ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

MATHICSE

Institute of Mathematics

School of Basic Sciences

MATHICSE Technical Report

Nr. 24.2017

November 2017

A posteriori error estimation for the stochastic collocation finite element method

Diane Guignard, Fabio Nobile

A posteriori error estimation for the stochastic collocation finite element method

Diane Guignard[1] and Fabio Nobile[1]

[1] Ecole Polytechnique Fédérale Lausanne, SB-MATH, Station 8, 1015 Lausanne, Switzerland

November 3, 2017

Abstract

In this work, we consider an elliptic partial differential equation with a random coefficient solved with the stochastic collocation finite element method. The random diffusion coefficient is assumed to depend in an affine way on independent random variables. We derive a residual-based *a posteriori* error estimate that is constituted of two parts controlling the stochastic collocation (SC) and the finite element (FE) errors, respectively. The SC error estimator is then used to drive an adaptive sparse grid algorithm. Several numerical examples are given to illustrate the efficiency of the error estimator and the performance of the adaptive algorithm.

1 Introduction

Partial differential equations (PDEs for short) are the mathematical formulation of many physical and engineering phenomena. For such problems, the input data are often affected by uncertainty, either due to a lack of knowledge or to an inherent variability of the system. Probability theory offers a possible way to describe the uncertainties, characterizing the uncertain input data with random variables or random fields and yielding PDEs with random inputs.

The development of efficient methods to tackle the numerical approximation of such problems has thus been of great interest and has attracted the attention of many scientists over the past decades. In this work, we will consider the stochastic collocation (SC) method [1–3] for the stochastic approximation and the finite element (FE) method for the physical space discretization. As sampling methods of Monte-Carlo or Quasi and multilevel Monte-Carlo type [4–7], and contrary to *intrusive* methods like stochastic Galerkin [8, 9], the stochastic collocation method requires only the solution of decoupled deterministic problems and thus allows the re-use of deterministic solvers. Moreover, exploiting the possible regularity of the solution with respect to the random parameters, the stochastic collocation method has the advantage to have a potentially much faster convergence rate than the Monte-Carlo method. It is also suitable for large uncertainties, contrary to perturbation type methods as considered in our previous works [10, 11].

Whenever a numerical method is used to approximate the solution of the problem under consideration, an error analysis should be performed to estimate the numerical error thus introduced. The derivation of *a priori* error estimates for the stochastic collocation finite element method is done e.g. in [1, 12, 13] but, to our knowledge, no *a posteriori* error estimate for the whole solution in suitable norms has been derived yet. It is of great importance to have *a posteriori* error estimators at

disposal since such estimators are the foundation of many adaptive strategies which aim at reaching a numerical solution with prescribed accuracy while keeping the computational cost as low as possible. Here, the numerical solution is affected by two sources of error, namely the SC and the FE errors, and the estimator should not only provide an upper bound of the error but also furnish an estimation of the contribution of each error component to the total error, so that it can be used for balancing errors in an adaptive algorithm. We mention that recently, *a posteriori* error estimates for a specific quantity of interest, usually referred to as *goal-oriented* error estimates, have been developed, see for instance [14].

The main drawback of the stochastic collocation method is that is suffer from the so-called *curse of dimensionality* when tensor grids are used, namely the performance of the method deteriorates as the number of random variables increases. A remedy is then to exploit the possible anisotropy of the solution, in the sense that the different random variables might not have the same influence on the solution. Examples of works in this direction are the anisotropic sparse grid method proposed in [15] or the quasi-optimal sparse grids method introduced in [12]. In the latter, the adaptive algorithm is based on *a priori* error estimates whose constants are numerically tuned during the process, yielding what the authors called an *a priori/a posteriori* strategy. A proof of convergence has been obtained in [16] for the pure *a priori* algorithm. An *a posteriori* sparse grid adaptive algorithm has first been proposed in [17] and then used for instance in [13, 18–22]. In [20], the adaptive process is driven by profit indicators obtained by solving additional PDEs. The method is applicable to a wide range of problems, including for instance the case of unbounded random variables or non-nested grids and can be combined with a Monte Carlo sampling, using a control variate technique, to handle rough random field [23]. However, the error indicators proposed so far are heuristic and do not provide a certified control of the error.

We mention that adaptive strategies have also been investigated when a different method is used for the stochastic space approximation. For instance in [24], the solution is approximated via a Taylor series and an adaptive algorithm is proposed with a proof of its convergence. In [25, 26], where the random PDEs are solved with the Stochastic Galerkin FEM, the convergence is proved when the adaptation is performed in both physical and stochastic spaces. In this case, the extension of the results obtained for the AFEM in [27] is feasible and strongly uses the so-called Galerkin orthogonality property. So far, at least to our knowledge, there is no proof of convergence for adaptive stochastic collocation methods.

The main goal of this paper is to derive an *a posteriori* error estimate that controls both the FE and the SC errors. We consider an elliptic diffusion problem with random coefficient that depends in an affine way on a finite number of independent random variables. Moreover, we restrict to the case where the source term is deterministic and the stochastic collocation scheme is interpolatory. The error estimate we obtain is residual-based, provides an upper bound of the total error and is localizable, hence suitable for adaptive algorithms. We use then the SC error estimator to drive an adaptive sparse grid algorithm in which the collocation points are iteratively selected based on a criteria that uses the error estimator. It is important to mention that so far, we have no proof of convergence of the adaptive algorithm proposed here. Moreover, this algorithm is only suitable for random spaces of moderate dimension (or if the anisotropy in the problem is significant). An alternative procedure should be used for high-dimensional problems, adapting for instance the *dimension adaptive* strategy proposed in [20] to our context. We stress that in this work we have focused only on adaptive strategies in the stochastic dimension and selected, in our numerical experiments, a sufficiently fine spatial mesh. The next step would be to propose an adaptive strategy with refinements in both the physical and random spaces, combining for instance the algorithm proposed here for the

adaptive selection of the collocation points with a standard AFEM for the physical mesh refinement.

The outline of the paper is the following. We give in section 2 the statement of the problem, namely an elliptic diffusion PDE with random coefficient. We present in section 3 the stochastic collocation finite element method we use to solve this problem approximatively. The section 4 is devoted to the *a posteriori* error analysis, more precisely to the derivation of a residual-based *a posteriori* error estimate that control the two error components. We give in section 5 a possible strategy to adaptively construct the sparse grid using the stochastic error estimator. We perform several numerical experiments in section 6 to test the efficiency of the error estimator and the performance of the proposed adaptive strategy. Finally, section 7 contains some insight on how to deal with high-dimensional problems and some conclusions are presented in section 8.

2 Problem statement

Let $D \subset \mathbb{R}^d$ be an open bounded domain with Lipschitz continuous boundary ∂D and let $(\Omega, \mathcal{F}, P)$ be a complete probability space. We seek for $u : \bar{D} \times \Omega \to \mathbb{R}$ that solves P-almost everywhere in Ω, or in other words almost surely (a.s.),

$$\left\{ \begin{array}{rcll} -\nabla \cdot (a(\cdot,\omega)\nabla u(\cdot,\omega)) & = & f(\cdot) & \text{in } D \\ u(\cdot,\omega) & = & 0 & \text{on } \partial D \end{array} \right. \tag{1}$$

with deterministic forcing term $f \in L^2(D)$ and random field a on $(\Omega, \mathcal{F}, P)$ over $W^{1,\infty}(D)$. Moreover, we make the following assumptions on the random diffusion coefficient a:

$$\text{there exist } a_{min}, a_{max} : \; P(\omega \in \Omega : \; 0 < a_{min} \leq a(\mathbf{x},\omega) \leq a_{max} < \infty \; \forall \mathbf{x} \in D) \tag{2}$$

and

$$a(\mathbf{x},\omega) = a_0(\mathbf{x}) + \sum_{n=1}^{N} a_n(\mathbf{x}) Y_n(\omega), \tag{3}$$

where $(Y_n)_{n=1}^N$ are real-valued independent random variables. Thanks to the Doob-Dynkin Lemma, the solution u depends on the same random variables as the diffusion coefficient a, i.e. we have $u(\mathbf{x},\omega) = u(\mathbf{x}, Y_1(\omega), \ldots, Y_N(\omega))$. Let us introduce $\Gamma = \Gamma_1 \times \ldots \times \Gamma_N$ with $\Gamma_n = Y_n(\Omega)$ for $n = 1, \ldots, N$. Moreover, let $\rho : \Gamma \to \mathbb{R}_+$ be the joint probability density function of the random vector $\mathbf{Y} = (Y_1, \ldots, Y_N)$, which factorizes as $\rho(\mathbf{y}) = \Pi_{n=1}^N \rho_n(y_n)$ for all $\mathbf{y} = (y_1, \ldots, y_N) \in \Gamma$. We can then replace the probability space $(\Omega, \mathcal{F}, P)$ by $(\Gamma, B(\Gamma), \rho(\mathbf{y})d\mathbf{y})$, where $B(\Gamma)$ denotes the Borel σ-algebra defined on Γ and $\rho(\mathbf{y})d\mathbf{y}$ the probability measure of $\mathbf{Y}$. Finally, for a given Banach space V with norm $\|\cdot\|_V$ and for $p \in [1, \infty]$ we define the Bochner space

$$L^p_\rho(\Gamma; V) := \{v : \Gamma \to V \,|\, v \text{ is strongly measurable and } \|v\|_{L^p_\rho(\Gamma;V)} < \infty\}$$

with

$$\|v\|_{L^p_\rho(\Gamma;V)} := \left\{ \begin{array}{cl} \left(\int_\Gamma \|v(\mathbf{y})\|_V^p \rho(\mathbf{y}) d\mathbf{y}\right)^{\frac{1}{p}} & \text{if } p < \infty \\ \rho - \text{ess sup}_{\mathbf{y}\in\Gamma} \|v(\mathbf{y})\|_V & \text{if } p = \infty. \end{array} \right.$$

