A posteriori error estimation for the stochastic collocation finite element approximation of the heat equation with random coefficients

Fabio Nobile and Eva Vidličková

Abstract In this work we present a residual based a posteriori error estimation for a heat equation with a random forcing term and a random diffusion coefficient which is assumed to depend affinely on a finite number of independent random variables. The problem is discretized by a stochastic collocation finite element method and advanced in time by the θ-scheme. The a posteriori error estimate consists of three parts controlling the finite element error, the time discretization error and the stochastic collocation error, respectively. These estimators are then used to drive an adaptive choice of FE mesh, collocation points and time steps. We study the effectiveness of the estimate and the performance of the adaptive algorithm on a numerical example.

1 Introduction

Many physical and engineering applications are modeled by partial differential equations (PDEs) with input data often subject to uncertainty due to measurement errors or insufficient knowledge. These uncertainties can be modeled by means of probability theory by introducing a set of random variables into the system. The PDEs are then solved approximately by numerical schemes, which brings along a high interest in reliable error estimation. In this work we consider a heat equation with random diffusion coefficient and forcing term discretized by the θ-scheme in time, the finite element method (FEM) in physical space and a stochastic collocation (SC) method in the random variables.

Fabio Nobile
École polytechnique fédérale de Lausanne, Institute of Mathematics, e-mail: fabio.nobile@epfl.ch

Eva Vidličková
École polytechnique fédérale de Lausanne, Institute of Mathematics ,e-mail: eva.vidlickova@epfl.ch

There is a vast literature on a posteriori error estimation for deterministic problems, mostly in the context of finite element discretizations. This subject was initiated in [6, 7] and improved in e.g. [1, 4, 26, 38] for elliptic problems. Concerning parabolic problems, a posteriori error estimations have been proposed in [21, 22] when using the Discontinuous Galerkin (DG) method as a space-time discretization. In [33] an implicit Euler scheme together with continuous, piecewise linear finite elements is considered and an adaptive algorithm is presented. In [36] the authors present a posteriori estimations for a more general $\theta-$scheme and finite elements of arbitrary order. This result is improved in [30, 2] for $\theta = 1/2$, i.e. the Crank-Nicholson scheme, by introducing a continuous, piecewise quadratic interpolation in time rather than a piecewise linear one. In [30] an adaptive algorithm is presented as well.

There is much less literature available for the a posteriori error estimation for random PDEs. When uncertainties are treated by the stochastic Galerkin method [23, 29], a posteriori error estimations together with an adaptive algorithm have been proposed in [27, 28] for a linear elasticity equation and in [19, 18, 20, 13, 16] for the case of an elliptic PDE with diffusion coefficient affinely dependent on a countable number of random variables and, recently, in [14] beyond the affine case. Concerning parabolic equations, the only work we are aware of considers uncertainty only in the Robin boundary condition, solved by the perturbation approach in [24].

In this work we will consider the stochastic collocation (SC) method [40, 5, 32] as one of the ways to treat uncertainties numerically. As opposed to an intrusive method such as stochastic Galerkin, the SC method requires only the solution of decoupled deterministic problems and thus allows the re-use of deterministic solvers. A priori error estimates have been derived in [5, 15, 11]. We will focus on sparse grid collocation methods which partially cure the curse of dimensionality, especially in the case of an anisotropic behaviour of the solution [31]. Concerning a reliable estimation of the discretization error, the work [25] proposes a residual based a posteriori error estimation for an elliptic problem discretized by a stochastic collocation finite element method and presents and numerically studies an algorithm that adaptively builds the sparse grid based on the a posteriori estimation of the SC error.

This work extends the results obtained in [25] to a heat equation with random right hand side and random diffusion coefficient that depends affinely on a finite number of random variables. We adopt the setting from [36] to treat the spatial and time discretization errors. Our estimator allows spatial meshes and sparse grids to change in time and provides estimates of the norm of the error in L^2 in stochastic space, L^2 in time and H^1 in physical space. The estimator naturally splits into a spatial discretization estimator, time discretization estimator and stochastic discretization estimator, which are then used to drive the adaptivity with respect to all three types of discretizations. We then propose an adaptive algorithm to build a suitable time discretization and a FE mesh and sparse gird common to all time steps, so as to achieve a prescribed tolerance on a global norm of the error. We the apply this to a problem with a deterministic right hand side and a diffusion coefficient depending affinely on few random variables.

The outline of the paper is the following. In section 2 we introduce the problem, namely a heat equation with a random diffusion coefficient and right hand side.

Section 3 is dedicated to defining the spatial, time and stochastic discretization. In section 4 we derive a first residual based a posteriori estimation for the general case of sparse grids and spatial meshes that change in time and a second, simpler, estimation for the case of spatial mesh and sparse grid kept fixed over the time iterations. Section 5 introduces an adaptive algorithm to build nonuniform time discretizations, as well as nonuniform meshes and anisotropic sparse grids that are fixed in time for the case of a deterministic right hand side. In section 6 we first study the behaviour and sharpness of all three components of the estimator (spatial, temporal and stochastic) and then apply the adaptive algorithm to propose a time discretization, mesh and sparse grid that guarantee the error to be bellow a prescribed tolerance.

2 Problem statement

Let $D \subset \mathbb{R}^d$ be an open polygonal domain with Lipschitz boundary ∂D and $(\Omega, \mathscr{F}, P)$ be a complete probability space. Given a final time T, random forcing term $f : D \times \Omega \times (0,T) \to \mathbb{R}$, initial condition $u_0 : D \times \Omega \to \mathbb{R}$ and a diffusion coefficient $a : D \times \Omega \to \mathbb{R}$, the problem states: find a solution $u :\ D \times \Omega \times (0,T] \to \mathbb{R}$ satisfying $P-$almost everywhere in Ω

$$\begin{aligned} \frac{\partial u}{\partial t} - \nabla \cdot (a \nabla u) &= f && \text{in } D \times \Omega \times (0,T], \\ u &= 0 && \text{on } \partial D \times \Omega \times (0,T], \\ u(\cdot,\cdot,0) &= u_0 && \text{in } D \times \Omega. \end{aligned} \tag{1}$$

Suppose that $f \in L^2(0,T;L^2(\Omega,H^{-1}(D)))$, $u_0 \in L^2(\Omega,H^1_0(D))$ and a is a random variable on $(\Omega,\mathscr{F},P)$ taking values in $W^{1,\infty}(D)$ (random field) satisfying

$$\exists\, a_{min}, a_{max} : P(\omega \in \Omega : 0 < a_{min} \leq a(x,\omega) \leq a_{max} < \infty \quad \forall x \in D) = 1. \tag{2}$$

In addition we require that the diffusion coefficient as well as the forcing term and the initial condition can be parametrized by a finite number of independent, real-valued random variables $\{Y_m\}_{m=1}^M$ defined on Ω, i.e. $f(x,\omega,t) = f(x,Y_1(\omega),\ldots,Y_M(\omega),t)$, $u_0(x,\omega) = u_0(x,Y_1(\omega),\ldots,Y_M(\omega))$ and the dependence of a on $\{Y_m\}_{m=1}^M$ is affine, i.e.

$$a(x,\omega) = a_0(x) + \sum_{m=1}^{M} a_m(x) Y_m(\omega). \tag{3}$$

The solution u then depends on the same random variables as well, i.e. $u(x,\omega,t) = u(x,Y_1(\omega),\ldots,Y_M(\omega),t)$, and we can recast the probability space $(\Omega,\mathscr{F},P)$ into $(\Gamma, B(\Gamma), \rho(y)\mathrm{d}y)$ by introducing $\Gamma = \Gamma_1 \times \cdots \times \Gamma_M$ with $\Gamma_m = Y_m(\Omega)$ for $m = 1,\ldots,M$. $B(\Gamma)$ denotes the Borel $\sigma-$algebra defined over Γ. The joint probability density function of the random vector $Y = (Y_1,\ldots,Y_M)$ is denoted by $\rho : \Gamma \to \mathbb{R}_+$

and factorizes as $\rho(y) = \Pi_{m=1}^{M} \rho_m(y_m)$ for all $y = (y_1, \dots, y_M) \in \Gamma$.

In what follows we consider the following two Bochner spaces: for a given Banach space $(V, \|\cdot\|_V)$ and for any $t_1, t_2 \in [0,T]$, $t_1 < t_2$ we define

$$L^2(t_1,t_2;V) = \{v : (t_1,t_2) \to V | v \text{ is strongly measurable and } \|v\|_{L^2(t_1,t_2;V)} < \infty\}$$

where $\|v\|^2_{L^2(t_1,t_2;V)} = \int_{t_1}^{t_2} \|v(t)\|_V^2 \,\mathrm{d}t$ and

$$L^2_\rho(\Gamma; L^2(t_1,t_2;V)) = \{v : \Gamma \to L^2(t_1,t_2;V) | \, v \text{ is strongly measurable and } \|v\|_{L^2_\rho(\Gamma;L^2(t_1,t_2;V))} < \infty\}$$

with $\|v\|^2_{L^2_\rho(\Gamma,L^2(t_1,t_2;V))} = \int_\Gamma \|v(y)\|^2_{L^2(t_1,t_2;V)} \rho(y) \mathrm{d}y$. It holds

$$L^2_\rho(\Gamma; L^2(t_1,t_2;V)) \cong L^2(t_1,t_2; L^2_\rho(\Gamma;V)),$$

i.e. this Lebesgue-Bochner space is isometrically isomorphic to the Bochner space $L^2(t_1,t_2;L^2_\rho(\Gamma;V))$ [35, p. 12].

The (pointwise in Γ) weak formulation of problem (1) then reads: Find $u \in W$ where

$$W = \left\{ w \in L^2_\rho\left(\Gamma; L^2\left(0,T;H^1_0(D)\right)\right) \text{ and } \frac{\partial w}{\partial t} \in L^2_\rho\left(\Gamma; L^2\left(0,T,H^{-1}(D)\right)\right)\right\}$$

s.t.

$$\int_D \frac{\partial u(x,y,t)}{\partial t} v(x)\,\mathrm{d}x + \int_D a(x,y) \nabla u(x,y,t) \cdot \nabla v(x)\,\mathrm{d}x = \int_D f(x,y,t) v(x)\,\mathrm{d}x \qquad \forall v \in H^1_0(D),\ \rho - \text{a.e. } y \in \Gamma, \text{ and a.e. } t \in (0,T) \tag{4}$$

with initial and boundary conditions:

$$\begin{aligned} u(x,y,0) &= u_0(x,y) && \rho\text{-a.e. } y \in \Gamma \\ u(x,y,t) &= 0 && x \in \partial D,\ \rho\text{-a.e. } y \in \Gamma, \text{a.e. } t \in (0,T). \end{aligned}$$

We endow the Sobolev space $H^1_0(D)$ with the gradient norm $\|v\|_{H^1_0} = \|\nabla v\|_{L^2(D)}$. Based on the existence result of the deterministic problem [17, p.513], the assumption (2) ensures the well-posedness of problem (4), i.e. there exists a unique solution $u \in W$ which moreover satisfies

$$\|u\|_{L^2_\rho(\Gamma;L^2(0,T;H^1_0(D)))} \le \frac{C}{\sqrt{a_{min}}} \left[\|u_0\|^2_{L^2_\rho(\Gamma;H^1_0(D))} + \frac{1}{a_{min}} \|f\|^2_{L^2_\rho(\Gamma;L^2(0,T;L^2(D)))} \right]^{1/2}.$$

3 Discretization aspects

In the following sub-sections we describe the techniques used for the discretization of problem (4) and corresponding assumptions necessary for a rigorous a posteriori estimation. We will closely follow the techniques used in [37, 36] for the time and space discretization and [25, 32, 31, 5] for the stochastic discretization by the stochastic collocation method.

