Probabilistic methods for elliptic partial differential equations

Assyr Abdulle* Giacomo Garegnani*

Abstract.

AMS subject classification.

Keywords.

1 Introduction

TO DO [1, 3, 6, 14, 15]

$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) &= f, \quad \text{in } D, \\ u &= g, \quad \text{on } \partial D. \end{aligned} \tag{1.1}$$

important: review of probabilistic methods for PDEs and ODEs. Have PDEs really been treated already? How? Inverse problems: what is the current state of things? Has anyone gone infinite dimensional?

2 Notation and method

2.1 Notation

We now introduce some basic notation which will be employed throughout this paper. Most of the notation is classic, but we report it here for completeness. The symbol $D \subset \mathbb{R}^2$ is employed for a bounded domain with smooth boundary, or for a convex polygon. The following symbols are employed for function spaces

- $L^p(D) = \{v \colon D \to \mathbb{R}, \int_D v^p \,\mathrm{d}x < \infty\}$,
- $W^{q,p}(D) = \Big\{v \in L^p(D), \sum_{|\alpha| \leq q} |D^\alpha v| \in L^p(D)\Big\}$,
- $H^q(D) \equiv W^{q,2}(D)$,
- $H_0^q(D) = \{v \in H^q(D), v\big|_{\partial D} = 0\}$,
- $C_0^l(D) = \{v \in \mathcal{C}^l(D), v\big|_{\partial D} = 0\}$.

For a function $v \in \mathcal{X}$ where $\mathcal{X}$ is any of the spaces above, we denote by $\|v\|_{\mathcal{X}}$ and $|v|_{\mathcal{X}}$ the usual norms and seminorms. Furthermore, for L^p and $W^{q,p}$, the usual meaning is given for $p = \infty$. For a vector $x \in \mathbb{R}^2$ we denote simply by $|x|$ its Euclidean norm. Moreover, we will employ the following symbols

*Institute of Mathematics, École Polytechnique Fédérale de Lausanne ({assyr.abdulle, giacomo.garegnani}@epfl.ch)

- $\mathcal{T}_h$ is a triangulation of D satisfying assumption, and V_h is the space of linear finite elements with zero boundary conditions defined on $\mathcal{T}_h$,
- $\widetilde{\mathcal{T}}_h$ is a perturbation of $\mathcal{T}_h$ as for Definition 2.1 such that Assumption 2.3 holds, and $\widetilde{V}_h$ is the space of linear finite elements with zero boundary conditions defined on $\widetilde{\mathcal{T}}_h$.

Finally, if a function $v\colon D \to \mathbb{R}$ or $v\colon D \to \mathbb{R}^2$ is constant over a set $K \subset D$, we denote by $v\big|_K$ its constant value.

2.2 Method

Weak formulation: bilinear form $a\colon V \times V \to \mathbb{R}$ and a linear functional $F\colon V \to \mathbb{R}$ satisfying the usual continuity and coercivity constraints, look for $u \in V$ satisfying

$$a(u, v) = F(v), \tag{2.1}$$

for all functions $v \in V$. Galerkin formulation: for $V_h \subset V$ such that $\dim V_h < \infty$, find $u_h \in V_h$ such that

$$a(u_h, v_h) = f(v_h), \quad \forall v_h \in V_h, \tag{2.2}$$

for all $v_h \in V_h$. Given a triangulation $\mathcal{T}_h$ of the domain D, we choose V_h to be the space of linear finite elements, i.e., $V_h = X_h^1 \cap V$, where

$$X_h^1 = \{v_h \in C^0(\overline{D})\colon v_h|_K \in \mathcal{P}_1, \text{ for all } K \in \mathcal{T}_h\},$$

and where $\mathcal{P}_1$ is the space of polynomials of degree at most one. The finite element space can be written then as $V_h = \operatorname{span}\{\varphi_i\}_{i=1}^N$, where the basis $\{\varphi_i\}_{i=1}^N$ are the Lagrange basis functions. Hence, each $v_h \in V_h$ can be written as $v_h = \sum_{i=1}^N v_i \varphi_i$, where v_i are the coefficients of v_h on the basis $\{\varphi_i\}_{i=1}^N$. Our probabilistic method is based on a randomly perturbed mesh $\widetilde{\mathcal{T}}_h$ which is defined as follows.