The (parametric, pointwise) weak formulation of problem (1) reads: find $u : \Gamma \to H_0^1(D)$ such that

$$\int_D a(\mathbf{x},\mathbf{y}) \nabla u(\mathbf{x},\mathbf{y}) \cdot \nabla v(\mathbf{x}) d\mathbf{x} = \int_D f(\mathbf{x}) v(\mathbf{x}) d\mathbf{x} \quad \forall v \in H_0^1(D), \; \rho\text{-a.e. in } \Gamma, \tag{4}$$

where $H_0^1(D)$ is the usual Sobolev space that we endow with the gradient norm $\|v\|_{H_0^1(D)} = \|\nabla v\|_{L^2(D)}$. By a straightforward application of Lax-Milgram's lemma, assumption (2) ensures the well-posedness of problem (4), namely that there exists a unique solution $u \in L^2_\rho(\Gamma; V)$, with $V = H_0^1(D)$, which satisfies the *a priori* estimate

$$\|u\|_{L^2_\rho(\Gamma;V)} \leq \frac{C_P}{a_{min}} \|f\|_{L^2(D)}.$$

In particular, we have $u \in L^p_\rho(\Gamma; V)$ for any $p \in [1, \infty]$. Moreover, it has been shown (see for instance [1]) that the parametric solution u of problem (4) is analytic with respect to each parameter $y_n \in \Gamma_n$, $n = 1, \ldots, N$. Finally, we mention that imposing $a(\cdot, \omega) \in L^\infty(D)$ is enough for the well-posedness of the problem. We assume $W^{1,\infty}(D)$ regularity for ease of derivation of our *a posteriori* error estimate, see (16) below.

3 Stochastic collocation finite element method

In this section, we briefly present the stochastic collocation finite element method (SC-FEM for short) for solving numerically PDEs with random input data, following closely [16] and focusing on the model problem (1). We also refer to [1, 3] for a complete discussion on this method. The idea is to proceed in two steps: first a semi-discretization of problem (4) using the FEM for the physical space approximation and then the application of a collocation method for the stochastic space approximation using global polynomials in $\mathbf{y}$. We thus seek for an approximate solution in a space $\mathbb{P}(\Gamma) \otimes V_h$, with $\mathbb{P}(\Gamma) \subset L^2_\rho(\Gamma)$ a polynomial space on Γ and V_h a FE subspace of V.

More precisely, for any $h > 0$, let $\mathcal{T}_h$ be a regular triangulation of D with elements T of diameter $h_T \leq h$. We assume that there exists a constant $c > 0$ satisfying

$$\frac{h_T}{\rho_T} \leq c \quad \forall T \in \mathcal{T}_h, \forall h > 0 \tag{5}$$

where $\rho_T = \sup\{diam(B) : B \text{ is a ball contained in } T\}$. Let $V_h \subset V$, with $\dim(V_h) = N_h$, be the space of continuous, piecewise linear finite element functions associated to $\mathcal{T}_h$ that vanish on ∂D. The semi-discretized problem is therefore given by: find $u_h : \Gamma \to V_h$ such that

$$\int_D a(\mathbf{x}, \mathbf{y}) \nabla u_h(\mathbf{x}, \mathbf{y}) \cdot \nabla v_h(\mathbf{x}) d\mathbf{x} = \int_D f(\mathbf{x}) v_h(\mathbf{x}) d\mathbf{x} \quad \forall v_h \in V_h, \rho\text{-a.e. in } \Gamma. \tag{6}$$

The problem (6) is then further discretized by considering a set $\{\mathbf{y}_1, \ldots, \mathbf{y}_{N_c}\}$ of N_c collocation points in Γ and building the global polynomial approximation

$$u_{h,N_c}(\mathbf{y}) = \sum_{k=1}^{N_c} u_h(\mathbf{y}_k) L_k(\mathbf{y}) \tag{7}$$

for appropriate multivariate (for instance Lagrange) polynomials L_k, where $u_h(\mathbf{y}_k)$ is the solution of problem (6) with $\mathbf{y} = \mathbf{y}_k$. A possible choice for the collocation points $\mathbf{y}_k \in \Gamma$ is to take the Cartesian product of certain abscissas in each direction. However, using such tensor grid would rapidly become computationally unaffordable due to the *curse of dimensionality*: the number of nodes increases exponentially with N. To alleviate this drawback, the idea is to use a so-called *sparse grid*, first introduced by Smolyak in [28]. Let us define

$$\mathcal{U}_n^{m(i_n)} : C^0(\Gamma_n) \to \mathbb{P}_{m(i_n)-1}(\Gamma_n) \tag{8}$$

a sequence of univariate polynomial interpolant operators along each direction Γ_n for $n = 1, \dots, N$, using abscissas $\{\xi_j^{n,i_n}\}_{j=1}^{m(i_n)}$. Here, $m(i_n)$ denotes the number of collocation points used to build the interpolant of level i_n and $\mathbb{P}_q(\Gamma_n)$ is the space of polynomials in y_n of degree at most q. The function m should satisfy $m(0) = 0$, $m(1) = 1$ and $m(i) < m(i+1)$ for any $i \geq 1$. Moreover, let $I \subset \mathbb{N}_+^N$ be a multi-index set, where $\mathbb{N}_+ = \{1, 2, \dots\}$ denotes the positive integers. In what follows, the only restriction on I will be that it is a downward closed set (a.k.a. lower set), i.e. it satisfies

$$\forall \mathbf{i} \in I, \quad \mathbf{i} - \mathbf{e}_j \in I \quad \forall j = 1, \dots, N \text{ such that } i_j > 1. \tag{9}$$

This condition is necessary to get good approximation properties, see for instance [17]. Setting $\mathcal{U}_n^0 = 0$ for $n = 1, \dots, N$, we define then the sparse grid interpolant S_I by

$$u_{h,I}(\mathbf{y}) = S_I[u_h](\mathbf{y}) = \sum_{\mathbf{i} \in I} \Delta^{\mathbf{m(i)}}(u_h)(\mathbf{y}) \tag{10}$$

where

$$\Delta^{\mathbf{m(i)}} = \bigotimes_{n=1}^{N} \Delta_n^{m(i_n)} = \bigotimes_{n=1}^{N} \left(\mathcal{U}_n^{m(i_n)} - \mathcal{U}_n^{m(i_n-1)} \right)$$

and $\mathbf{m(i)} = (m(i_1), \dots, m(i_N))$. The operators $\Delta_n^{m(i_n)}$ and $\Delta^{\mathbf{m(i)}}$ are often referred to as *difference* (or *detail*) and *hierarchical surplus* operators, respectively. In what follows, we assume that

$$u_h(\mathbf{y}) = \sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta^{\mathbf{m(i)}}(u_h)(\mathbf{y}) \quad \rho\text{-a.e. in } \Gamma, \tag{11}$$

where the series converges absolutely in V, which holds if u is sufficiently smooth in $\mathbf{y}$ and if the operators $\mathcal{U}_n^{m(i_n)}$ in (8) are such that $\bigotimes_{n=1}^N \mathcal{U}_n^{m(i_n)} u \to u$ in V as $\mathbf{i} \to \infty$. Finally, we mention that the operator S_I in (10) can be equivalently written as a linear combination of tensor grid interpolations, see for instance [29], as

$$S_I[u_h](\mathbf{y}) = \sum_{\mathbf{i} \in I} c_{\mathbf{i}} \bigotimes_{n=1}^{N} \mathcal{U}_n^{m(i_n)}(u_h)(\mathbf{y}), \quad c_{\mathbf{i}} = \sum_{\substack{\mathbf{j} \in \{0,1\}^N \\ (\mathbf{i}+\mathbf{j}) \in I}} (-1)^{|\mathbf{j}|} \tag{12}$$

with $|\mathbf{j}| = \sum_{n=1}^N j_n$ for $\mathbf{j} = (j_1, ..., j_N)$. Notice that many of the coefficients $c_{\mathbf{i}}$ are actually zero: for instance, for $\mathbf{i} \in I$, $c_{\mathbf{i}} = 0$ if $(\mathbf{i}+\mathbf{j}) \in I$ for all $\mathbf{j} \in \{0,1\}^N$. We then call *sparse grid* the set of N_c collocation points needed by (12) to compute $S_I[u_h]$. To summarize, the sparse grid interpolant S_I is characterized by the multi-index set I, the function m defining the number of collocation points on each level and the type of univariate nodes. Remark that I must contain the multi-index $\mathbf{1}$, which allows to approximate constant functions.

Our error estimate will only be valid in the case S_I is interpolatory, i.e. it satisfies $S_I[f](\mathbf{y}_k) = f(\mathbf{y}_k)$ for $k = 1, \dots, N_c$ where $\{\mathbf{y}_1, \dots, \mathbf{y}_{N_c}\}$ are the collocation points in the sparse grid underlying the multi-index set I and function m. Notice that such property requires the use of nested sequencies of univariate nodes $\{\xi_j^{n,i_n-1}\} \subset \{\xi_j^{n,i_n}\}$, see for instance [30, p.277]. Finally, we introduce the notion of margin M_I and reduced margin, defined respectively by

$$\begin{aligned} M_I &= \{\mathbf{i} \in \mathbb{N}_+^N \setminus I : \mathbf{i} - \mathbf{e}_n \in I \text{ for some } n \in \{1, \dots, N\}\} \\ R_I &= \{\mathbf{i} \in M_I : \mathbf{i} - \mathbf{e}_n \in I \text{ for all } n = 1, \dots, N \text{ with } i_n > 1\}. \end{aligned}$$

4 Residual-based *a posteriori* error estimate

We will now derive an *a posteriori* error estimate for the error $u - S_I[u_h]$ which consists of two parts controlling the finite element and stochastic collocation errors, respectively. We first give two results that we will use in the derivation of the error estimate.