3.1 *Time discretization*

For the time discretization we divide the time interval into N subintervals $0 = t_0 < t_1 < \cdots < t_N = T$. By τ we will denote the discretization $\tau = \{t_n\}_{n=1}^N$ and τ_{n+1} will denote the length of the $(n+1)$-th interval $\tau_{n+1} = t_{n+1} - t_n$. We will also assume that f is continuous w.r.t. time. We will use the abbreviations

$$g^n(x,y) = g(x,y,t_n), \quad g^{n\theta} = (1-\theta)g^n + \theta g^{n+1}.$$

The numerical scheme considered here for the time discretization is the $\theta-$scheme with $\theta \in [0,1]$.

3.2 *Space discretization*

The spatial discretization will be performed by the finite element method. To each time instant $t_n, 0 \leq n \leq N$, we associate a triangulation $\mathscr{T}_{h_n}$ of D which satisfies $\bigcup_{K\in\mathscr{T}_{h_n}} K = D$ and a corresponding conforming finite element space V_{h_n}. For a rigorous estimation we require the following conditions to be satisfied, which are taken from [36].

1. *Affine equivalence:* there is an invertible affine mapping for every element $K \in \mathscr{T}_{h_n}$ onto the standard reference d-simplex or the standard unit cube in $\mathbb{R}^d$.
2. *Admissibility:* any two elements either share a vertex or a complete edge $(d = 2)$ or a complete face $(d = 3)$ or are disjoint.
3. *Shape regularity:* the ratio of the diameter of any element to the diameter of its largest inscribed ball is bounded uniformly with respect to all partitions $\mathscr{T}_{h_n}$ and to N.
4. *Transition condition:* for every $n = 1,\ldots,N$ there is a refinement of both $\mathscr{T}_{h_n}$ and $\mathscr{T}_{h_{n-1}}$, denoted by $\tilde{\mathscr{T}}_{h_n}$, which is an affinely equivalent, admissable and shape-regular triangulation and such that

$$\sup_{1\leq n\leq N} \sup_{K\in\tilde{\mathscr{T}}_{h_n}} \sup_{K'\in\mathscr{T}_{h_n};K'\supset K} \frac{h_{K'}}{h_K} < \infty. \tag{5}$$

5. For every $n = 1, \dots, N$, V_{h_n} consists of continuous functions which are piecewise polynomials of degree $\leq p_n,\ p_n \geq 1$ where p_n is uniformly bounded with respect to N.

3.3 Stochastic discretization

The stochastic discretization is performed by a sparse grid collocation method, first introduced in [34]. We will briefly recall this method and refer the reader to [32, 31, 5] for more details.

Let us define a sequence of univariate polynomial interpolant operators

$$\mathscr{U}_j^{m(i_j)} : C^0(\Gamma_j) \to \mathbb{P}_{m(i_j)-1}(\Gamma_j), \quad j = 1, \dots, M,$$

where $m(i_j)$, called a level function, denotes the number of collocation points for level i_j and $\mathbb{P}_q(\Gamma_j)$ is the space of polynomials over Γ_j with degree at most q. The function m is a strictly increasing function satisfying $m(0) = 0, m(1) = 1$. For a multi-index $q = (q_1, \dots, q_M) \in \mathbb{N}^M$, we denote by $\mathbb{P}_q(\Gamma)$ the tensor product polynomial space $\mathbb{P}_q(\Gamma) = \bigoplus_{j=1}^M \mathbb{P}_{q_j}(\Gamma_j)$.

A sparse grid is built over a multi-index set $I \subset \mathbb{N}_+^M$ with the only assumption being that I is downward-closed (called also admissibility condition), i.e.

$$\forall i \in I,\ i - e_j \in I \quad \forall j \in \{1, 2, \dots, M\} \quad \text{s.t.} \quad i_j > 1,$$

where e_j is the $j-$th canonical unit vector.

By setting $\mathscr{U}_j^0 = 0$ for $j = \{1, \dots, M\}$ we can define the sparse grid interpolant $S_I : L^2_\rho(\Gamma) \cap C^0(\Gamma) \to \mathbb{P}_I := \bigoplus_{i \in I} \mathbb{P}_{m(i)-1}(\Gamma)$ of a continuous function $f : \Gamma \to \mathbb{R}$ by

$$S_I[f](y) = \sum_{i \in I} \triangle^{m(i)}[f](y), \tag{6}$$

where

$$\triangle^{m(i)} = \bigotimes_{j=1}^M \big(\mathscr{U}_j^{m(i_j)} - \mathscr{U}_j^{m(i_j-1)}\big).$$

The operator S_I can be equivalently expressed as a linear combination of tensor grid interpolations (see [31])

$$S_I[f](y) = \sum_{i \in I} c_i \bigotimes_{j=1}^M \mathscr{U}_j^{m(i_j)}(f)(y), \qquad c_i = \sum_{\substack{k \in \{0,1\}^M \\ (i+k) \in I}} (-1)^{|k|} \tag{7}$$

with $|k| = \sum_{j=1}^{M} k_j$. We then call a sparse grid the collection of $N_c(I)$ points $\mathscr{X}(I) = \{y_1, \ldots, y_{N_c(I)}\}$ that are used in (7) to build the interpolant $S_I[f]$. The collocation points are called nested if we have $\mathscr{X}(I) \subset \mathscr{X}(J)$ whenever $I \subset J$. Since $S_I[f]$ is linear in the point evaluations $\{f(y_k),\ y_k \in \mathscr{X}(I)\}$, it can be written in the form

$$S_I[f](y) = \sum_{k=1}^{N_c(I)} f(y_k) L_k(y) \tag{8}$$

for suitable functions L_k. Finally, we introduce the notion of margin M_I of the index set I defined by

$$M_I = \{i \in \mathbb{N}_+^M \setminus I : \quad i - e_j \in I \text{ for some } j \in \{1, \ldots, M\}\}. \tag{9}$$

Equation (4) will be collocated on the grid $\mathscr{X}(I_n) = \{y_1, \ldots, y_{N_c(I_n)}\}$ defined by an index set I_n that is allowed to change between the time steps. In particular, we allow for both refinement and coarsening of the index set. The collocation points are assumed to be nested. This condition implies, in particular, that S_{I_n} is interpolatory, i.e.

$$S_{I_n}[f](y_k) = f(y_k), \quad k = 1, \ldots, N_c(I_n), \quad n = 0, \ldots, N, \tag{10}$$

see [10, p. 277]. By $\tilde{I}_{n+1}$ we will denote the index set

$$\tilde{I}_{n+1} = I_n \cup I_{n+1}. \tag{11}$$

The following proposition will be useful for the derivation of the error estimates.

Proposition 1. *Let S_I be an interpolatory sparse grid interpolant, as defined in* (6)*. Then*

1. $\forall f, g \in C^0(\Gamma)$: $\qquad S_I[f\,g] = S_I[f\,S_I[g]]$,
2. $\forall f \in C^0(\Gamma)$: $\qquad S_I[f] \in \mathbb{P}_I$,
3. $\forall p \in \mathbb{P}_I(\Gamma)$: $\qquad S_I[p] = p$.

A proof can be found in [25, p.3126] for part 1. and in [8, p.52] for part 2.,3.

If S_I is interpolatory, then the functions L_k in (8) are Lagrangian, i.e. $L_k(y_j) = \delta_{jk}$ and form a basis of $\mathbb{P}_I$.

3.4 Fully discrete problem

We allow the spatial and the stochastic grid to change over time and we define the discrete solution for each $n = 0, \ldots, N$ as a function belonging to $V_{h_n} \otimes \mathbb{P}_{I_n}$:

$$u^n_{h_n,I_n} = \sum_{k=1}^{N_c(I_n)} u^n_{h_n,I_n,k} L_k(y),$$

where $u^n_{h_n,I_n,k} = u^n_{h_n,I_n}(y_k) \in V_{h_n}$ and $u^{n+1}_{h_{n+1},I_{n+1}}$ satisfies for $\forall v_{h_{n+1}} \in V_{h_{n+1}}$ and for $\forall k = 1,\dots,N_c(I_{n+1})$ the equation

$$\begin{aligned}\int_D \frac{u^{n+1}_{h_{n+1},I_{n+1},k}(x) - u^n_{h_n,I_n}(x,y_k)}{\tau_{n+1}} v_{h_{n+1}}(x)\mathrm{d}x \\ + \int_D a(x,y_k)\big(\theta\nabla u^{n+1}_{h_{n+1},I_{n+1},k}(x) + (1-\theta)\nabla u^n_{h_n,I_n}(x,y_k)\big)\nabla v_{h_{n+1}}(x)\mathrm{d}x \\ = \int_D f^{n\theta}(x,y_k)\, v_{h_{n+1}}(x)\mathrm{d}x\end{aligned} \tag{12}$$

with initial condition

$$u^0_{h_0,I_0}(x,y) = \sum_{k=1}^{N_c(I_0)} \Pi_{h_0} u_0(x,y_k) L_k(y) \tag{13}$$

where Π_{h_0} is a Lagrange interpolation operator into V_{h_0}. The Lax Milgram lemma implies the existence of a unique sequence of solutions $\{u^n_{h_n,I_n}\}_{n=0}^N$. Based on this sequence we build a piecewise affine function $\tilde{u}$ on $[0,T]$ which equals $u^n_{h_n,I_n}$ at times t_n, $n = 0,\dots,N$, i.e.

$$\tilde{u}(t) = \frac{t_{n+1}-t}{\tau_{n+1}} u^n_{h_n,I_n} + \frac{t-t_n}{\tau_{n+1}} u^{n+1}_{h_{n+1},I_{n+1}}, \qquad t \in [t_n,t_{n+1}]. \tag{14}$$

Note that

$$\frac{\partial \tilde{u}}{\partial t} = \frac{1}{\tau_{n+1}}\big(u^{n+1}_{h_{n+1},I_{n+1}} - u^n_{h_n,I_n}\big) \qquad \text{on } (t_n,t_{n+1}].$$

With this construction, for every $n = 0,\dots,N-1$, the discretized solution belongs to the space

$$\tilde{u} \in L^2(t_n,t_{n+1};\mathbb{P}_{\tilde{I}_{n+1}} \otimes \tilde{V}_{h_{n+1}}) \subset L^2(t_n,t_{n+1};L^2_\rho(\Gamma;H^1_0(D)))$$

where $\tilde{V}_{h_{n+1}}$ is the FE space corresponding to the refined triangulation $\tilde{\mathscr{T}}_{h_{n+1}}$, see (5), and $\tilde{I}_{n+1}$ is the union of the index sets defined in (11).

4 Residual based a posteriori error estimation

In this section we will derive an a posteriori error estimate for $u - \tilde{u}$ which consists of three error contributors: space, time and stochastic. First we shall start by stating the equation satisfied by $\tilde{u}$.

