 51–150. Cambridge Univ. Press, Cambridge, 1998.

[13] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, 46(02):317–339, 2012.

[14] B. Ganapathysubramanian and N. Zabaras. Sparse grid collocation schemes for stochastic natural convection problems. *Journal of Computational Physics*, 225(1):652–685, 2007.

[15] W. Gautschi. *Orthogonal Polynomials: Computation and Approximation.* Oxford University Press, Oxford, 2004.

[16] R. G. Ghanem and P. D. Spanos. *Stochastic finite elements: a spectral approach.* Springer-Verlag, New York, 1991.

[17] C.J. Gittelson. Uniformely convergent adaptive methods for parametric operator equations. SAM-Report 2011-19, Seminar für Angewandte Mathematik, ETH, Zurich, 2011.

[18] O. P. Le Maître and O. M. Knio. *Spectral methods for uncertainty quantification.* Scientific Computation. Springer, New York, 2010. With applications to computational fluid dynamics.

[19] H. G. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Comput. Methods Appl. Mech. Engrg.*, 194(12-16):1295–1331, 2005.

[20] R. B. Nelsen. *An introduction to copulas.* Springer Series in Statistics. Springer, New York, second edition, 2006.

[21] F. Nobile and R. Tempone. Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients. *Internat. J. Numer. Methods Engrg.*, 80(6-7):979–1006, 2009.

[22] F. Nobile, R. Tempone, and C.G. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2309–2345, 2008.

[23] M. F. Pellissetti and R. G. Ghanem. Iterative solution of systems of linear equations arising in the context of stochastic finite elements. *Adv. Eng. Software*, 31:607–616, 2000.

[24] G. Stefanou. The stochastic finite element method: past, present and future. *Comput. Methods Appl. Mech. Engrg.*, 198:1031–1051, 2009.

[25] R. A. Todor and C. Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J Numer Anal*, 27(2):232–261, 2007.

[26] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM J. Sci. Comput.*, 27(3):1118–1139, 2005.

[27] D. Xiu and G.E. Karniadakis. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.*, 24(2):619–644, 2002.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

30/2012 Beck, J.; Nobile, F.; Tamellini, L.; Tempone, R.;
Convergence of quasi-optimal Stochastic Galerkin Methods for a class of PDES with random coefficients

29/2012 Chen, P.; Quarteroni, A.; Rozza, G.
Stochastic Optimal Robin Boundary Control Problems of Advection-Dominated Elliptic Equations

28/2012 Canuto, C.; Nochetto, R.H.; Verani, M.
Contraction and optimality properties of adaptive Legendre-Galerkin methods: the 1-dimensional case

27/2012 Pigoli, D.; Secchi,P.
Estimation of the mean for spatially dependent data belonging to a Riemannian manifold

26/2012 Tamellini, L.; Le Maitre, O.; Nouy, A.
Model reduction based on Proper Generalized Decomposition for the Stochastic steady incompressible Navier-Stokes equations

25/2012 Manfredini, F.; Pucci, P.; Secchi, P.; Tagliolato, P.; Vantini, S.; Vitelli, V.
Treelet decomposition of mobile phone data for deriving city usage and mobility pattern in the Milan urban region

24/2012 Antonietti, P.F.; Giani, S.; Houston, P.
hp Version Composite Discontinuous Galerkin Methods for Elliptic Problems on Complicated Domains

23/2012 Fabio Nobile, Christian Vergara
Partitioned algorithms for fluid-structure interaction problems in haemodynamics

22/2012 Ettinger, B.; Passerini, T.;Perotto, S.; Sangalli, L.M.
Regression models for data distributed over non-planar domains

21/2012 Guglielmi, A.; Ieva, F.; Paganoni, A.M.; Ruggeri, F.; Soriano, J.
Semiparametric Bayesian models for clustering and classification in presence of unbalanced in-hospital survival