Proposition 4.1. *Let S_I be the operator defined in (10). Then for any $f, g \in C^0(\Gamma)$ we have*

$$S_I[fg] = S_I[fS_I[g]].$$

Proof. Since S_I is assumed to be interpolatory, we have $S_I[g](\mathbf{y}_k) = g(\mathbf{y}_k)$ for all $k = 1, \ldots, N_c$. By the definition of S_I, we get then for any $\mathbf{y} \in \Gamma$

$$\begin{aligned} S_I[fS_I[g]](\mathbf{y}) &= \sum_{k=1}^{N_c} \left(fS_I[g]\right)(\mathbf{y}_k) L_k(\mathbf{y}) = \sum_{k=1}^{N_c} f(\mathbf{y}_k) S_I[g](\mathbf{y}_k) L_k(\mathbf{y}) \\ &= \sum_{k=1}^{N_c} f(\mathbf{y}_k) g(\mathbf{y}_k) L_k(\mathbf{y}) = S_I[fg](\mathbf{y}). \end{aligned}$$

□

For any downward closed multi-index set I, let us define the polynomial space $\mathbb{P}_I$ by

$$\mathbb{P}_I = \sum_{\mathbf{i} \in I} \mathbb{P}_{\mathbf{m(i)-1}} \quad \text{with} \quad \mathbb{P}_{\mathbf{m(i)-1}} = \mathbb{P}_{m(i_1)-1} \otimes \ldots \otimes \mathbb{P}_{m(i_N)-1}. \tag{13}$$

We have the following approximation properties that will be crucial in the derivation of our error estimate.

Proposition 4.2. *Let S_I be the operator defined if (10). Then*

1. $S_I[f] \in \mathbb{P}_I \quad \forall f \in C^0(\Gamma)$
2. *S_I is exact on $\mathbb{P}_I$, i.e.* $S_I[f] = f \quad \forall f \in \mathbb{P}_I$.

Proof. See Proposition 1 in [31]. □

Finally, we introduce the (generalized) jump of a function φ across an edge ($d = 2$) or a face ($d = 3$) e in the direction $\mathbf{n}_e$ orthogonal to e by

$$[\varphi]_{\mathbf{n}_e}(\mathbf{x}) := \begin{cases} \lim_{t \to 0^+} \left(\varphi(\mathbf{x} + t\mathbf{n}_e) - \varphi(\mathbf{x} - t\mathbf{n}_e)\right) & \text{if } e \not\subset \partial D \\ 0 & \text{if } e \subset \partial D. \end{cases}$$

We can now state our residual-based *a posteriori* error estimate.

Proposition 4.3. *Let u and u_h be the solutions of (4) and (6), respectively and let $S_I[u_h]$ be the sparse grid approximation of u_h computed using the multi-index set I. There exists a constant $C > 0$ depending only on the mesh aspect ratio c such that for any $p \in [1, \infty]$ we have*

$$\|u - S_I[u_h]\|_{L^p_\rho(\Gamma;V)} \leq \frac{1}{a_{min}} \left[C\eta_{FE} + \zeta_{SC}\right], \tag{14}$$

where

$$\eta_{FE} = \sum_{k=1}^{N_c} \eta_k \|L_k\|_{L^p_\rho(\Gamma)}, \quad \eta_k := \left(\sum_{T \in \mathcal{T}_h} \eta_{k,T}^2\right)^{\frac{1}{2}} \tag{15}$$

with

$$\eta_{k,T} := h_T^2 \|f + \nabla \cdot (a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k))\|^2_{L^2(T)} + \sum_{e \subset \partial T} h_e \|\frac{1}{2}[a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k) \cdot \mathbf{n}_e]_{\mathbf{n}_e}\|^2_{L^2(e)} \tag{16}$$

and

$$\zeta_{SC} = \sum_{\mathbf{i} \in M_I} \zeta_{\mathbf{i}}, \quad \zeta_{\mathbf{i}} := \|\Delta^{\mathbf{m(i)}} \left(a \nabla S_I[u_h]\right)\|_{L^p_\rho(\Gamma; L^2(D))}. \tag{17}$$

Proof. In what follows, all equations hold ρ-a.e. in Γ without specifically mentioning it. Moreover, the dependence of each function on variables will not necessarily be indicated, unless ambiguity arises. For any $v \in V$ we have

$$\begin{aligned} \int_D a\nabla(u - S_I[u_h]) \cdot \nabla v &= \int_D fv - \int_D a \nabla S_I[u_h] \cdot \nabla v \\ &= \underbrace{S_I\left[\int_D fv - \int_D a\nabla u_h \cdot \nabla v\right]}_{=:A_1} \\ &\quad + \underbrace{S_I\left[\int_D a\nabla u_h \cdot \nabla v\right] - \int_D a \nabla S_I[u_h] \cdot \nabla v}_{=:A_2}. \end{aligned} \tag{18}$$

For the second equality, we have used that f is deterministic and thus $S_I[f] = f$ for any multi-index set I. We analyse the terms A_1 and A_2 separately. For the first term, thanks to the Galerkin orthogonality we have

$$\begin{aligned} A_1 &= \sum_{k=1}^{N_c} \left[\int_D fv - \int_D a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k) \cdot \nabla v\right] L_k(\mathbf{y}) \\ &= \sum_{k=1}^{N_c} \left[\int_D f(v - v_h) - \int_D a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k) \cdot \nabla(v - v_h)\right] L_k(\mathbf{y}) \end{aligned} \tag{19}$$

for any $v_h \in V_h$. We take $v_h = I_h v$ the Clément interpolant of v for which we have the following interpolation error bounds [32]

$$\|v - I_h v\|_{L^2(T)} \leq C h_T \|\nabla v\|_{L^2(N(T))} \quad \text{and} \quad \|v - I_h v\|_{L^2(e)} \leq C h_e^{\frac{1}{2}} \|\nabla v\|_{L^2(N(T_e))} \tag{20}$$

for any element T and any edge or face e. Here, for an internal edge or face e, T_e is the union of the two elements sharing e. Moreover, $N(T)$ denotes the patch of elements associated to T, i.e. all $K \in \mathcal{T}_h$ with $\bar{K} \cap \bar{T} \neq \emptyset$ (the definition of $N(T_e)$ being analogous). After splitting the integral in (19) over each element T and integrating by part, we obtain

$$A_1 \leq C \sum_{k=1}^{N_c} |L_k(\mathbf{y})| \eta_k \|\nabla v\|_{L^2(D)} \tag{21}$$

with η_k defined in (15). Notice that this term η_k is *deterministic*, namely it does not depend on $\mathbf{y}$. It controls the FE error made when solving approximately the problem for the collocation point $\mathbf{y}_k$.

We now bound the second term A_2. We first notice that, thanks to Proposition 4.1, we have $S_I\left[a\nabla u_h\right] = S_I\left[a\nabla S_I[u_h]\right]$ since S_I is assumed to be interpolatory.

Therefore, using relation (11) we get

$$\begin{aligned} A_2 &= \int_D \left(S_I\left[a\nabla S_I[u_h]\right] - a\nabla S_I[u_h]\right)\cdot\nabla v = -\int_D \sum_{\mathbf{i}\notin I} \Delta^{\mathbf{m(i)}}(a\nabla S_I[u_h])\cdot\nabla v \\ &= -\int_D \sum_{\mathbf{i}\in M_I} \Delta^{\mathbf{m(i)}}(a\nabla S_I[u_h])\cdot\nabla v \\ &\leq \|\sum_{\mathbf{i}\in M_I} \Delta^{\mathbf{m(i)}}(a\nabla S_I[u_h])\|_{L^2(D)}\|\nabla v\|_{L^2(D)}. \end{aligned} \tag{22}$$

We have used the fact that a depends in an affine way on the random variables, see (2), to restrict the summation over the multi-indices of the margin M_I of I. Indeed, by Proposition 4.2 we have $S_I[u_h] \in \mathbb{P}_I$, with $\mathbb{P}_I$ defined in (13), and by assumption

$$a \in \mathbb{P}_{\mathbf{0}} + \sum_{n=1}^{N} \mathbb{P}_{\mathbf{e}_n}, \quad \text{with } \mathbb{P}_{\mathbf{e}_n} = \mathbb{P}_0 \otimes \ldots \otimes \mathbb{P}_0 \otimes \underbrace{\mathbb{P}_1}_{n^{th}index} \otimes \mathbb{P}_0 \ldots \otimes \mathbb{P}_0.$$

Therefore, we have $a\nabla S_I[u_h] \in \sum_{n=1}^{N}\sum_{\mathbf{i}\in I} \mathbb{P}_{\mathbf{m(i)}-\mathbf{1}+\mathbf{e}_n} \subset \mathbb{P}_{I\cup M_I}$ and thus

$$\Delta^{\mathbf{m(i)}}\left(a\nabla S_I[u_h]\right) = 0 \quad \forall\, \mathbf{i} \notin I \cup M_I \tag{23}$$

using again Proposition 4.2, namely that $S_{I\cup M_I}$ is exact on $\mathbb{P}_{I\cup M_I}$. Thanks to the uniform lower bound a_{min} on a, taking then $v = u(\mathbf{y}) - S_I[u_h](\mathbf{y})$ in (18) and using the bounds (21) and (22) for the terms A_1 and A_2, respectively, yields

$$\|\nabla(u(\mathbf{y})-S_I[u_h](\mathbf{y}))\|_{L^2(D)} \leq \frac{1}{a_{min}}\left(C\sum_{k=1}^{N_c}|L_k(\mathbf{y})|\eta_k + \|\sum_{\mathbf{i}\in M_I}\Delta^{\mathbf{m(i)}}(a\nabla S_I[u_h])(\mathbf{y})\|_{L^2(D)}\right). \tag{24}$$

To conclude the proof, it only remains to take the $L^p_\rho(\Gamma)$ norm on both sides of the last inequality and to use the triangle inequality for the norm $L^p_\rho(\Gamma; L^2(D))$ to take out the sum over the multi-indices $\mathbf{i} \in M_I$. □

Notice that in this proof, we have strongly used the fact that S_I is interpolatory and that a depends in an affine way on the random variables. The latter allows us to restrict the summation over all the multi-indices outside I in the bound of A_2 to the multi-indices belonging to the margin M_I, see (22). Moreover, it is worth mentioning that equation (24) yields a pointwise (in $\mathbf{y}$) error estimate.