Definition 2.1. Let us consider a domain $D \subset \mathbb{R}^d$ and a triangulation $\mathcal{T}_h$ characterised by the maximum diameter $h > 0$ of its elements and by the set of vertices $\mathcal{N}_h = \{x_i\}_{i=1}^N$ such that $\mathcal{N}_h = \mathcal{N}_h^I \cup \mathcal{N}_h^B$, where $\mathcal{N}_h^I$ and $\mathcal{N}_h^B$ are the vertices in the interior of D and on ∂D respectively, and where we denote $N_I = |\mathcal{N}_h^I|$ and $N_B = |\mathcal{N}_h^B|$. Given a probability space (Ω, Σ, μ), the random mesh $\widetilde{\mathcal{T}}_h$ is defined by a sequence of random variables $\{\alpha_i\}_{i=1}^{N_I}$, $\alpha_i\colon \Omega \to \mathbb{R}^d$, which are used to perturb the internal nodes as

$$\tilde{x}_i = x_i + h^p \alpha_i, \quad x_i \in \mathcal{N}_h^I,$$

where $p \geq 1$. Let us remark that the vertices laying on ∂D in $\mathcal{T}_h$ are unperturbed in $\widetilde{\mathcal{T}}_h$.

Once the perturbed mesh $\widetilde{\mathcal{T}}_h$ is obtained, let us denote by $\widetilde{V}_h$ the piecewise linear finite element space defined on $\widetilde{\mathcal{T}}_h$. Let us remark that the space $\widetilde{V}_h = \widetilde{V}_h(\omega)$ is random itself, i.e., for each realisation of the random variables $\{\alpha_i\}_{i=1}^{N_I}$ we obtain a different perturbed finite element space.

Definition 2.2. With the notation above, the probabilistic solution $\widetilde{u}_h\colon \Omega \times D \to \mathbb{R}$ is a random field satisfying for all $\omega \in \Omega$

$$\widetilde{u}_h(\omega, \cdot) \in \widetilde{V}_h(\omega), \text{ s.t. } a(\widetilde{u}_h(\omega, \cdot), \widetilde{v}_h) = F(\widetilde{v}_h), \text{ for all } \widetilde{v}_h \in \widetilde{V}_h(\omega).$$

Let us finally introduce the following assumption on the random variables defining the mesh perturbation.

Assumption 2.3. The random variables α_i satisfy

$$|\alpha_i| \leq \frac{1}{2}\left(\frac{\rho_i}{h}\right)^p,$$

almost surely, where

$$\rho_i = \min_{K \in \Delta(x_i)} \rho_K,$$

with $\Delta(x_i)$ the set of all elements having x_i as a vertex. With this choice, the topology of the mesh $\widetilde{\mathcal{T}}_h$ is almost surely the same as the topology of $\mathcal{T}_h$, i.e., the edges of $\widetilde{\mathcal{T}}_h$ do not cross.

3 A priori error analysis

In this section, we analyse the convergence a priori of our method. In particular, we wish the family of probabilistic solutions to be close to the solution obtained with the original mesh, i.e., we will prove that

$$\|u_h - \widetilde{u}_h\|_{\mathcal{X}} \leq C\eta(h), \quad \text{a.s.}, \tag{3.1}$$

where $\eta\colon \mathbb{R} \to \mathbb{R}$ is such that $\eta(h) \to 0$ for $h \to 0$ and where $\mathcal{X} = \{H^1(D), L^\infty(D)\}$. Similarly to standard error analysis, we first introduce an interpolation result and then prove convergence in the above sense.

3.1 Interpolation analysis

In this section we consider the Legendre piecewise linear interpolants and their properties when they are employed to pass from the space V_h to the space $\widetilde{V}_h$. Let us first recall the definition of the Legendre interpolant.

Definition 3.1. Let $V = \mathcal{C}_0^0(D)$. We denote by $\Pi_h\colon V \to V_h$ and $\widetilde{\Pi}_h\colon V \to \widetilde{V}_h$ the Legendre piecewise linear interpolation operators on V_h and $\widetilde{V}_h$ respectively, i.e., for $v \in V$

$$\Pi_h v(x) = \sum_{x_j \in \mathcal{N}_h^I} v(x_j)\varphi_j(x), \quad \widetilde{\Pi}_h v(x) = \sum_{\tilde{x}_j \in \widetilde{\mathcal{N}}_h^I} v(x_j)\tilde{\varphi}_j(x),$$

where $\{\varphi_i\}_{i=1}^{N^I}$ and $\{\tilde{\varphi}_i\}_{i=1}^{N^I}$ are the basis functions of V_h and $\widetilde{V}_h$ respectively.

In the following lemma we characterise the value that the Legendre interpolant $\widetilde{\Pi}_h$ assumes on the nodes of the original mesh $\mathcal{T}_h$. Let us remark that $V_h \subset C_0^0(D)$, thus the interpolant above can be employed on V_h.