From (12) it is easy see that the discretized solution $\tilde{u}$ satisfies the following equation in $(t_n,t_{n+1}]$ and for each $n = 0,\dots,N-1$

$$\int_D S_{I_{n+1}}\left[\frac{\partial \tilde{u}}{\partial t}\right] v_{h_{n+1}} + \int_D S_{I_{n+1}}\left[a\nabla\tilde{u}\right]\nabla v_{h_{n+1}} = \int_D S_{I_{n+1}}\left[f\right] v_{h_{n+1}}$$
$$+\int_D S_{I_{n+1}}\left[a\nabla\tilde{u} - a\nabla\tilde{u}^{n\theta}\right]\nabla v_{h_{n+1}} + \int_D S_{I_{n+1}}\left[f^{n\theta} - f\right] v_{h_{n+1}} \tag{15}$$
$$\forall v_{h_{n+1}} \in V_{h_{n+1}}, \text{ everywhere in } \Gamma.$$

For any element, face or edge S, h_S denotes its diameter. With every edge ($d=2$) or face ($d=3$) E, we identify a unit vector η_E orthogonal to it and denote the jump across E in direction η_E by $[\cdot]_E$. The assumption (2) ensures that the energy norm and the H_0^1 norm are equivalent for every $y \in \Gamma$, i.e. there exists $0 < c_{min} \leq c_{max}$ s.t.

$$c_{min}\|\nabla v\|_{L^2(D)} \leq \|a^{1/2}(y)\nabla v\|_{L^2(D)} \leq c_{max}\|\nabla v\|_{L^2(D)}, \quad \rho-\text{a.e. in } \Gamma$$

for any $v \in H_0^1(D)$. The constants c_{min}, c_{max} can be bounded by $c_{min} \geq \frac{1}{\sqrt{a_{min}}}$ and $c_{max} \leq \sqrt{a_{max}}$.

Now we can proceed to state the a posteriori error estimate.

Theorem 1. *Let u be the solution of* (4) *and $\tilde{u}$ be defined as in* (14)*. Then there exists a constant $C > 0$ independent of the time step, mesh size, the sparse grid index set such that*

$$\|(u-\tilde{u})(T)\|^2_{L^2_\rho(\Gamma;L^2(D))} + c^2_{min}\|u-\tilde{u}\|^2_{L^2(0,T;L^2_\rho(\Gamma;H^1_0(D)))}$$
$$\leq \|(u-\tilde{u})(0)\|^2_{L^2_\rho(\Gamma;L^2(D))} + \varepsilon^2_{spa} + \varepsilon^2_{tem} + \varepsilon^2_{sto},$$

where

$$\varepsilon^2_{spa} = \frac{C}{c^2_{min}}\sum_{n=0}^{N-1}\Lambda_{I_{n+1}}\sum_{k=1}^{N_c(I_{n+1})}$$
$$\Big(\sum_{K\in\tilde{\mathscr{T}}_{h_{n+1}}} h_K^2\left\|f(y_k) - \frac{\partial\tilde{u}}{\partial t}(y_k) + \nabla\cdot\big(a(y_k)\nabla\tilde{u}(y_k)\big)\right\|^2_{L^2(t_n,t_{n+1};L^2(K))} \tag{16}$$
$$+\sum_{E\subset\partial K} h_E\left\|\frac{1}{2}[a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\right\|^2_{L^2(t_n,t_{n+1};L^2(E))}\Big)\|L_k\|_{L^1_\rho(\Gamma)}$$

and

$$\varepsilon^2_{tem} = \frac{C}{c^2_{min}}\sum_{n=0}^{N-1}\Lambda_{I_{n+1}}\sum_{k=1}^{N_c(I_{n+1})} 2\Big(\|f(y_k) - f^{n\theta}(y_k)\|^2_{L^2(t_n,t_{n+1};L^2(D))}$$
$$+\tau_{n+1}\frac{\theta^3+(1-\theta)^3}{3}\|a(y_k)\nabla(u^{n+1}_{h_{n+1},I_{n+1}} - u^n_{h_n,I_n})\|^2_{L^2(D)}\Big)\|L_k\|_{L^1_\rho(\Gamma)} \tag{17}$$

and

$$
\begin{aligned}
\varepsilon_{sto}^2 = \frac{C}{c_{min}^2}\sum_{n=0}^{N-1}\tau_{n+1}\Big(&\sum_{i\in I_{n+1}^C\cap(I_n\cup M_{I_n})} \left\|\triangle^{m(i)}(a\nabla u_{h_n,I_n}^n)\right\|_{L^2_\rho(\Gamma;L^2(D))}^2 \\
&+ \sum_{i\in M_{I_{n+1}}} \left\|\triangle^{m(i)}(a\nabla u_{h_{n+1},I_{n+1}}^{n+1})\right\|_{L^2_\rho(\Gamma;L^2(D))}^2 \Big) \\
&+ \sum_{i\in I_{n+1}^C} \left\|\triangle^{m(i)}(f)\right\|_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma,L^2(D)))}^2 \\
&+ \frac{1}{\tau_{n+1}} \sum_{i\in \tilde{I}_{n+1}\setminus I_{n+1}} \left\|\triangle^{m(i)}(u_{h_n,I_n}^n)\right\|_{L^2_\rho(\Gamma;L^2(D))}^2 .
\end{aligned}
\tag{18}
$$

where $\Lambda_{I_{n+1}}$ denotes the Lebesgue constant corresponding to the index set I_{n+1}.

Proof. In what follows all equations hold a.e. in (t_n,t_{n+1}), $n=0,\dots,N-1$ and ρ-a.e. in Γ and we will omit the dependence on the variables x,y,t. We will start by dividing the estimate into a stochastic and a deterministic part. For every $v\in H_0^1(D)$ we have

$$
\begin{aligned}
\int_D \Big(\frac{\partial u}{\partial t} - \frac{\partial \tilde{u}}{\partial t}\Big) v + \int_D a\nabla\big(u-\tilde{u}\big)\nabla v &= \int_D fv - \int_D \frac{\partial \tilde{u}}{\partial t} v - \int_D a\nabla\tilde{u}\nabla v = \\
&= \underbrace{S_{I_{n+1}}\Big[\int_D fv - \int_D \frac{\partial \tilde{u}}{\partial t} v - \int_D a\nabla\tilde{u}\nabla v\Big]\Big\}}_{=:A_{det}} \\
&\quad + \underbrace{S_{I_{n+1}}\Big[\int_D a\nabla\tilde{u}\nabla v + \int_D \frac{\partial \tilde{u}}{\partial t} - \int_D fv\Big] - \Big(\int_D a\nabla\tilde{u}\nabla v + \int_D \frac{\partial \tilde{u}}{\partial t} - \int_D fv\Big)}_{=:A_{sto}}.
\end{aligned}
$$

We analyze A_{det} and A_{sto} separately. The term A_{det} accounts for both spatial and temporal error contribution and we can use standard techniques for a posteriori error estimation of deterministic heat equations, see [37, 36].

For any $v_{h_{n+1}}\in V_{h_{n+1}}$ we have

$$
\begin{aligned}
A_{det} &= S_{I_{n+1}}\Big[\int_D f(v-v_{h_{n+1}}) - \int_D \frac{\partial \tilde{u}}{\partial t}(v-v_{h_{n+1}}) - \int_D a\nabla\tilde{u}\nabla(v-v_{h_{n+1}})\Big] \\
&\quad + S_{I_{n+1}}\Big[\int_D f v_{h_{n+1}} - \int_D \frac{\partial \tilde{u}}{\partial t} v_{h_{n+1}} - \int_D a\nabla\tilde{u}\nabla v_{h_{n+1}}\Big] \\
&= \underbrace{S_{I_{n+1}}\Big[\int_D f(v-v_{h_{n+1}}) - \int_D \frac{\partial \tilde{u}}{\partial t}(v-v_{h_{n+1}}) - \int_D a\nabla\tilde{u}\nabla(v-v_{h_{n+1}})\Big]}_{=:A_{spa}} \\
&\quad + \underbrace{S_{I_{n+1}}\Big[\int_D \big(a\nabla\tilde{u}^{n\theta} - a\nabla\tilde{u}\big)\nabla v_{h_{n+1}}\Big] + S_{I_{n+1}}\Big[\int_D \big(f-f^{n\theta})\big) v_{h_{n+1}}\Big]}_{=:A_{tem}}
\end{aligned}
\tag{19}
$$

where in the second equality we employed the equation (15) for $\tilde{u}$. Now we have divided A_{det} into a spatial A_{spa} and a temporal A_{tem} error contributor.

For the spatial part we will follow the estimation provided in [36]. We denote by J_{h_n} any of the quasi interpolation operators of [39] defined on $H_0^1(D)$ and with values in the space of continuous, piecewise linear finite element functions corresponding to $\mathscr{T}_{h_n}$. Then, combining the interpolation error estimates of [39], a standard trace theorem [39, Lemma 3.2] and the condition 4. stated in (5), the following estimates hold for every $v \in H_0^1$ and for any element $K \in \tilde{\mathscr{T}}_{h_n}$ and interior edge/face $E \in \tilde{\mathscr{E}}_{h_n}$

$$
\begin{aligned}
&\|\nabla(v - J_{h_n} v)\|_{L^2(K)} \le \|\nabla(v - J_{h_n} v)\|_{L^2(K')} \le c_0 \|\nabla v\|_{L^2(\tilde{\omega}_K)}, \\
&\|v - J_{h_n} v\|_{L^2(K)} \le \|v - J_{h_n} v\|_{L^2(K')} \le c_1 h_{K'} \|\nabla v\|_{L^2(\tilde{\omega}_K)} \le \tilde{c}_1 h_K \|\nabla v\|_{L^2(\tilde{\omega}_K)}, \\
&\|v - J_{h_n} v\|_{L^2(E)} \le c_2 \big\{ h_E^{-1/2} \|v - J_{h_n} v\|_{L^2(K)} + h_E^{1/2} \|\nabla(v - J_{h_n} v)\|_{L^2(K)} \big\} \\
&\qquad\qquad\qquad \le \tilde{c}_2 h_E^{1/2} \|\nabla v\|_{L^2(\tilde{\omega}_K)},
\end{aligned}
\tag{20}
$$

where K' denotes the element of $\mathscr{T}_{h_n}$ that contains K and $\tilde{\omega}_K$ denotes the subset that consists of all elements of $\tilde{\mathscr{T}}_{h_n}$ sharing at least a vertex with K'. The constants c_0, c_1, c_2 only depend on the maximal ratio of the diameter of any element to the diameter of its largest inscribed ball. The constants $\tilde{c}_1, \tilde{c}_2$ in addition depend on the maximal ratio $h_{K'}/h_K$.