Remark 4.4. *The spatial error estimate η_{FE} in (15) depends on $\|L_k(\mathbf{y})\|_{L^p_\rho(\Gamma)}$, $k = 1, \ldots, N_c$, i.e. on the stability constant of the operator S_I. These quantities can be bounded using the Lebesgue constant for S_I, whose growth depends on the choice of the function m and the family of interpolation points used by $\mathcal{U}_n^{m(i)}$, $n = 1, \ldots, N$. For instance, when using a doubling rule for m as in [30], defined by $m(1) = 1$ and $m(i) = 2^{i-1} + 1$ if $i > 1$, and Clenshaw-Curtis nodes, the Lebesgue constant associated with the operator S_I can be bounded by $|I|^2$ [33]. As an alternative, we could bound the term A_1 in (19) as follows*

$$\begin{aligned} A_1 &= \sum_{T\in\mathcal{T}_h}\left[\int_T \sum_{k=1}^{N_c} L_k(\mathbf{y})(f + \nabla\cdot(a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k)))(v - v_h) + \right. \\ &\qquad \left. \frac{1}{2}\sum_{e\subset\partial T}\int_e \sum_{k=1}^{N_c} L_k(\mathbf{y})[a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k)\cdot\mathbf{n}_e]_{\mathbf{n}_e}(v - v_h)\right] \\ &\leq C\left(\sum_{T\in\mathcal{T}_h}\eta_T^2\right)^{\frac{1}{2}}\|\nabla v\|_{L^2(D)} \end{aligned}$$

with

$$\eta_T(\mathbf{y})^2 := h_T^2 \| \sum_{k=1}^{N_c} L_k(\mathbf{y})(f + \nabla \cdot (a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k)))\|^2_{L^2(T)} + \sum_{e \subset \partial T} h_e \| \frac{1}{2} \sum_{k=1}^{N_c} L_k(\mathbf{y})[a(\mathbf{y}_k)\nabla u_h(\mathbf{y}_k) \cdot \mathbf{n}_e]_{\mathbf{n}_e} \|^2_{L^2(e)}. \quad (25)$$

Since $\left(\sum_{T\in\mathcal{T}_h} \eta_T^2\right)^{\frac{1}{2}} \leq \sum_{T\in\mathcal{T}_h} \eta_T$, *we can then replace (14) by*

$$\|u - S_I[u_h]\|_{L^p_\rho(\Gamma;V)} \leq \frac{1}{a_{min}} \left[C \sum_{T\in\mathcal{T}_h} \|\eta_T\|_{L^p_\rho(\Gamma)} + \zeta_{SC} \right]. \quad (26)$$

Mesh refinement, using the error estimate of Proposition 4.3 or the one proposed here, would lead to different adaptive strategies. The estimator in (15) gives an estimation of the spatial error for each collocation point, that is further localized on each element $T \in \mathcal{T}_h$. *Indeed, the estimator* $\eta_{k,T}$ *in (16) is an estimator of the FE error for the element* T *and the collocation point* $\mathbf{y}_k$. *Therefore, different spatial meshes could be considered for different collocation point. On the contrary, the estimator in (25) gives an estimation of the spatial error for each element* $T \in \mathcal{T}_h$ *and contains the contribution of all the collocation points. In this case, the same spatial mesh would then be used for all the collocation points.*

5 Adaptive algorithm

The error estimator deduced from Proposition 4.3 can be used to adaptively refine the mesh and increase the multi-index set. Such an adaptive strategy aims at reaching a given accuracy of the (FE and stochastic) error with computational cost as low as possible. Since the theory for mesh adaptation, often referred to as adaptive finite element method (AFEM), is well-developed and studied, we will focus on the stochastic collocation error. More precisely, we will consider an adaptive construction of the multi-index set I proceeding similarly to what has been originally proposed in [17] and further used for instance in [13, 20].

We give below a possible adaptive strategy which uses the error estimators $\zeta_{\mathbf{i}}$ given in (17) to drive the process, with the requirement that the multi-index set I must remain downward closed during the adaptation. Basically, at each iteration we select the multi-index in the margin M_I of the current set I that has the largest profit, the latter being defined as follows. For any $\mathbf{i} \in M_I$, we define

$$P_{\mathbf{i}} := \frac{\sum_{\mathbf{j}\in A_{\mathbf{i}}} \zeta_{\mathbf{j}}}{\sum_{\mathbf{j}\in A_{\mathbf{i}}} W_{\mathbf{j}}} \quad (27)$$

where $A_{\mathbf{i}} = J_{\mathbf{i}} \backslash I$ and $J_{\mathbf{i}}$ is the downward closed set of minimal cardinality containing $I \cup \{\mathbf{i}\}$, i.e. $A_{\mathbf{i}}$ is the set containing $\mathbf{i}$ plus all the multi-indices $\mathbf{j} \in M_I$ that must also be included in I if $\mathbf{i}$ is added to I so that the set remains downward closed. Moreover, we have denoted by $W_{\mathbf{i}}$ the *work* contribution of the multi-index $\mathbf{i}$, which can be defined as in [20] by

$$W_{\mathbf{i}} = \Pi_{n=1}^{N}(m(i_n) - m(i_n - 1)). \quad (28)$$

In the case of nested sets of points, as considered here, it corresponds to the number of new points in Γ introduced if $\mathbf{i}$ is added to I. We could also choose to set $W_{\mathbf{i}} = 1$ if we want to drive the adaptation only based on the error estimators. Finally, notice that for any $\mathbf{i} \in R_I$, since $I \cup \{\mathbf{i}\}$ is always downward closed, we have $A_{\mathbf{i}} = \{\mathbf{i}\}$ and the profit is simply given by $P_{\mathbf{i}} = \frac{\zeta_{\mathbf{i}}}{W_{\mathbf{i}}}$.

We can now introduce the adaptive algorithm we are considering in this work.

Algorithm 1 Adaptive algorithm (stochastic space adaptation)

Require: $Tol > 0$

Ensure: multi-index set I such that $\zeta_{SC} \leq Tol$

1: $I = \{\mathbf{1}\}$, $u_{h,I} = S_I[u_h]$, $\zeta_{SC} = \zeta_{\mathbf{1}}$
2: **while** $\zeta_{SC} > Tol$ **do**
3: $\quad \mathbf{i}^* = \text{argmax}_{\mathbf{i} \in M_I} P_{\mathbf{i}}$ — select the most profitable multi-index
4: $\quad I \leftarrow I \cup A_{\mathbf{i}^*}$ — update the multi-index set
5: $\quad u_{h,I} = S_I[u_h]$ — compute the new sparse grid approximation
6: $\quad \zeta_{SC} = \sum_{\mathbf{i} \in M_I} \zeta_{\mathbf{i}}$ — compute the error estimator (17)
7: **end while**

Remark 5.1. *Algorithm 1 is one possible adaptive strategy. In particular, we choose to select only one multi-index at each iteration, see line 3. Another possibility would be to allow the selection of several multi-indices, for instance to satisfy a Dörfler-type criterion. Moreover, the selection of the most profitable element is made on the full margin in Algorithm 1. To reduce the computational cost, we could alternatively drive the adaptive process only by the profit of the elements of the reduced margin. In such a case, we do not need to compute $\zeta_{\mathbf{i}}$ for each $\mathbf{i} \in M_I \setminus R_I$. However, the global error estimator ζ_{SC} would then no longer be available and another stopping criterion must be used.*

6 Numerical results

We consider here numerical examples to test the efficiency of the SC error estimator derived in Proposition 4.3 and, in particular, to test the performance of Algorithm 1. In all what follows, the FE error is not accounted for. Moreover, we consider the case $p = \infty$ and we thus consider the error and estimator defined by respectively

$$\|u_h - S_I[u_h]\|_{L^\infty_\rho(\Gamma;H^1_0(D))} \quad \text{and} \quad \sum_{\mathbf{i} \in M_I} \|\Delta^{\mathbf{m(i)}}(a \nabla S_I[u_h])\|_{L^\infty_\rho(\Gamma;L^2(D))}.$$

In order to compute the $L^\infty_\rho(\Gamma)$ norm approximately, we use a set $\theta \subset \Gamma$ of finite cardinality, that is we use the approximation

$$\|g\|_{L^\infty_\rho(\Gamma)} \approx \max_{\mathbf{y} \in \theta} |g(\mathbf{y})|$$

for any $g \in C^0(\Gamma)$. In what follows, we set θ to be constituted of 500 points randomly sampled according to the distribution ρ.

Remark 6.1. *Notice that we do not include the constant a_{min}^{-1} in the estimator. Even though this constant is required to have a guaranteed upper bound on the error, see (14), it will usually lead to a (large) over estimation of the error if taken into account. If the goal is to get a numerical approximation with prescribed accuracy, we can proceed as follows to detect if it would be preferable to include the factor a_{min}^{-1} or not: take a (small) set I, compute the approximation $S_I[u_h]$ and compare $\|\nabla e\|_{L^\infty_\rho(\Gamma;L^2(D))}$ and $\|a^{1/2}\nabla e\|_{L^\infty_\rho(\Gamma;L^2(D))}$ with $e = u_h - S_I[u_h]$. If the H^1 semi-norm and the energy norm of the error are comparable, then a_{min}^{-1} should not be included in the estimator.*

Before performing sparse grid adaptation, we will test the efficiency of the SC estimator considering different approximation spaces chosen a priori. We will consider

both cases $m(i) = i$ and

$$m(i) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } i = 1 \\ 2^{i-1} + 1 & \text{if } i > 1. \end{cases} \tag{29}$$

Since we need nested sequences of points, we use Leja points for the linear case $m(i) = i$ and Clenshaw-Curtis (CC) points if m is defined by (29). We recall that for a generic compact set X and a given initial point $y^0 \in X$, the (standard) Leja points are defined recursively by [34]

$$y^k = \underset{y \in X}{\operatorname{argmax}} \prod_{j=1}^{k-1} (y - y^j), \quad k = 1, 2, ...$$

In what follows, when using Leja points on an interval $\Gamma_i = [a_i, b_i] \subset \mathbb{R}$, we will set the initial point to the endpoint b_i. To test the efficiency of the estimator, we will consider an arbitrary (downward closed) multi-index set I or, for a given level of approximation w, the classical approximation spaces [31] given in Table 1.

Approximation space	m	I	points
Tensor product (TP)	$m(i) = i$	$I(w) = \{\mathbf{i} \in \mathbb{N}_+^N : \max_n(i_n - 1) \leq w\}$	Leja
Total degree (TD)	$m(i) = i$	$I(w) = \{\mathbf{i} \in \mathbb{N}_+^N : \sum_n(i_n - 1) \leq w\}$	Leja
Hyperbolic cross (HC)	$m(i) = i$	$I(w) = \{\mathbf{i} \in \mathbb{N}_+^N : \Pi_n(i_n) \leq w + 1\}$	Leja
Smolyak (SM)	m in (29)	$I(w) = \{\mathbf{i} \in \mathbb{N}_+^N : \sum_n(i_n - 1) \leq w\}$	CC

Table 1: Approximation spaces for testing the efficiency of the SC error estimator.

6.1 First example

We start with the analysis of an inclusion problem, first with $N = 2$ inclusions and then with $N = 8$ inclusions, see [31]. The physical domain is the unit square $D = (0, 1)^2$ in which we identify the subdomains F and C_n, $n = 1, ..., N$ as depicted on Figure 1-left for the case $N = 2$ and on Figure 7-left for $N = 8$. The square subdomain F has a side length of 0.2 while the radius of each circular subdomain C_n is equal to 0.13. We set the forcing term to $f(\mathbf{x}) = 100\chi_F(\mathbf{x})$ and we define the random diffusion coefficient by

$$a(\mathbf{x}, \mathbf{Y}(\omega)) = a_0(\mathbf{x}) + \sum_{n=1}^{N} \gamma_n \chi_n(\mathbf{x}) Y_n(\omega) \quad \text{with } a_0 = 1 \tag{30}$$

where χ_F and χ_n, $n = 1, ..., N$, denote the indicator function of each subdomain. The parameters γ_n, $n = 1, ..., N$ are used to introduce anisotropy in the problem, assigning more importance to one or another direction y_n.

Case $N = 2$

We start with $N = 2$ and we take $Y_n \sim \mathcal{U}[-0.99, 0.99]$ for $n = 1, 2$. The FE mesh we are using consists of 4961 vertices and 9696 triangles with minimal and maximal diameter h_T of about 7.367e-3 and 2.854e-2, respectively. The mean and the standard deviation of the solution is given in Figure 1 for the isotropic case $\gamma_1 = \gamma_2 = 1$.

We give in Figure 2 the error and the estimator with respect to the number of points in the sparse grids for the four types of approximation spaces defined in Table

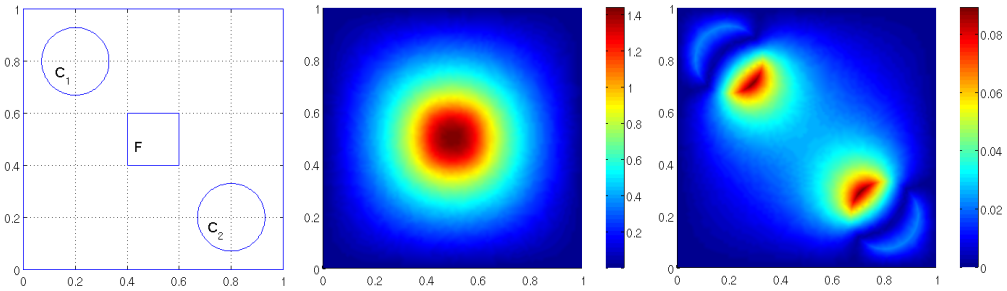

Figure 1: Geometry of the problem (left), expected value (middle) and standard deviation (right) of the solution for the case $\gamma_1 = \gamma_2 = 1$.

1. We consider the isotropic case $\gamma_1 = \gamma_2 = 1$ but also an anisotropic one, namely $\gamma_1 = 1$ and $\gamma_2 = 0.1$. The maximum level of approximation w is set to 10 for TP, 14 for TD, 29 for HC and 5 for SM, which corresponds to a sparse grid of 121, 120, 111 and 145 points in Γ, respectively.

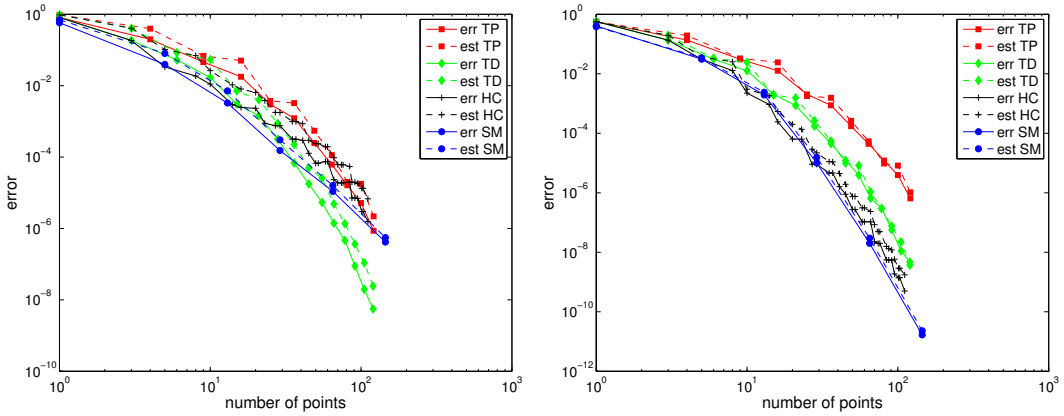

Figure 2: Error and estimator w.r.t. the number of points for the four approximation spaces given in Table 1. Left: isotropic case $\gamma_1 = \gamma_2 = 1$; right: anisotropic case $\gamma_1 = 1$ and $\gamma_2 = 0.1$.

We can see that the estimator provides a good control of the error for all the considered approximation spaces and for both the isotropic and the anisotropic cases. This is also the case when an arbitrary multi-index set is considered. Indeed, let us take for instance $I_1 = [(1,1),(1,2),(1,3),(1,4),(1,5),(2,1),(2,2),(2,3),(3,1)]$, which is a priori not a good set for the considered values of γ_1 and γ_2 as it uses more point for y_2 rather than y_1, and $I_2 = [(1,1),(1,2),(2,1),(2,2),(3,1),(4,1),(5,1)]$. The results we obtain for the two cases $m(i) = i$ with Leja points and m in (29) with CC points are presented in Table 2. Finally, we mention that we observe similar behaviours for all the numerical examples presented below.

We now consider the adaptive strategy proposed in Algorithm 1. From now on, we restrict to Clenshaw-Curtis nodes and m defined in (29). We start with the isotropic case $\gamma_1 = \gamma_2 = 1$. We set the tolerance to $Tol = 10^{-6}$. The evolution of the multi-index set I during the adaptive process is presented in Figure 3. The multi-index in green denote the selected element at the current iteration of Algorithm 1, i.e. the one with the highest profit, before it is added to I.

		$m(i) = i$ and Leja points			m in (29) and CC points		
		♯ pts	error	estimator	♯ pts	error	estimator
iso	I_1	9	3.5977e-2	4.8307e-2	29	2.2348e-3	3.7151e-3
	I_2	7	1.4035e-1	2.1013e-1	23	3.1768e-2	3.5094e-2
aniso	I_1	9	3.0888e-2	3.3697e-2	29	1.9359e-3	2.3869e-3
	I_2	7	2.4784e-3	4.5417e-3	23	8.3280e-5	1.3008e-4

Table 2: Number of points, error and estimator for the given multi-index sets I_1 and I_2. Case iso: $\gamma_1 = \gamma_2 = 1$; case aniso: $\gamma_1 = 1$ and $\gamma_2 = 0.1$.

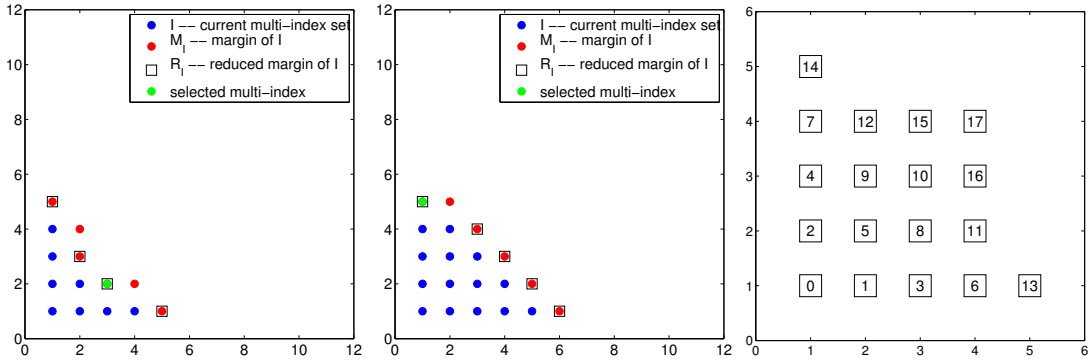

Figure 3: Evolution of I during the adaptive process for the case $\gamma_1 = \gamma_2 = 1$. From left to right: iterations 8 and 14 and order of selection of the multi-indices.