With η_K denoting a unit outward pointing normal we further derive

$$
\begin{aligned}
A_{spa}(y,t) = &\sum_{k=1}^{N_c(I_{n+1})} \Big[\int_D f(y_k)(v - v_{h_{n+1}}) - \int_D \frac{\partial \tilde{u}}{\partial t}(y_k)\big(v - v_{h_{n+1}}\big) \\
&- \int_D a(y_k) \nabla \tilde{u}(y_k) \nabla (v - v_{h_{n+1}}) \Big] L_k(y) \\
= &\sum_{k=1}^{N_c(I_{n+1})} \Big[\sum_{K \in \tilde{\mathscr{T}}_{h_{n+1}}} \int_K \big[f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla \cdot \big(a(y_k) \nabla \tilde{u}(y_k)\big)\big](v - v_{h_{n+1}}) \Big] \\
&- \Big[\sum_{E \in \tilde{\mathscr{E}}_{h_{n+1}}} \int_E [a(y_k) \nabla \tilde{u}(y_k) \cdot \eta_E]_E (v - v_{h_{n+1}}) \Big] L_k(y).
\end{aligned}
$$

Considering $v_{h_{n+1}} = J_{h_{n+1}}(v)$ leads us to

$$
A_{spa}(y,t) \le \sum_{k=1}^{N_c(I_{n+1})} \Big[\sum_{K \in \tilde{\mathscr{T}}_{h_{n+1}}} \tilde{c}_1 h_K \Big\| f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla \cdot \big(a(y_k) \nabla \tilde{u}(y_k)\big) \Big\|_{L^2(K)} \|\nabla v\|_{L^2(\tilde{\omega}_K)}
$$

$$+ \sum_{E \in \tilde{\mathscr{E}}_{h_{n+1}}} \tilde{c}_2 h_E^{1/2} \left\| [a(y_k)\nabla \tilde{u}(y_k) \cdot \eta_E]_E \right\|_{L^2(E)} \|\nabla v\|_{L^2(\tilde{\omega}_K)} \Big] \left|L_k(y)\right|.$$

Now, using the discrete Cauchy–Schwarz inequality and the fact that the domains $\tilde{\omega}_K$ only consist of a finite number of elements, this number being bounded by the maximal ratio of the diameter of any element to the diameter of its largest inscribed ball and on the ratios $h_{K'}/h_K$, we derive

$$\begin{aligned} A_{spa}(y,t) \le C_1 \sum_{k=1}^{N_c(I_{n+1})} \Big[\Big(\sum_{K \in \tilde{\mathscr{T}}_{h_{n+1}}} h_K^2 \left\| f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla \cdot \big(a(y_k)\nabla \tilde{u}(y_k)\big) \right\|_{L^2(K)}^2 \Big)^{1/2} \\ + \Big(\sum_{E \in \tilde{\mathscr{E}}_{h_{n+1}}} h_E \left\| [a(y_k)\nabla \tilde{u}(y_k) \cdot \eta_E]_E \right\|_{L^2(E)}^2 \Big)^{1/2} \Big] \left|L_k(y)\right| \; \|\nabla v\|_{L^2(D)} \\ = C_1 \sum_{k=1}^{N_c(I_{n+1})} \mathscr{E}_{spa,k}^{n+1}(t) \left|L_k(y)\right| \; \|\nabla v\|_{L^2(D)} \,. \end{aligned}$$

As for the temporal part A_{tem}, we proceed in a similar manner

$$\begin{aligned} A_{tem}(y,t) = \sum_{k=1}^{N_c(I_{n+1})} \Big[\int_D a(y_k) \big(\nabla \tilde{u}^{n\theta}(y_k) - \nabla \tilde{u}(y_k)\big) \nabla v_{h_{n+1}} \\ + \int_D \big(f(y_k) - f^{n\theta}(y_k)\big) v_{h_{n+1}} \Big] L_k(y) \\ \le C_2 \sum_{k=1}^{N_c(I_{n+1})} \Big[\left\| a(y_k) \nabla \big(\tilde{u}^{n\theta}(y_k) - \tilde{u}(y_k)\big) \right\|_{L^2(D)} \\ + \left\| \big(f(y_k) - f^{n\theta}(y_k)\big) \right\|_{L^2(D)} \Big] \left|L_k(y)\right| \; \|\nabla v\|_{L^2(D)} \\ = C_2 \sum_{k=1}^{N_c(I_{n+1})} \mathscr{E}_{tem,k}^{n+1}(t) \left|L_k(y)\right| \; \|\nabla v\|_{L^2(D)} \,, \end{aligned} \tag{21}$$

where C_2 depends on the $J_{h_{n+1}}$ interpolation operator norm and the Poincaré constant.

Now we focus on the term A_{sto} describing the stochastic part of the error. We will use the fact that $S_{I_{n+1}}[a\nabla u_{h_{n+1}}] = S_{I_{n+1}}[a\nabla S_{I_{n+1}}[u_{h_{n+1}}]]$ given by Proposition 1. We derive

$$A_{sto}(y,t) = S_{I_{n+1}} \Big[\int_D a\nabla \tilde{u} \nabla v + \int_D \frac{\partial \tilde{u}}{\partial t} - \int_D fv \Big] - \Big(\int_D a\nabla \tilde{u} \nabla v + \int_D \frac{\partial \tilde{u}}{\partial t} - \int_D fv \Big)$$

$$
\begin{aligned}
&= \int_D \big(S_{I_{n+1}}[a\nabla\tilde{u}] - a\nabla\tilde{u}\big)\nabla v + \int_D \big(f - S_{I_{n+1}}[f]\big)\, v \\
&\quad + \int_D \Big(\frac{S_{I_{n+1}}[u^{n+1}_{h_{n+1},I_{n+1}} - u^n_{h_n,I_n}] - (u^{n+1}_{h_{n+1},I_{n+1}} - u^n_{h_n,I_n})}{\tau_{n+1}}\Big)\, v \\
&= \int_D \big(S_{I_{n+1}}[a\nabla\tilde{u}] - a\nabla\tilde{u}\big)\nabla v + \int_D \big(f - S_{I_{n+1}}[f]\big)\, v \\
&\quad + \int_D \frac{S_{\tilde{I}_{n+1}}[u^n_{h_n,I_n}] - S_{I_{n+1}}[u^n_{h_n,I_n}]}{\tau_{n+1}}\, v \\
&\le \Big\| \sum_{i\in I^C_{n+1}} \triangle^{m(i)}(a\nabla\tilde{u}) \Big\|_{L^2(D)} \|\nabla v\|_{L^2(D)} + \Big\| \sum_{i\in I^C_{n+1}} \triangle^{m(i)}(f) \Big\|_{L^2(D)} \|v\|_{L^2(D)} \\
&\quad + \frac{1}{\tau_{n+1}} \Big\| \sum_{i\in \tilde{I}_{n+1}\setminus I_{n+1}} \triangle^{m(i)}(u^n_{h_n,I_n}) \Big\|_{L^2(D)} \|v\|_{L^2(D)} \\
&\le C_3 \Big(\Big\| \sum_{i\in (I_n\setminus I_{n+1})\cup M_{\tilde{I}_{n+1}}} \triangle^{m(i)}(a\nabla\tilde{u}) \Big\|_{L^2(D)} + \Big\| \sum_{i\in I^C_{n+1}} \triangle^{m(i)}(f) \Big\|_{L^2(D)} \\
&\quad + \frac{1}{\tau_{n+1}} \Big\| \sum_{i\in \tilde{I}_{n+1}\setminus I_{n+1}} \triangle^{m(i)}(u^n_{h_n,I_n}) \Big\|_{L^2(D)} \Big) \|\nabla v\|_{L^2(D)} \\
&= C_3\, \mathscr{E}^{n+1}_{sto}\, \|\nabla v\|_{L^2(D)}.
\end{aligned}
$$

In the last inequality we used the affine dependence of a on the random variables, stated in the assumption (3), which allows us to restrict the sum over I^C_{n+1} to the index set $M_{\tilde{I}_{n+1}} \cup (I_n \setminus I_{n+1}) = I^C_{n+1} \cap (\tilde{I}_{n+1} \cup M_{\tilde{I}_{n+1}})$. This comes from the fact that $\tilde{u} \in \mathbb{P}_{\tilde{I}_{n+1}}$ which implies $a\nabla\tilde{u} \in \mathbb{P}_{\tilde{I}_{n+1}\cup M_{\tilde{I}_{n+1}}}$. Since $S_{\tilde{I}_{n+1}\cup M_{\tilde{I}_{n+1}}}$ is exact on $\mathbb{P}_{\tilde{I}_{n+1}\cup M_{\tilde{I}_{n+1}}}$ (Proposition 1.3), we obtain

$$
\triangle^{m(i)}(a\nabla\tilde{u}) = 0, \quad \forall i \notin \tilde{I}_{n+1} \cup M_{\tilde{I}_{n+1}}.
$$

Altogether we obtained for every $n = 0, \ldots, N-1$

$$
\begin{aligned}
&\int_D \frac{\partial(u-\tilde{u})}{\partial t} v + \int_D a\nabla(u-\tilde{u})\cdot\nabla v \\
&\qquad \le C_4 \Big(\mathscr{E}^{n+1}_{sto} + \sum_{k=1}^{N_c(I_{n+1})} (\mathscr{E}^{n+1}_{tem,k} + \mathscr{E}^{n+1}_{spa,k}) \,|L_k(y)| \Big) \|\nabla v\|_{L^2(D)}.
\end{aligned}
$$

Taking $v = u - \tilde{u} \in H^1_0$ and using the Young inequality we have

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u-\tilde{u}\|^2_{L^2(D)}(y,t)+c^2_{min}\|\nabla(u-\tilde{u})\|^2_{L^2(D)}(y,t)$$
$$\leq \frac{1}{2\,c^2_{min}}C_4^2\Big(\mathscr{E}^{n+1}_{sto}(y,t)+\sum_{k=1}^{N_c(I_{n+1})}\big(\mathscr{E}^{n+1}_{tem,k}(t)+\mathscr{E}^{n+1}_{spa,k}(t)\big)\big|L_k(y)\big|\Big)^2$$
$$+\frac{c^2_{min}}{2}\|\nabla(u-\tilde{u})\|^2_{L^2(D)}(y,t)$$

which holds for a.e. $t\in(t_n,t_{n+1}]$. The last step is to integrate the last inequality w.r.t t over $(0,T)$ and w.r.t y over Γ. Using the discrete Cauchy-Schwarz inequality we derive

$$\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_\Gamma\Big(\sum_{k=1}^{N_c(I_{n+1})}\mathscr{E}^{n+1}_{tem,k}(t)\big|L_k(y)\big|\Big)^2\rho(y)\,\mathrm{d}y\,\mathrm{d}t$$
$$\leq\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_\Gamma\sum_{k=1}^{N_c(I_{n+1})}\mathscr{E}^{n+1}_{tem,k}(t)^2|L_k(y)|\sum_{k=1}^{N_c(I_{n+1})}\big|L_k(y)\big|\rho(y)\,\mathrm{d}y\,\mathrm{d}t$$
$$\leq\sum_{n=0}^{N-1}\Big(\sum_{k=1}^{N_c(I_{n+1})}\int_{t_n}^{t_{n+1}}\mathscr{E}^{n+1}_{tem,k}(t)^2\,\mathrm{d}t\int_\Gamma\big|L_k(y)\big|\rho(y)\,\mathrm{d}y\Big)\Big(\sup_{y\in\Gamma}\sum_{k=1}^{N_c(I_{n+1})}|L_k(y)|\Big)$$
$$\leq\sum_{n=0}^{N-1}\Lambda_{I_{n+1}}\Big(\sum_{k=1}^{N_c(I_{n+1})}2\big(\|f(y_k)-f^{n\theta}(y_k)\|^2_{L^2(t_n,t_{n+1};L^2(D))}$$
$$+\tau_{n+1}\frac{\theta^3+(1-\theta)^3}{3}\|a(y_k)\nabla(u^{n+1}_{h_{n+1},I_{n+1}}-u^n_{h_n,I_n})\|^2_{L^2(D)}\big)\|L_k\|_{L^1_\rho(\Gamma)}\Big),$$

where in the last inequality we employed the observation

$$\tilde{u}^{n\theta}-\tilde{u}=\big(\theta-\frac{t-t_n}{\tau_{n+1}}\big)\,(u^{n+1}-u^n).$$