We can detect the isotropy of the problem by the *symmetrical* construction of the multi-index set. For instance, at iteration 11 the point $(4, 2)$ is added while $(2, 4)$ is selected at the next iteration. Moreover, we see that the estimator provides a good control of the error as shown in Figure 4, where the final sparse grid is also given. It has been obtained after 17 iterations, yielding a grid of 97 points and an error and an estimator of about 3.8464e-7 and 7.6980e-7, respectively. The error in energy norm at this final stage, namely $\|a^{\frac{1}{2}}\nabla e\|_{L^\infty_\rho(\Gamma;L^2(D))}$, is about 3.0020e-7 and thus close to the error in H^1 semi-norm. Finally, we mention that the highest profit of the elements of the margin of this final stage is about 2.3220e-8 and is achieved at $(2, 5)$, which belongs to the reduced margin.

We now set different values for γ_1 and γ_2 in (30) to see if the adaptive algorithm is able to capture the anisotropy of the problem. We thus set $\gamma_1 = 1$ and $\gamma_2 = 0.1$. We present in Figure 5 the set I at various steps of the adaptive construction. As expected, the algorithm clearly identifies a preferred direction, namely the horizontal direction which corresponds to y_1.

The final sparse grid for a tolerance of $Tol = 10^{-6}$ in Algorithm 1 is given in Figure 6 and has been reached in 10 iterations. In this case, there are 41 points in the sparse grid, the error and estimator are 6.9851e-8 and 1.2506e-7, respectively, and the maximal profit among the elements of the margin is of about 2.0030e-8 at $(3, 3)$, which belongs to the reduced margin. Finally, the error in H^1 semi-norm is comparable to the error in energy norm, which is about 6.3569e-8.

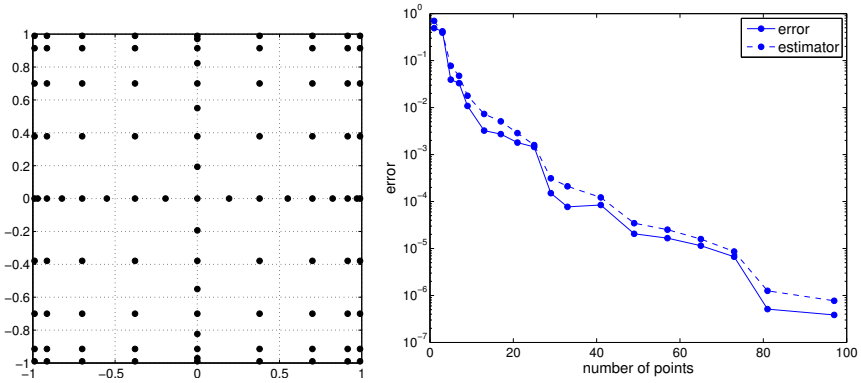

Figure 4: Final sparse grid (left) and error and estimator with respect to the number of points in semi-logarithmic scale (right) for the case $\gamma_1 = \gamma_2 = 1$.

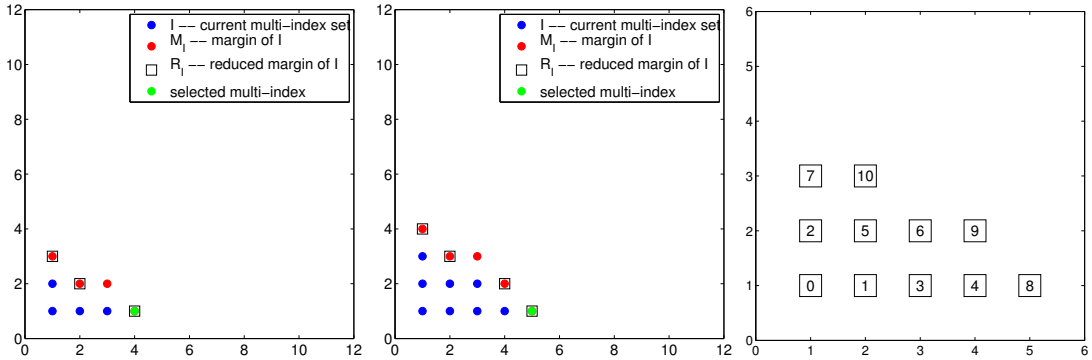

Figure 5: Evolution of the multi-index set I during the adaptive process for the case $\gamma_1 = 1$ and $\gamma_2 = 0.1$. From left to right: iterations 4 and 8 and order of selection of the multi-indices.

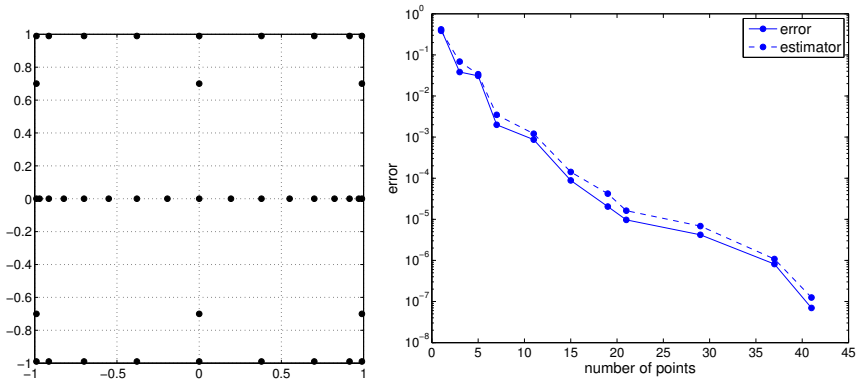

Figure 6: Final sparse grid (left) and error and estimator with respect to the number of points in semi-logarithmic scale (right) for the case $\gamma_1 = 1$ and $\gamma_2 = 0.1$.

Case $N = 8$

To conclude on this inclusion problem, we consider the case $N = 8$ as in [31] and we choose $Y_n \sim \mathcal{U}[-0.99, 0.2]$ for $n = 1, \ldots, 8$ in (30). The geometry is given in Figure 7-left, where the value of the coefficients γ_n, $n = 1, \ldots, 8$, is also given. The FE mesh we are using contains 3805 vertices and 7416 triangles with minimal and maximal diameter h_T of about 1.0041e-2 and 3.1153e-2, respectively. For this case, we set the tolerance to $Tol = 10^{-3}$ in Algorithm 1.

In Figure 7-right, we give the error and the estimator with respect to the number of points in the grid. At the final stage, obtained in 79 iterations, the sparse grid contains 363 points and the error and estimator are about 1.0852e-4 and 9.9014e-4, respectively. Moreover, the error in energy norm is about 8.7246e-5. Finally, the maximum profit among the elements of the margin is about 5.4553e-6 and is achieved at $(1, 1, 1, 2, 1, 2, 2, 1)$.

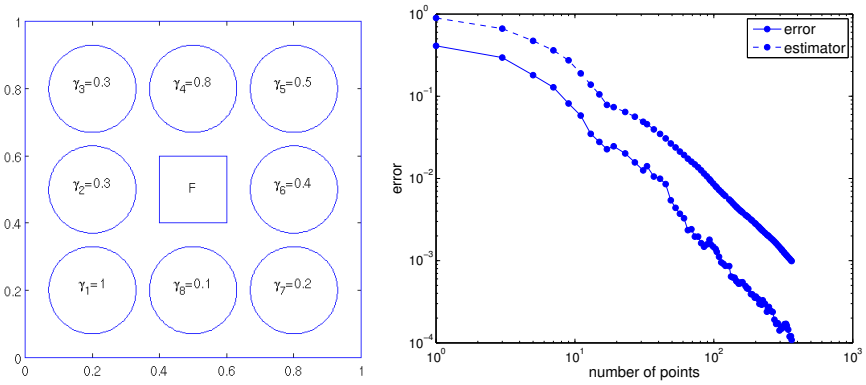

Figure 7: Geometry of the problem for $N = 8$ with indication of the coefficients γ_n, $n = 1, \ldots, 8$ (left) and error and estimator with respect to the number of points in logarithmic scale (right).

In this case, the estimator still provides a reasonable control of the error, even though it is less efficient than for the case $N = 2$. We see several possible explanations for this behaviour and we give a non-exhaustive list below. First of all, we have not been able to prove that the error estimator provides a lower bound for the error. The difficulties arise, among other, from the lack of *Galerkin orthogonality* but also from the use of the triangle inequality to localize the estimator on each multi-index of the margin. Moreover, we are not taking into account the error due to the approximation of the $L^\infty_\rho(\Gamma)$ norm and further investigation should be made in this direction, namely trying to quantify this additional error and perform additional tests with other training sets Θ. The size of the training set could also be adapted with respect to the number of points in the sparse grid and not be fixed once for all as considered here.

The projection of the obtained multi-index set I over two directions, namely y_1 and y_4, y_1 and y_5 and y_1 and y_8, is presented in Figure 8. These results are consistent with the choice we made for the value of the coefficients γ_n, $n = 1, 2, ..., 8$, see Figure 7-left.

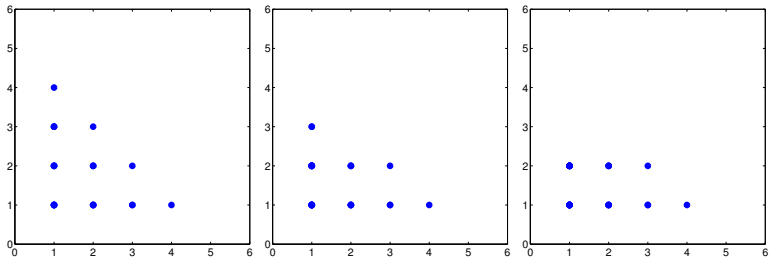

Figure 8: Projection of the multi-index set I, obtained for $Tol = 10^{-3}$ in Algorithm 1, on (y_1, y_4) (left), (y_1, y_5) (middle) and (y_1, y_8) (right).

6.2 Second example

As a second numerical experiment, we consider problem (1) with $D = (0,1)^2$, $f(\mathbf{x}) = 32(x_1(1-x_1) + x_2(1-x_2))$ and

$$a(\mathbf{x}, \mathbf{Y}(\omega)) = 1 + \sum_{n=1}^{N} \frac{\cos(2\pi n x_1) + \cos(2\pi n x_2)}{(\pi n)^2} Y_n(\omega) \quad \text{with} \quad Y_n \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$$

for $\mathbf{x} = (x_1, x_2) \in D$. We use a spatial mesh consisting of 2673 vertices and 5184 triangles with minimum and maximum diameter h_T of about 0.01 and 0.04, respectively. Finally, we consider the two cases $N = 3$ and $N = 5$ and we set the tolerance to $Tol = 10^{-6}$ in Algorithm 1.

The results for the case $N = 3$ are given in Figure 9. We plot the error and the estimator with respect to the number of collocation points. We also give the projection of the final multi-index set I over two directions, namely y_1 and y_3. For this final state, obtained in 27 iterations, the error and the estimator are about 4.3746e-7 and 9.2363e-7, respectively, and the grid contains 141 points. The error in energy norm is about 3.5904e-7. Finally, we mention that the multi-index that has been added in the last iteration to the final set I is $(4, 3, 1)$ and that the maximum profit among the elements of M_I is about 3.0550e-8 and is reached at $(3, 2, 3)$ which belongs to R_I.

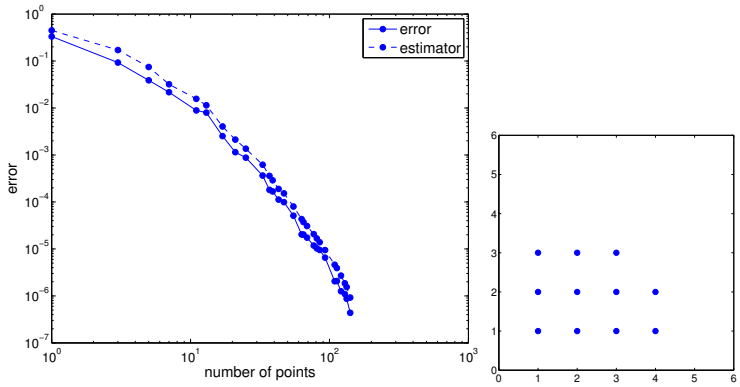

Figure 9: Error and estimator with respect to the number of points in logarithmic scale (left) and projection of the final multi-index set on (y_1, y_3) (right) for the case $N = 3$.

The Figure 10 contains the results for the case $N = 5$. The final multi-index set I is projected on y_1 and y_3 and on y_1 and y_5. The final grid has 973 points,

for an error and estimator of about 1.7666e-7 and 9.9454e-7, respectively, and has been reached in 110 iterations. The error in energy norm at this final stage is about 1.4942e-7. Finally, the last multi-index added to the set is $(4, 2, 1, 2, 2)$ and the maximum profit among the elements of the margin of the final set is about 3.9748e-9 at $(4, 3, 1, 2, 1) \in R_I$.

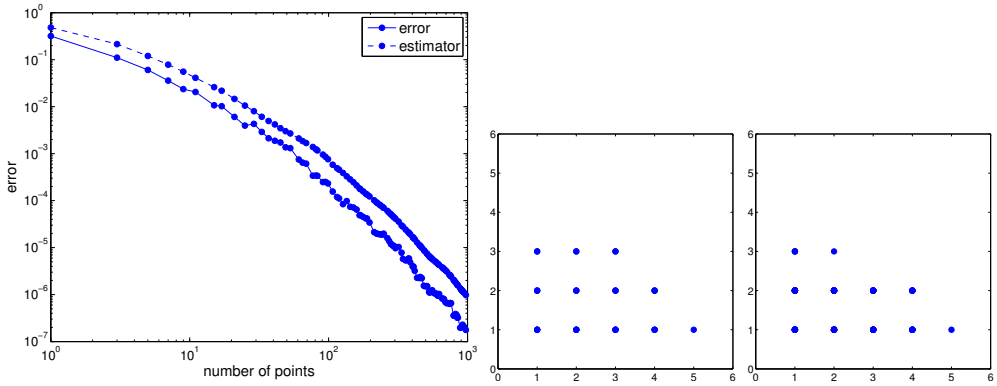

Figure 10: Error and estimator with respect to the number of points in logarithmic scale (left) and projection of the final multi-index set on (y_1, y_3) (middle) and on (y_1, y_5) (right) for the case $N = 5$.

In both cases $N = 3$ and $N = 5$, the error estimator provides a good control of the error, the overestimation being slightly bigger for $N = 5$ than $N = 3$. Moreover, due to the decay of the a_n in n^{-2}, the random variables Y_n should have less and less influence as n increases. The adaptive algorithm is able to capture this feature, as seen for instance when projecting the obtained multi-index set over two different directions. From this experiment, together with the numerical results obtained for the inclusion problems, we see that the efficiency of the stochastic error estimator seems to be linked to the number of random variables. Further investigation should be made in this direction to determine whether this is indeed the case or if the reason is elsewhere, for instance the error due to the approximation of the $L^\infty_\rho(\Gamma)$ norm.

Remark 6.2. *For all the numerical examples given above, the selected multi-index at each iteration of Algorithm 1 belongs to R_I. In what follows, we consider a 1D example for which the optimal set is not downward closed, as observed in [12]. The goal is then to see if our adaptive algorithm capture this feature.*

6.3 1D numerical example

We consider the problem (1) with $D = (0, 1)$ the unit interval, $f(x) = 1$ and $a(x, \mathbf{Y}(\omega)) = 1 + 0.1Y_1(\omega) + 0.5Y_2(\omega)$ where $Y_n \sim \mathcal{U}[-1, 1]$ for $n = 1, 2$.

For the FE mesh, we consider a uniform partition of the unit interval with mesh size $h = 2^{-12}$, that is we discretize $[0, 1]$ taking the nodes $x_i = ih$ with $i = 0, ..., 2^{12}$. We give in Figure 11 three different examples for which the selected multi-index belongs to $M_I \setminus R_I$. For such multi-index, more than one element is added to I because of the constraint that I remains downward closed during the adaptive process. If we set the tolerance to $Tol = 10^{-8}$ in Algorithm 1, the adaptive process stops after 16 iterations and the sparse grid contains 153 points in Γ. Moreover, the corresponding error and estimator are about 2.4882e-9 and 4.2305e-9, respectively, while the error in energy norm is about 2.0389e-9.

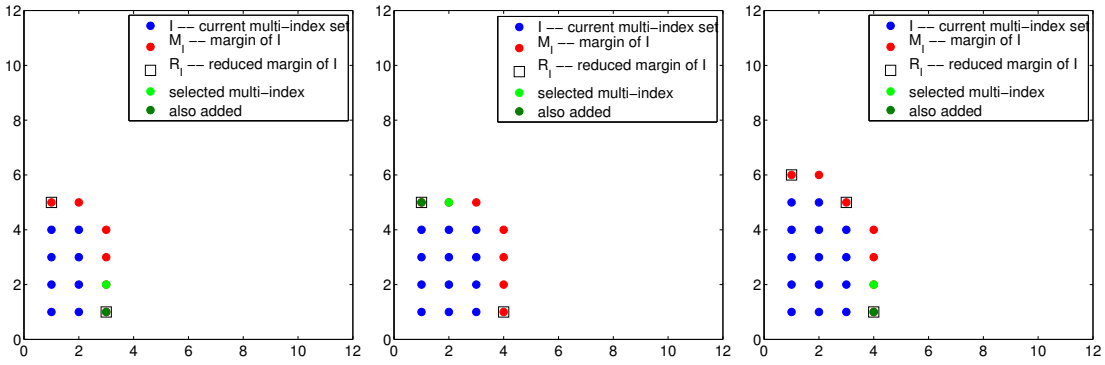

Figure 11: Three examples for which the selected multi-index belongs to $M_I \setminus R_I$ which correspond to iterations 8 (left), 11 (middle) and 12 (right).