Analogously for the spatial part

$$\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_\Gamma\Big(\sum_{k=1}^{N_c(I_{n+1})}\mathscr{E}^{n+1}_{spa,k}(t)\big|L_k(y)\big|\Big)^2\rho(y)\,\mathrm{d}y\,\mathrm{d}t$$
$$\leq\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_\Gamma\sum_{k=1}^{N_c(I_{n+1})}\mathscr{E}^{n+1}_{spa,k}(t)^2\big|L_k(y)\big|\sum_{k=1}^{N_c(I_{n+1})}\big|L_k(y)\big|\rho(y)\,\mathrm{d}y\,\mathrm{d}t$$
$$\leq\sum_{n=0}^{N-1}\Lambda_{I_{n+1}}\Big(\sum_{k=1}^{N_c(I_{n+1})}\Big(\sum_{K\in\tilde{\mathscr{T}}_{h_{n+1}}}h_K^2\Big\|f(y_k)-\frac{\partial\tilde{u}}{\partial t}(y_k)+\nabla\cdot\big(a(y_k)\nabla\tilde{u}(y_k)\big)\Big\|^2_{L^2(t_n,t_{n+1};L^2(K))}$$
$$+\sum_{E\subset\partial K}h_E\Big\|\frac{1}{2}[a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\Big\|^2_{L^2(t_n,t_{n+1};L^2(E))}\Big)\|L_k\|_{L^1_\rho(\Gamma)}\Big)$$

As for the stochastic part we derive

$$
\begin{aligned}
&\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_{\Gamma}\mathscr{E}_{sto}^{n+1}(y,t)^2\rho(y)\,\mathrm{d}y\,\mathrm{d}t\\
&\leq 3\sum_{n=0}^{N-1}\int_{t_n}^{t_{n+1}}\int_{\Gamma}\sum_{i\in(I_n\setminus I_{n+1})\cup M_{\tilde{I}_{n+1}}}\left\|\triangle^{m(i)}(a\nabla\tilde{u})\right\|^2_{L^2(D)}\\
&\quad+\sum_{i\in I_{n+1}^C}\left\|\triangle^{m(i)}(f)\right\|^2_{L^2(D)}+\sum_{i\in\tilde{I}_{n+1}\setminus I_{n+1}}\frac{1}{\tau_{n+1}^2}\left\|\triangle^{m(i)}(u^n_{h_n,I_n})\right\|^2_{L^2(D)}\rho(y)\,\mathrm{d}y\,\mathrm{d}t\\
&\leq C_5\sum_{n=0}^{N-1}\Big[\int_{t_n}^{t_{n+1}}\Big(\frac{t-t_n}{\tau_{n+1}}\Big)^2\mathrm{d}t\sum_{i\in I_{n+1}^C\cap(I_n\cup M_{I_n})}\left\|\triangle^{m(i)}(a\nabla u^n_{h_n,I_n})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\\
&\quad+\int_{t_n}^{t_{n+1}}\Big(\frac{t_{n+1}-t}{\tau_{n+1}}\Big)^2\mathrm{d}t\sum_{i\in M_{I_{n+1}}}\left\|\triangle^{m(i)}(a\nabla u^{n+1}_{h_{n+1},I_{n+1}})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\\
&\quad+\sum_{i\in I_{n+1}^C}\left\|\triangle^{m(i)}(f)\right\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma,L^2(D)))}\\
&\quad+\frac{\tau_{n+1}}{\tau_{n+1}^2}\sum_{i\in\tilde{I}_{n+1}\setminus I_{n+1}}\left\|\triangle^{m(i)}(u^n_{h_n,I_n})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\Big]\\
&=C_5\sum_{n=0}^{N-1}\Big[\frac{\tau_{n+1}}{3}\Big(\sum_{i\in I_{n+1}^C\cap(I_n\cup M_{I_n})}\left\|\triangle^{m(i)}(a\nabla u^n_{h_n,I_n})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\\
&\quad+\sum_{i\in M_{I_{n+1}}}\left\|\triangle^{m(i)}(a\nabla u^{n+1}_{h_{n+1},I_{n+1}})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\Big)\\
&\quad+\sum_{i\in I_{n+1}^C}\left\|\triangle^{m(i)}(f)\right\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma,L^2(D)))}\\
&\quad+\frac{1}{\tau_{n+1}}\sum_{i\in\tilde{I}_{n+1}\setminus I_{n+1}}\left\|\triangle^{m(i)}(u^n_{h_n,I_n})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\Big]
\end{aligned}
$$

□

The spatial and time estimators in (16), (17) depend on the Lebesgue constant Λ_{n+1}. The growth of the Lebesgue constant depends on the choice of the level function m and the type of the collocation points, which for the nested Clenshaw-Curtis points yields an estimate $\Lambda_I \sim |I|^2$ and for the projected Leja points an estimate $\Lambda_I \sim |I|^{3+\varepsilon}$ for any $\varepsilon > 0$ (see [15]). Such estimation can cause the estimator to be too conservative. This issue was addressed in [25, Rem 4.4]. The following theorem provides an alternative way of estimating the spatial and time estimator without involving the Lebesgue constant and is an extension of the results from [25, Rem 4.4].

Theorem 2. *The spatial estimator* ε_{spa} *from* (16) *and the time estimator* ε_{tem} *from* (17) *can be alternatively expressed as*

$$\varepsilon_{tem}^2 = \frac{C}{c_{min}^2} \sum_{n=0}^{N-1} \Big[\Big\| \sum_{k=1}^{N_c(I_{n+1})} \Big[\big(f(y_k) - f^{n\theta}(y_k) \big) \Big] L_k(y) \Big\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma;L^2(D)))} + \tau_{n+1} \Big\| \sum_{k=1}^{N_c(I_{n+1})} \Big[a(y_k) \nabla \big(u^{n+1}_{h_{n+1},I_{n+1}}(y_k) - u^n_{h_n,I_n}(y_k) \big) \Big] L_k(y) \Big\|^2_{L^2_\rho(\Gamma;L^2(D))} \Big] \tag{22}$$

and

$$\varepsilon_{spa}^2 = \sum_{n=0}^{N-1} \sum_{K \in \tilde{\mathscr{T}}_{h_{n+1}}} (\varepsilon^n_{spa,K})^2 \tag{23}$$

with

$$(\varepsilon^n_{spa,K})^2 = \frac{C}{c_{min}^2} h_K^2 \Big\| \sum_{k=1}^{N_c(I_{n+1})} \Big[f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla \cdot \big(a(y_k) \nabla \tilde{u}(y_k) \big) \Big] L_k(y) \Big\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma;L^2(K)))} + \sum_{E \subset \partial K} h_E \Big\| \sum_{k=1}^{N_c(I_{n+1})} \Big(\frac{1}{2} [a(y_k) \nabla \tilde{u}(y_k) \cdot \eta_E]_E \Big) L_k(y) \Big\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma;L^2(E)))}.$$

Proof. We follow by estimating the term A_{tem} from (21) by

$$\begin{aligned} A_{tem}(y,t) &= \int_D \sum_{k=1}^{N_c(I_{n+1})} a(y_k) \big(\nabla \tilde{u}^{n\theta}(y_k) - \nabla \tilde{u}(y_k) \big) L_k(y) \nabla v_{h_{n+1}} \\ &\quad + \int_D \sum_{k=1}^{N_c(I_{n+1})} \big(f(y_k) - f^{n\theta}(y_k) \big) L_k(y)\, v_{h_{n+1}} \\ &\le C \Big(\Big\| \sum_{k=1}^{N_c(I_{n+1})} a(y_k) \big(\nabla \tilde{u}^{n\theta}(y_k) - \nabla \tilde{u}(y_k) \big) L_k(y) \Big\|_{L^2(D)} \\ &\quad + \Big\| \sum_{k=1}^{N_c(I_{n+1})} \big(f(y_k) - f^{n\theta}(y_k) \big) L_k(y) \Big\|_{L^2(D)} \Big) \|\nabla v\|_{L^2(D)} \\ &= C \mathscr{E}^{n+1}_{tem} \|\nabla v\|_{L^2(D)} \end{aligned}$$

where we applied the same interpolation results as proposed in (20). Analogously for the spatial estimation we derive

$$A_{spa}(y,t) = \sum_{K \in \tilde{\mathscr{T}}_{h_{n+1}}} \Big[\int_K \sum_{k=1}^{N_c(I_{n+1})} \big[f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla \cdot \big(a(y_k) \nabla \tilde{u}(y_k) \big) \big] L_k(y) (v - v_{h_{n+1}}) \Big]$$

$$- \sum_{E\in\tilde{\mathscr{E}}_{h_{n+1}}} \int_E \sum_{k=1}^{N_c(I_{n+1})} [a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\, L_k(y)(v - v_{h_{n+1}})$$

$$\leq \sum_{K\in\tilde{\mathscr{T}}_{h_{n+1}}} \left\| \sum_{k=1}^{N_c(I_{n+1})} \left[f(y_k) - \frac{\partial\tilde{u}}{\partial t}(y_k) + \nabla\cdot\big(a(y_k)\nabla\tilde{u}(y_k)\big)\right] L_k(y) \right\|_{L_2(K)} \left\| v - v_{h_{n+1}} \right\|_{L^2(K)}$$

$$+ \sum_{E\in\tilde{\mathscr{E}}_{h_{n+1}}} \left\| \sum_{k=1}^{N_c(I_{n+1})} [a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\, L_k(y) \right\|_{L^2(E)} \left\| v - v_{h_{n+1}} \right\|_{L^2(E)}.$$

Using again the interpolation results (20) and the discrete Cauchy-Schwarz inequality we obtain

$$A_{spa}(y,t) \leq C\Bigg[\Bigg(\sum_{K\in\tilde{\mathscr{T}}_{h_{n+1}}} h_K^2 \left\| \sum_{k=1}^{N_c(I_{n+1})} \left[f(y_k) - \frac{\partial\tilde{u}}{\partial t}(y_k) + \nabla\cdot\big(a(y_k)\nabla\tilde{u}(y_k)\big)\right] L_k(y) \right\|_{L_2(K)}^2 \Bigg)^{1/2}$$

$$+ \Bigg(\sum_{E\in\tilde{\mathscr{E}}_{h_{n+1}}} h_E \left\| \sum_{k=1}^{N_c(I_{n+1})} [a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\, L_k(y) \right\|_{L^2(E)} \Bigg)^{1/2}\Bigg] \|\nabla v\|_{L^2(D)}$$

$$= C\mathscr{E}_{spa}^{n+1} \|\nabla v\|_{L^2(D)}$$

The rest of the proof follows the same steps as in the proof of Theorem 1. □

Remark 1. Note that the spatial estimator from Theorem 1 allows for different FE meshes for different collocation points. This property is sacrificed in the spatial estimator from Theorem 2. We shall also note that the error estimator derived in this work is of suboptimal order in the case $\theta = 1/2$, which corresponds to the Crank-Nicolson scheme. In order to restore the second order convergence one shall work with a piecewise quadratic polynomial function in time instead of the linear one defined in (14). We refer an interested reader to [30, 2].