7 Dimension adaptive

We provide here some hints about how we could proceed to deal with the case where the number N of random variables in the system is large, possibly infinite. For such problem, the cardinality of the margin of the multi-index set I becomes large and the computation of the error estimator ζ_{SC} is no longer feasible. The idea would then be to activate only a fraction of the directions y_n, the more important ones, as proposed in [20]. The error committed by neglecting some directions should then be appropriately estimated. A first step in this direction is proposed below, namely we provide the main relation which separates the various sources of error. Let use rewrite the diffusion coefficient a defined in (3) as a_N to highlight its dependence on the N random variables Y_n, $n = 1, \ldots, N$. Similarly, we write u_N the solution of the diffusion problem with diffusion coefficient a_N. For $1 \leq M \leq N$, let u_M be the solution of the diffusion problem with coefficient $a_M(x, \mathbf{y}) = a_0(x) + \sum_{n=1}^{M} a_n(x) y_n$. The goal is to estimate the error $u_N - S_I[u_{M,h}]$ where $u_{M,h}$ is the FE approximation of u_M and S_I is the sparse grid interpolant based on $I \subset \mathbb{N}_+^M$, i.e. with M *active variables*. We can easily show that for any $v \in H_0^1(D)$ and a.s. in Ω we have

$$\int_D a_N \nabla(u_N - S_I[u_{M,h}]) \cdot \nabla v = A_1 + A_2 + A_3$$

with

$$A_1 := S_I \left[\int_D fv - \int_D a_M \nabla u_{M,h} \cdot \nabla v \right] \tag{31}$$

$$A_2 := -\int_D \sum_{\mathbf{i} \in M_I} \Delta^{m(\mathbf{i})} \left(a_M \nabla S_I[u_{M,h}] \right) \cdot \nabla v \tag{32}$$

$$A_3 := -\int_D (a_N - a_M) \nabla S_I[u_{M,h}] \cdot \nabla v. \tag{33}$$

Indeed, we have

$$\begin{aligned} \int_D a_N \nabla(u_N - S_I[u_{M,h}]) \cdot \nabla v &= \int_D fv - \int_D a_N \nabla S_I[u_{M,h}] \cdot \nabla v \\ &= \int_D fv - \int_D a_M \nabla S_I[u_{M,h}] \cdot \nabla v - \int_D (a_N - a_M) \nabla S_I[u_{M,h}] \cdot \nabla v \end{aligned}$$

and the first two terms of the right-hand side can be split into A_1 and A_2 defined in (31) and (32), respectively, proceeding exactly as in the proof of Proposition

4.3. The terms A_1 and A_2, which correspond to the errors due to FE and SC, respectively, can be estimated proceeding exactly as in section 4. Finally, for the term A_3, which account for the neglection of some directions, we can use the relation

$$A_3 = -\int_D \sum_{n=M+1}^{N} a_n y_n \nabla S_I[u_{M,h}] \cdot \nabla v.$$

8 Conclusions

In this work we have derived a residual-based *a posteriori* error estimate for the stochastic collocation finite element method, focusing on an elliptic model problem with a random diffusion coefficient. Our error estimate is valid under the assumptions that the diffusion coefficient depends affinely on the random variables and that the sparse grid approximation is interpolatory, which requires the use of nested points. The error estimate, which provides an upper bound of the total error, is constituted of two parts accounting for the finite element and the stochastic collocation errors, respectively. We have then used the stochastic collocation estimator to drive an adaptive strategy in which the multi-index set characterizing the sparse grid is constructed step by step. We assign a profit to each element of the margin of the set and, at each iteration, we select the most profitable one to enter the set. We have provided several numerical examples of moderate dimension to illustrate the theoretical findings. More precisely, we have compared the error and the estimator for various given multi-index sets and we have then test the efficiency of the proposed adaptive algorithm. The latter, which uses the stochastic collocation estimator to drive the adaptive process, is one possible strategy. Several other versions could be considered as well, for instance by selecting more than one multi-index at each iteration using a so-called Dörfler or maximum marking strategy with, ideally, a proof of convergence. In the case of high-dimensional problems, that is when the coefficient depends on (possibly infinitely) many random variables, the computation of the profit of each element of the margin become prohibitive and an alternative should be used. We have given some insight in this direction but further investigations, including numerical experiments, should be done.

References

[1] 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.

[2] 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.

[3] 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.

[4] G.S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer Ser. Oper. Res. Financ. Eng. Springer, New York, 1996.

[5] J. Dick and F. Pillichshammer. *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration.* Cambridge University Press, Cambridge, 2010.

[6] J. Dick, F.Y. Kuo, and I.H. Sloan. High-dimensional integration: The quasi-Monte Carlo way. *Acta Numer.*, 22:133–288, 2013.

[7] M.B. Giles. Multi-level Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[8] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer, New York, 1991.

[9] O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Sci. Comput. Springer Netherlands, 2010.

[10] D. Guignard, F. Nobile, and M. Picasso. A posteriori error estimation for elliptic partial differential equations with small uncertainties. *Numer. Methods Partial Differential Equations*, 32(1):175–212, 2016.

[11] D. Guignard, F. Nobile, and M. Picasso. A posteriori error estimation for the steady Navier-Stokes equations in random domains. *Comput. Methods Appl. Mech. Engrg.*, 313:483–511, 2017.

[12] J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 22(9):1250023–1–1250023–33, 2012.

[13] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric pdes. *Found. Comput. Math.*, pages 1–33, 2013.

[14] R.C. Almeida and J.T. Oden. Solution verification, goal-oriented adaptive methods for stochastic advection-diffusion problems. *Comput. Methods Appl. Mech. Engrg.*, 199:2472–2486, 2010.

[15] F. Nobile, R. Tempone, and C.G. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM J. Numer. Anal.*, 46(5):2411–2442, 2008.

[16] F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse grid approximation of Hilbert-valued functions: application to random elliptic PDEs. *Numer. Math.*, 134(2):343–388, 2016.

[17] T. Gerstner and M. Griebel. Dimension-adaptive tensor-product quadrature. *Computing*, 71(1):65–87, 2003.

[18] H.J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[19] A. Klimke. *Uncertainty modeling using fuzzy arithmetic and sparse grids*. PhD thesis, Universität Stuttgart, 2006.

[20] F. Nobile, L. Tamellini, F. Tesei, and R. Tempone. An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient. In J. Garcke and D. Pflüger, editors, *Sparse Grids and Applications - Stuttgart 2014*, pages 191–220. Springer International Publishing, 2016.

[21] M. Griebel and S. Knapek. Optimized general sparse grid approximation spaces for operator equations. *Math. Comp.*, 78(268):2223–2257, 2009.

[22] C. Schillings and C. Schwab. Sparse, adaptive smolyak quadratures for bayesian inverse problems. *Inverse Probl.*, 29(6):065011, 2013.

[23] F. Nobile and F. Tesei. A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients. *Stoch. PDE: Anal. Comp.*, 3(3):398–444, 2015.

[24] A. Chkifa, A. Cohen, R. DeVore, and C. Schwab. Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic PDEs. *ESAIM: Math. Model. Numer. Anal.*, 47(01):253–280, 2013.

[25] M. Eigel, C.J. Gittelson, C. Schwab, and E. Zander. Adaptive stochastic Galerkin FEM. *Comput. Methods Appl. Mech. Engrg.*, 270:247–269, 2014.

[26] M. Eigel, C.J. Gittelson, C. Schwab, and E. Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM: Math. Model. Numer. Anal.*, 49(5):1367–1398, 2015.

[27] J.M. Cascon, C. Kreuzer, R.H. Nochetto, and K.G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.

[28] S.A. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Math. Dokl.*, 4:240–243, 1963. [Russian original in Dokl. Akad. Nauk SSSR, 148:1042–1045, 1963].

[29] G.W. Wasilkowski and H. Wozniakowski. Explicit cost bounds of algorithms for multivariate tensor product problems. *J. Complexity*, 11(1):1–56, 1995.

[30] V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Adv. Comput. Math.*, 12(4):273–288, 2000.

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

[32] P. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.*, 9(R2):77–84, 1975.

[33] A. Chkifa. *Méthodes polynomiales parcimonieuses en grande dimension. Application aux EDP paramétriques.* PhD thesis, Laboratoire Jacques Louis Lions, 2014.

[34] F. Leja. Sur certaines suites liées aux ensembles plans et leur application à la représentation conforme. *Ann. Polon. Math.*, 4(1):8–13, 1957.

Recent publications:

INSTITUTE of MATHEMATICS

MATHICSE Group

Ecole Polytechnique Fédérale (EPFL)

CH-1015 Lausanne

2017

12.2017 SEBASTIAN KRUMSCHEID, FABIO NOBILE:
Multilevel Monte Carlo approximation of functions,

13.2017 R.N. SIMPSON, Z. LIU, R. VÁZQUEZ, J.A. EVANS:
An isogeometric boundary element method for electromagnetic scattering with compatible B-spline discretization

14.2017 NICCOLO DAL SANTO, SIMONE DEPARIS, ANDREA MANZONI:
A numerical investigation of multi space reduced basis preconditioners for parametrized elliptic advection-diffusion

15.2017 ASSYR ABDULLE, TIMOTHÉE POUCHON:
Effective models for long time wave propagation in locally periodic media

16.2017 ASSYR ABDULLE, MARCUS J. GROTE, ORANE JECKER:
FE-HMM for elastic waves in heterogeneous media

17.2017 JOHN A. EVANS, MICHAEL A. SCOTT, KENDRICK SHEPHERD, DEREK THOMAS, RAFAEL VÁZQUEZ:
Hierarchical B-spline complexes of discrete differential forms

18.2017 ELEONORA MUSHARBASH, FABIO NOBILE:
Symplectic dynamical low rank approximation of wave equations with random parameters

19.2017 ASSYR ABDULLE, IBRAHIM ALMUSLIMANI, GILLES VILMART:
Optimal explicit stabilized integrator of weak order one for stiff and ergodic stochastic differential equations

20.2017 ABDUL-LATEEF HAJI-ALI, FABIO NOBILE, RAÚL TEMPONE, SÖREN WOLFERS:
Multilevel weighted least squares polynomial approximation

21.2017 NICCOLO DAL SANTO, SIMONE DEPARIS, ANDREA MANZONI, ALFIO QUARTERONI:
An algebraic least squares reduced basis method for the solution of parametrized Stokes equations

22.2017 ALEXEY CHERNOY, HAKON HOEL, KODY J. H. LAW, FABIO NOBILE, RAUL TEMPONE:
Multilevel ensemble Kalman filtering fors patio-temporal processes

23.2017 MICHELE PISARONI, SEBASTIAN KRUMSCHEID, FABIO NOBILE:
Quantifying uncertain system outputs via the multilevel Monte Carlo method – Part I: Central moment estimation

24.2017 DIANE GUIGNARD, FABIO NOBILE:
A posteriori error estimation for the stochastic collocation finite element method

***