5 Adaptive algorithm

The estimators from the preceding section provide us with an upper bound of the error that is naturally localized in all variables - time, space and stochastics. There are many possible choices of adaptive algorithms that can be constructed starting from these estimators. One could drive the adaptive choice of time-varying finite element and stochastic grids by a local in time error estimator, as was proposed in [33, 9] for time varying FE or DG meshes in the case of a deterministic heat equation. Also, the spatial estimator ε_{spa} in Theorem 1 is naturally localized over the collocation points so it allows for different adapted FE meshes in different collocation points. This idea has been explored e.g. in [18] in the context of a stochastic Galerkin polynomial chaos approximation of an elliptic problem with random coefficients. There are, however, many problems whose behaviour does not require FE meshes and sparse grids that dramatically change in time. Considering fixed in time FE meshes and sparse grids simplifies the estimators and the adaptive process. In this work we will restrict to adapted FE meshes and sparse grids which are fixed in time with the goal to obtain the overall error

$$\varepsilon = \|(u-\tilde{u})(T)\|^2_{L^2_\rho(\Gamma;L^2(D))} + c^2_{min}\|u-\tilde{u}\|^2_{L^2(0,T;L^2_\rho(\Gamma;H^1_0(D)))}$$

under a prescribed tolerance TOL. We will apply the global spatial and time estimators from Theorem 2, i.e. (22), (23) by localizing the spatial estimator into elements, the time estimator into time steps and the stochastic estimator (18) into indices. For a deterministic right hand side the corresponding error estimators become

$$\begin{aligned}\varepsilon^2_{spa,K} = \frac{c_1}{c^2_{min}} h^2_K \left\| \sum_{k=1}^{N_c(I)} \Big[f(y_k) - \frac{\partial \tilde{u}}{\partial t}(y_k) + \nabla\cdot\big(a(y_k)\nabla\tilde{u}(y_k)\big)\Big] L_k(y) \right\|^2_{L^2(0,T;L^2_\rho(\Gamma;L^2(K)))} \\ + \frac{c_2}{c^2_{min}} \sum_{E\subset\partial K} h_E \left\| \sum_{k=1}^{N_c(I)} \Big(\frac{1}{2}[a(y_k)\nabla\tilde{u}(y_k)\cdot\eta_E]_E\Big) L_k(y)\right\|^2_{L^2(0,T;L^2_\rho(\Gamma;L^2(E)))}\end{aligned} \tag{24}$$

for every element $K \in \mathscr{T}_h$,

$$\begin{aligned}\varepsilon^2_{tem,n} = \frac{c_3}{c^2_{min}} \left\| \sum_{k=1}^{N_c(I)} \Big[\big(f(y_k) - f^{n\theta}(y_k)\big)\Big] L_k(y)\right\|^2_{L^2(t_n,t_{n+1};L^2_\rho(\Gamma;L^2(D)))} \\ + \frac{c_4}{c^2_{min}} \tau_{n+1} \left\| \sum_{k=1}^{N_c(I)} \Big[a(y_k)\nabla\big(u^{n+1}_{h,I}(y_k) - u^n_{h,I}(y_k)\big)\Big] L_k(y)\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\end{aligned} \tag{25}$$

for every subinterval $[t_n, t_{n+1}]$, $n = 0, \ldots, N-1$, and

$$\varepsilon^2_{sto,i} = \frac{1}{c^2_{min}} \sum_{n=0}^{N-1} \tau_{n+1} \Big(\left\| \triangle^{m(i)}(a\nabla u^n_{h,I})\right\|^2_{L^2_\rho(\Gamma;L^2(D))} + \left\| \triangle^{m(i)}(a\nabla u^{n+1}_{h,I})\right\|^2_{L^2_\rho(\Gamma;L^2(D))}\Big) \tag{26}$$

for every multi index $i \in M_I$.

Then the overall error ε can be bounded by

$$\varepsilon^2 \le \sum_{K\in\mathscr{T}_h} \varepsilon^2_{spa,K} + \sum_{n=0}^{N-1} \varepsilon^2_{tem,n} + \sum_{i\in M_I} \varepsilon^2_{sto,i}.$$

The algorithm will start with fairly coarse grids and index set $\mathscr{T}_h$, τ, I, compute the numerical solution $\tilde{u}$ and compute the estimators (24), (25), (26) for every cell, time subinterval and index from the margin. Let $\mathcal{N} = |\mathcal{T}_h| + N + |M_I|$ denote the total number of elements in $\{\mathcal{T}_h, \tau, M_I\}$, i.e. number of cells + number of subintervals (N) + number of indices in the margin M_I. Then we will refine a cell K whenever $\varepsilon^2_{spa,K} > (\alpha TOL/\mathcal{N})^2$, divide a time interval $[t_n, t_{n+1}]$ into 2 equal subintervals whenever $\varepsilon^2_{tem,n} > (\alpha TOL/\mathcal{N})^2$ and add an index $i \in M_I$ into the index set I whenever $\varepsilon^2_{sto,i} > (\alpha TOL/\mathcal{N})^2$, where $\alpha > 1$. Note that adding an index i might result in adding more indices since we need to keep the index set I downward closed. With the new refined mesh, time grid and sparse grid we need to compute a new solution $\tilde{u}$ and continue until the stopping criterion

$$\varepsilon^2_{\mathcal{T}_h,\tau,I} := \sum_{K \in \mathcal{T}_h} \varepsilon^2_{spa,K} + \sum_{n=0}^{N-1} \varepsilon^2_{tem,n} + \sum_{i \in M_I} \varepsilon^2_{sto,i} < TOL^2$$

is satisfied. This procedure is described in Algorithm 1. We shall note that there is no proof of convergence for this algorithm.

Algorithm 1: Adaptive algorithm

```
Data: TOL > 0
Result: τ, I, 𝒯_h and ũ s.t. ε_{𝒯_h,τ,I} < TOL
Initialize τ, I, 𝒯_h;
compute ũ on τ, I, 𝒯_h;
compute ε_{spa,K}, ε_{tem,n}, ε_{sto,i};
while ε_{𝒯_h,τ,I} ≥ TOL do
    set 𝒩 = |𝒯_h| + N + |M_I|;
    for K ∈ 𝒯_h do
        if ε_{spa,K} > α TOL/𝒩 then
            refine K
        end
    end
    for n ∈ {0, ..., N−1} do
        if ε_{tem,n} > α TOL/𝒩 then
            refine [t_n, t_{n+1}]
        end
    end
    for i ∈ M_I do
        if ε_{sto,i} > α TOL/𝒩 then
            I = I ∪ i;
            add indices s.t. I is downward closed
        end
    end
    update τ, I, 𝒯_h;
    compute ũ on new τ, I, 𝒯_h;
    compute ε_{spa,K}, ε_{tem,n}, ε_{sto,i};
end
```

6 Numerical results

This section is dedicated to study the effectiveness of the estimators in (24), (25), (26) and the performance of the adaptive algorithm introduced in Section 5. The practical computation of these estimators requires some estimation of the constants $c_1, \dots, c_4, c_{min}$ as well as an approximate computation of the $L^2_\rho(\Gamma)$ norm. This is discussed hereafter.

Let us consider problem (1) set in a unit square $D = [0,1]^2$ with time domain $[0,1]$ and an uncertain diffusion coefficient

$$a(x,y) = a_0 + \sum_{m=1}^{2} \frac{cos(2\pi m x_1) + cos(2\pi m x_2)}{(\pi m)^2} y_m \qquad (27)$$

with $x = (x_1, x_2)$, $y = (y_1, y_2)$ and $a_0 > 0$ set to satisfy

$$\inf_{x \in D, y \in \Gamma} a(x,y) = 0.01.$$

The random variables are independent and uniformly distributed $y_m \sim U([-1,1])$ and the forcing term is deterministic and time-independent

$$f(x) = 20\, \mathbb{1}_F(x)$$

with $F = [0.4, 0.6] \times [0.4, 0.6]$ a square in the middle of the domain.

In all of our simulations we used the spatial and time estimators provided in Theorem 2 which do not require an explicit estimation of the Lebesgue constant. The norm $\|g\|_{L^2_\rho(\Gamma)}$ for $g \in C^0(\Gamma)$ is approximated using a set $\Theta \subset \Gamma$ of finite cardinality by

$$\|g\|_{L^2_\rho(\Gamma)} \approx \Big(\frac{1}{|\Theta|} \sum_{y \in \Theta} g(y)^2 \Big)^{1/2}.$$

We set Θ to consist of 500 randomly sampled points in $\Gamma = [-1,1]^2$ according to the distribution ρ, uniform on Γ. As suggested in [25], instead of setting $c_{min} = \sqrt{a_{min}}$, which may be too conservative, we will rather approximate it by

$$c_{min} := \min_{v \in U \subset L^2_\rho(\Gamma; H^1_0(D))} \min_{y \in \Xi} \frac{\|a^{1/2}(y) \nabla v(y)\|_{L^2(D)}}{\|\nabla v(y)\|_{L^2(D)}},$$

where we take $U = \{u^n_{h,I},\ n = 0, \dots, N\}$ and Ξ is a set of random samples of small cardinality (different from Θ). For the specific diffusion coefficient in (27) we estimated $c_{min} \approx 0.41$. The norm $\|g\|_{L^2(0,T)}$ is computed using the trapezoidal rule as suggested in [33]. We have considered $P1$ finite elements without fitting the FE mesh to the subdomain F and $\theta = 1$, namely the implicit Euler method. The sparse grid consists of Leja points built as symmetric Leja sequences within $[-1,1]$ (see e.g. [15]) with level function $m(i) = i$. This combination satisfies the interpolatory condition (10). For a sharp behaviour of the estimators and an efficient performance of the adaptive algorithm one needs a good estimation of the constants c_1, c_2, c_3, c_4 in (24), (25), (26). This requires a good estimation of the interpolation constants from (20) which is not an easy task and we refer the reader to [39] for ways to bound the interpolation constants. In our case, we adopted the strategy proposed in [33], i.e. estimated the constants by observing the behaviour of the estimators vs. the behaviour of the true error when refining individually uniform spatial grids, uniform time grids and isotropic sparse grids for different solutions u. This is done on relatively coarse FE meshes, sparse grids and time discretizations so that the overall cost of estimating the constants is much smaller than the cost of the adaptive process. We obtained the estimates $c_1 = 0.016, c_2 = 0.023, c_4 = 0.078$. The term including the constant c_3 is in our case equal to 0. Our simulations were done using the FEniCS library [3].

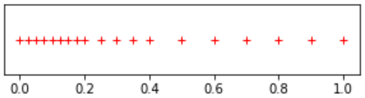
6.1 Numerical study of the performance of the estimators

This part is dedicated to study the effectiveness of the error estimator considering different non uniform FE meshes, time discretizations and index sets. We proceed by studying first a "marginalized" error and estimator, where by "marginalized" spatial error we mean an error caused by only spatial discretization, i.e. the numerical solution and the "true" solution are computed using the same ("overkilling") discretization for time and random variables. Analogously, for the marginalized time and stochastic errors.
In Figure 1 we show the convergence results for the marginalized time estimator. The numerical and true solution were computed on a uniform spatial grid consisting of 6400 triangles, with diameter 0.025 and a sparse grid having 113 collocation points. We considered both uniform and non uniform time discretizations for the numerical solution specified in Figure 1 (left). The true solution was computed on a much finer time grid. Figure 1 (right) shows the "true" error as well as the error estimator over the sequence of time grids obtained by refinement of the three grids shown in the left plot.

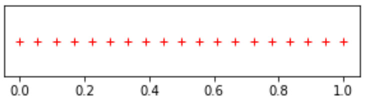
The convergence study of the stochastic estimator was performed on a triangulation with 6400 triangles with diameter 0.025 and with 200 uniform time steps. The results are shown in Figure 2. We considered sequences of anisotropic sparse grids with the index sets defined as

$$I(w) = \{i \in \mathbb{N}_+^M : \sum_n \beta_n (i_n - 1) \leq w\}$$

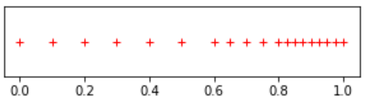
with $w = 1, \ldots, 8$ and $\beta = (\beta_1, \ldots, \beta_M), \beta_m \geq 1$. The weights β were fixed to $(1,2), (1,1), (2,1)$. Examples of such sparse grids with $w = 5$ can be seen in Figure 2 (left). The reference solution was computed with $w = 15$. Figure 2 (right) shows



(a) Time discretization 1



(b) Time discretization 2



(c) Time discretization 3

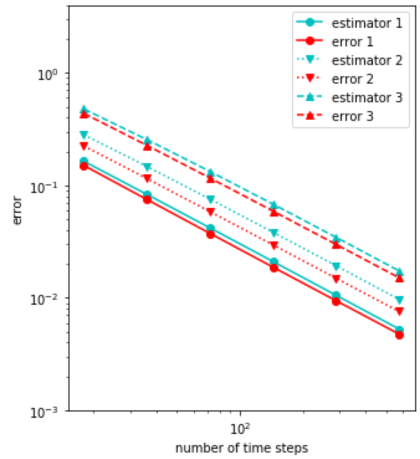
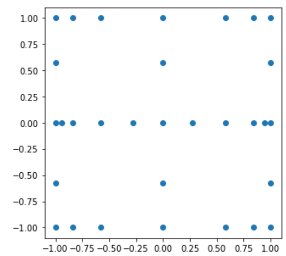


Fig. 1: Time error and estimator with respect to the number of time steps (right) for solutions computed on refinements of 3 time grids (left).

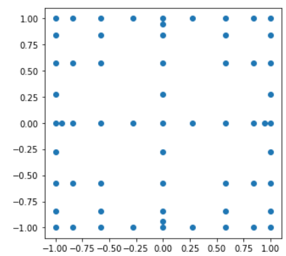
the error and the estimator for the 3 considered choices of sparse grids.

Concerning the spatial estimator, we fixed the number of collocation points to 113 built as an isotropic sparse grid, the number of uniform time steps to 200 and computed the numerical solution on non uniform triangulations specified in Figure 3 (left). The convergence in Figure 3 was achieved by uniformly refining every cell, i.e. halving the diameter of every cell at each iteration of the convergence study with the use of refinement by longest edge bisection [12].

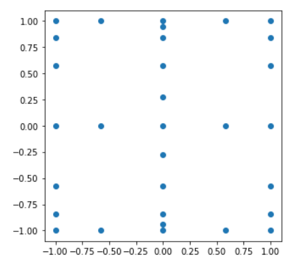
We now focus on the total error and consider several combinations of uniform refinements in the different components (spatial, temporal, stochastic). We report in Table 1 the behaviour of the estimator in all cases. From these results we conclude that the three components of the estimator behave in a fairly independent way. The only dependency we can observe is the stochastic estimator being dependent on the spatial discretization. If the stochastic error is negligible compared to the spatial error, the stochastic estimator grows as the spatial estimator decreases while refining the spatial grid. When they reach a similar magnitude, decreasing the spatial error does not influence the stochastic estimator anymore. All the numerical solutions were computed on uniform triangulations, uniform time grids and isotropic sparse grids.



(a) $\beta = (1,2)$, $w = 5$



(b) $\beta = (1,1)$, $w = 5$



(c) $\beta = (2,1)$, $w = 5$

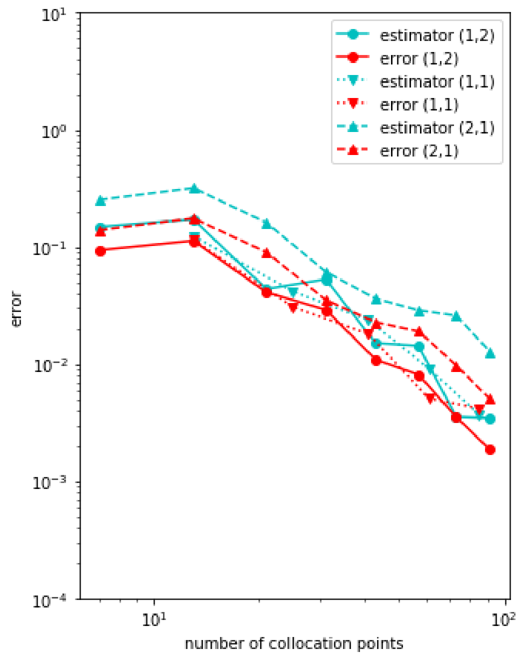
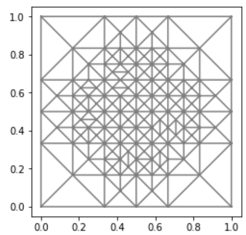


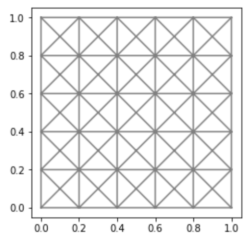
Fig. 2: Stochastic error and estimator with respect to the number of collocation points (right) for solutions computed on anisotropic sparse grids (left) of levels $w = 1, \ldots, 8$.

6.2 Numerical study of the performance of the adaptive algorithm

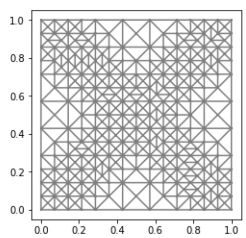
In this part we study the performance of the Algorithm 1 applied to problem (27). We set the tolerance to $TOL = 0.1$, $\alpha = 1.5$ and initialize the spatial grid as a uniform triangulation having 25 points and 100 triangles. The initial time discretization was set to have 25 equally spaced subintervals and the initial sparse grid was isotropic with 13 collocation points built over the index set $I = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}$. The initial discretizations are depicted in Figure 4 (left) with their corresponding error indicators (right). In Figure 5 (left) we show the final grids, the spatial triangulation having 7490 triangles, the time grid consisting of 155 steps and the stochastic sparse grid having 57 collocation points. As we can see, the algorithm was able to detect



(a) Triangulation 1



(b) Triangulation 2



(c) Triangulation 3

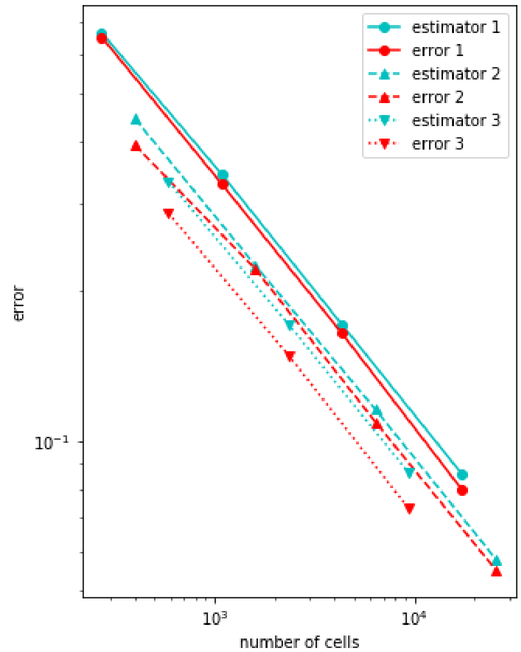
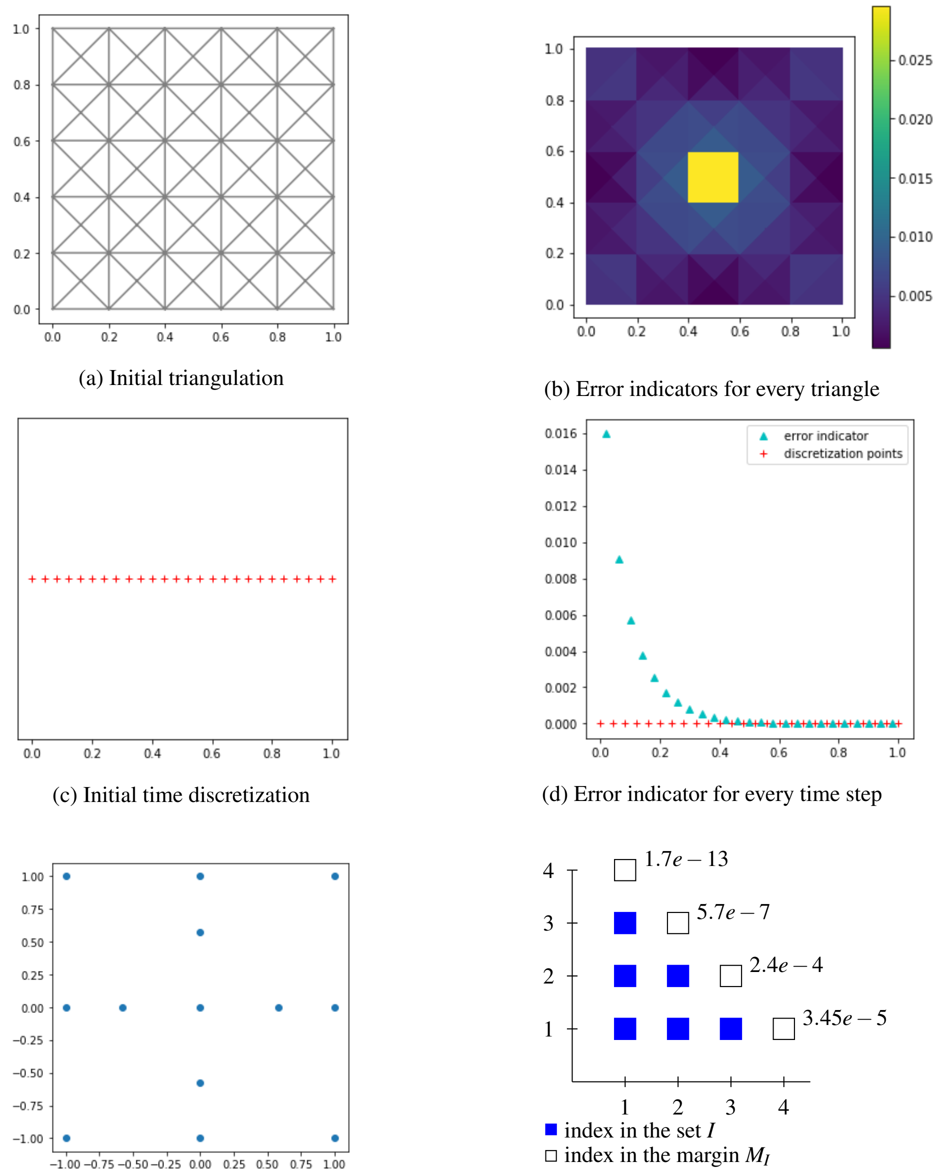


Fig. 3: Spatial error and estimator with respect to the number of cells (right) for solutions computed on refinements of 3 triangulations (left).

the location and discontinuity of the forcing term. The final time discretization is clearly consistent with the dissipative behaviour of this problem and the algorithm is also able to identify the dominant random variable Y_1. In Figure 5 (right) we can as well observe that the estimator provides a good control over the error throughout the whole process. As a last result we report in Table 2 the number of cells, time steps and collocation points in the final discretizations for different tolerances. We can see that halving the tolerance results in approximately twice more time steps and four times more cells which agrees with the expected order of convergence.



(a) Initial triangulation

(b) Error indicators for every triangle

(c) Initial time discretization

(d) Error indicator for every time step

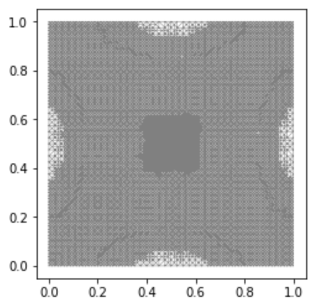
(e) Initial sparse grid

(f) Error indicator for every index in the margin

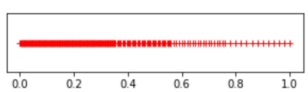
Fig. 4: Initial discretizations (left) when running the Algorithm 1 applied to problem (27) and corresponding error indicators (right) for elements, time intervals and multi-indices.

Table 1: Error and error estimation when using different combinations of uniform triangulations, uniform time steps and isotropic sparse grids

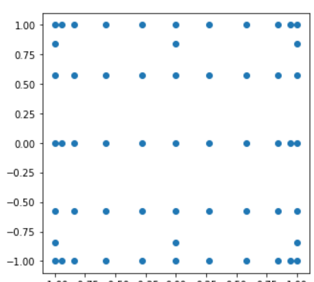
h	τ	coll. pts	ε_{spa}	ε_{tem}	ε_{sto}	$\varepsilon_{\mathscr{T}_h,\tau,I}$	ε
0.2	0.025	13	0.73	0.13	0.017	0.74	0.66
0.1	0.025	13	0.38	0.13	0.1	0.42	0.39
0.04	0.025	13	0.16	0.13	0.18	0.27	0.2
0.02	0.025	13	0.08	0.13	0.2	0.25	0.16
0.01	0.025	13	0.04	0.13	0.2	0.24	0.157
0.02	0.05	13	0.08	0.26	0.2	0.33	0.22
0.02	0.025	13	0.08	0.13	0.2	0.25	0.16
0.02	0.0125	13	0.081	0.068	0.2	0.22	0.15
0.02	0.025	5	0.08	0.13	0.34	0.37	0.18
0.02	0.025	13	0.08	0.13	0.2	0.25	0.16
0.02	0.025	25	0.08	0.13	0.07	0.15	0.11



(a) Final triangulation



(b) Final time discretization



(c) Final sparse grid

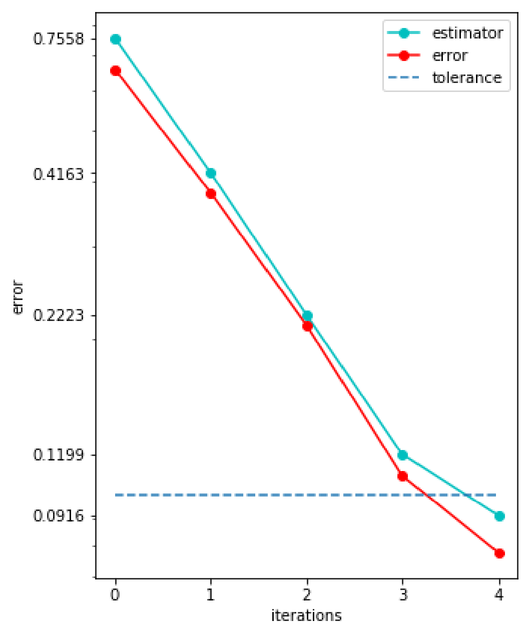


Fig. 5: Final discretizations (left) resulting from the Algorithm 1 applied to problem (27) and the evolution of the overall error and error estimation (right).

Table 2: Number of cells, time steps and collocation points when using the Algorithm 1 with different tolerances.

TOL	no. of cells	no. of time steps	no. of coll. points	$\varepsilon_{\mathscr{T}_h,\tau,I}$
0.4	492	36	17	0.3676
0.2	1808	70	33	0.1927
0.1	7490	155	57	0.0947
0.05	29106	357	89	0.0487

7 Conclusion

We have derived a residual based a posteriori error estimation for a heat equation with a random forcing term and a random diffusion coefficient dependent on a finite number of independent random variables. The dependency of the diffusion coefficient is, moreover, assumed to be affine. This problem was discretized by a θ-scheme in time, FEM in physical space and sparse grid collocation method in stochastic space which required the use of nested collocation points. The estimate consisted of three parts accounting for the FEM error, time discretization error and the stochastic error, respectively. The derivation is valid for the case of time-varying FE meshes and time-varying sparse grids allowing for both refinement and coarsening. We proposed an adaptive algorithm for the choice of time discretization, FE mesh and sparse grid, where the mesh and sparse gird are fixed in time which simplifies the computation of the estimators. The estimators are localized on each element of the FE mesh, each time step and each index from the margin of the sparse grid index set and we perform a refinement whenever the localized estimate is higher than a prescribed condition (see Algorithm 1). We studied the effectiveness of the estimators over non-uniform time discretizations, non-uniform meshes and anisotropic sparse grids, fixed in time and applied the adaptive algorithm to a problem with a deterministic, time independent forcing term. This algorithm is one possible strategy. Several other versions could be considered as well, for instance, to allow for coarsening in the adaptive process for a more uniform distribution of the error. We believe that the derived error estimates could provide a reliable basis for error estimation and adaptation strategies that include time-varying FE meshes and sparse grids. One could, for example, drive an adaptive choice of time-varying meshes and sparse girds by localizing the spatial and stochastic estimator for a specific time step, as was proposed in [33, 9] for time varying FE or DG meshes in the case of a deterministic heat equation.

Acknowledgements

This work has been supported by the Swiss National Science Foundation under the Project n. 172678 “Uncertainty Quantification Techniques for PDE constrained optimization and random evolution equations". The authors also acknowledge the support from the Center for ADvanced MOdeling Science (CADMOS).

References

1. M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods. *Numerische Mathematik*, 65(1):23–50, 1993.
2. G. Akrivis, C. Makridakis, and R. H. Nochetto. A posteriori error estimates for the Crank-Nicolson method for parabolic equations. *Math. Comput.*, 75:511–531, 2006.
3. M. S. Alnæs, J. Blechta, J. Hake, A. Johansson, B. Kehlet, A. Logg, C. Richardson, J. Ring, M. E. Rognes, and G. N. Wells. The FEnics project version 1.5. *Archive of Numerical Software*, 3(100), 2015.
4. I. Babuška, R. Duran, and R. Rodriguez. Analysis of the efficiency of an a posteriori error estimator for linear triangular finite elements. *Siam Journal on Numerical Analysis - SIAM J NUMER ANAL*, 29, 1992.
5. 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, 2010.
6. I. Babuška and W. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM Journal on Numerical Analysis*, 15(4):736–754, 1978.
7. I. Babuška and W. Rheinboldt. A posteriori error estimates in the finite element method. *Int. J. Numer. Methods Engrg.*, 12:1597–1615, 1978.
8. 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. Rønquist, editors, *Spectral and High Order Methods for Partial Differential Equations*, pages 43–62, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
9. G. Bangerth and R. Rannacher. *Adaptive Finite Element Methods for Solving Differential Equations*. Birkhäuser, Basel, 2003.
10. V. Barthelmann, E. Novak, and K. Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12(4):273–288, Mar 2000.
11. J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(9), 2012.
12. C. Bedregal and M.-C. Rivara. Longest-edge algorithms for size-optimal refinement of triangulations. *Computer-Aided Design*, 46:246 – 251, 2014. 2013 SIAM Conference on Geometric and Physical Modeling.
13. A. Bespalov, C. Powell, and D. Silvester. Energy norm a posteriori error estimation for parametric operator equations. *SIAM Journal on Scientific Computing*, 36(2):A339–A363, 2014.
14. A. Bespalov and F. Xu. A posteriori error estimation and adaptivity in stochastic Galerkin fem for parametric elliptic PDEs: beyond the affine case. *arXiv:1903.06520*, 2019.
15. A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Foundations of Computational Mathematics*, 14(4):601–633, Aug 2014.
16. A. J. Crowder, C. E. Powell, and A. Bespalov. Efficient adaptive multilevel stochastic Galerkin approximation using implicit a posteriori error estimation. *arXiv:1806.05987*, 2018.
17. R. Dautray, A. Craig, M. Artola, M. Cessenat, J. Lions, and H. Lanchon. *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 5 Evolution Problems I*. Mathematical Analysis and Numerical Methods for Science and Technology. Springer Berlin Heidelberg, 1999.
18. M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander. Adaptive stochastic Galerkin fem. *Computer Methods in Applied Mechanics and Engineering*, 270:247 – 269, 2014.

19. M. Eigel, C. J. Gittelson, C. Schwab, and E. Zander. Residual-based a posteriori error estimation for stochastic Galerkin finite element methods. *CMAME*, 11 2013.
20. Eigel, Martin, Gittelson, Claude Jeffrey, Schwab, Christoph, and Zander, Elmar. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM: M2AN*, 49(5):1367–1398, 2015.
21. K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems i: A linear model problem. *SIAM Journal on Numerical Analysis*, 28(1):43–77, 1991.
22. K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems ii: Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM Journal on Numerical Analysis*, 32(3):706–740, 1995.
23. R. Ghanem and P. D Spanos. *Stochastic Finite Element: a Spectral Approach*, volume 224. Springer, New York, 01 2003.
24. D. Guignard. Partial differential equations with random input data: A perturbation approach. *Archives of Computational Methods in Engineering*, 09 2018.
25. D. Guignard and F. Nobile. A posteriori error estimation for the stochastic collocation finite element method. *SIAM Journal on Numerical Analysis*, 56(5):3121–3143, 2018.
26. C. Johnson. Adaptive finite element methods for diffusion and convection problems. *Computer Methods in Applied Mechanics and Engineering*, 82(1):301 – 322, 1990. Proceedings of the Workshop on Reliability in Computational Mechanics.
27. A. Khan, A. Bespalov, C. E. Powell, and D. Silvester. Robust a posteriori error estimation for stochastic Galerkin formulations of parameter-dependent linear elasticity equations. *arXiv:1810.07440*, 2018.
28. A. Khan, C. E. Powell, and A. Bespalov. Robust error estimation for lowest-order approximation of nearly incompressible elasticity. *arXiv:1801.04122*, 2018.
29. O. Le Maître and O. Knio. *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer Netherlands, 01 2010.
30. A. Lozinski, M. Picasso, and V. Prachittham. An anisotropic error estimator for the Crank–Nicolson method: Application to a parabolic problem. *SIAM Journal on Scientific Computing*, 31(4):2757–2783, 2009.
31. F. Nobile, R. Tempone, and C. Webster. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2411–2442, 2008.
32. F. Nobile, R. Tempone, and C. Webster. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 46(5):2309–2345, 2008.
33. M. Picasso. Adaptive finite elements for a linear parabolic problem. *Computer Methods in Applied Mechanics and Engineering*, 167(3-4):223–237, 1998.
34. S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Dokl. Akad. Nauk SSSR*, 148(5):1042–1045, 1963. cited By 108.
35. J. van Neerven. *Stochastic Evolution Equations - Lecture Notes of the 11th Internet Seminar, 2007/08*. 01 2008.
36. R. Verfürth. A posteriori error estimates for finite element discretizations of the heat equation. *CALCOLO*, 40(3):195–212, Dec 2003.
37. R. Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods*. A Posteriori Error Estimation Techniques for Finite Element Methods. OUP Oxford, 2013.
38. R. Verfürth. A posteriori error estimates for nonlinear problems. finite element discretizations of elliptic equations. *Mathematics of Computation*, 62(206):445–475, 1994.
39. R. Verfürth. Error estimates for some quasi-interpolation operators. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(4):695–713, 1999.
40. D. Xiu and J. S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005.