Lemma 3.2. *With the notation of Definition 3.1, it holds for all $v_h \in V_h$ and all $x_i \in \mathcal{N}_h^I$*

$$\begin{aligned} v_h(\tilde{x}_i) &= v_h(x_i) + \bar{h}_i^p \alpha_i^\top \nabla v_h(\tilde{x}_i), \\ \widetilde{\Pi}_h v_h(x_i) - v_h(x_i) &= \bar{h}_i^p \alpha_i^\top \Big(\nabla v_h(\tilde{x}_i) - \nabla\widetilde{\Pi}_h v_h(x_i)\Big). \end{aligned}$$

Proof. We can now expand the function v_h, which is linear on the segment connecting x_i and $\tilde{x}_i$, as

$$v_h(\tilde{x}_i) = v_h(x_i) + \bar{h}_i^p \alpha_i^\top \nabla v_h(\tilde{x}_i), \tag{3.2}$$

which is the first equality. Let us now denote $e_h = \widetilde{\Pi}_h v_h - v_h$. An exact Taylor expansion of the linear basis function $\tilde{\varphi}_j$ gives

$$\begin{aligned} e_h(x_i) &= \sum_j v_h(\tilde{x}_j)\varphi_j(x_i) - v_h(x_i) \\ &= \sum_j v_h(\tilde{x}_j)\Big(\tilde{\varphi}_j(\tilde{x}_i) - \bar{h}_i^p \alpha_i^\top \nabla\tilde{\varphi}_j(x_i)\Big) - v_h(x_i) \\ &= v_h(\tilde{x}_i) - v_h(x_i) - \sum_j \bar{h}_i^p \alpha_i^\top v_h(\tilde{x}_j)\nabla\tilde{\varphi}_j(x_i). \end{aligned}$$

This, together with (3.2), yields

$$\begin{aligned} e_h(x_i) &= \bar{h}_i^p \alpha_i^\top \nabla v_h(\tilde{x}_i) - \bar{h}_i^p \alpha_i^\top \sum_j v_h(\tilde{x}_j)\nabla\tilde{\varphi}_j(x_i) \\ &= \bar{h}_i^p \alpha_i^\top \Big(\nabla v_h(\tilde{x}_i) - \nabla\widetilde{\Pi}_h v_h(x_i)\Big), \end{aligned}$$

which is the second desired equality and which therefore concludes the proof. □

We are not interested in all possible functions in V_h, but only in those which are close enough to a smooth function. The definition below sets the function space we consider in the following.

Definition 3.3. We denote by $V_h^{2,\infty} \subset V_h$ the space such that $v_h \in V_h^{2,\infty}$ if there exists $v \in W^{2,\infty}(D)$ satisfying

$$\|v - v_h\|_{W^{1,\infty}(D)} \leq Ch|\log h||v|_{W^{2,\infty}(D)},$$

where $C > 0$ is a constant independent of h.

In the following Lemma, we provide a property of functions in $V_h^{2,\infty}$ which is quite consequent from the definition of the space. Since we repeatedly employ this result in the following, let us highlight it here.

Lemma 3.4. *Let $v_h \in V_h^{2,\infty}$. Then, if two triangles $K, K' \in \mathcal{T}_h$ share a vertex, it holds*

$$|\nabla v_h\big|_K - \nabla v_h\big|_{K'}| \leq Ch|\log h||v|_{W^{2,\infty}(D)},$$

for a constant $C > 0$ independent of h.

Proof. The proof follows from the triangle inequality. In particular, let $x \in K$ and $x' \in K'$. Then, there exists $v \in W^{2,\infty}$ such that

$$\begin{aligned}
|\nabla v_h\big|_K - \nabla v_h\big|_{K'}| &\leq |\nabla v_h\big|_K - \nabla v(x)| + |\nabla v_h\big|_{K'} - \nabla v(x')| + |\nabla v(x) - \nabla v(x')| \\
&\leq 2|v_h - v|_{W^{1,\infty}(D)} + |v|_{W^{2,\infty}(D)}|x - x'|.
\end{aligned}$$

The desired result follows from the definition of $V_h^{2,\infty}$ and from the fact that since K and K' share a vertex, it is possible to bound $|x - x'| \leq Ch$. $\square$

In the following we consider the affine maps which define locally the geometry of the mesh. Let $K \in \mathcal{T}_h$ be an element of the original mesh, and $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ be the corresponding element in the perturbed mesh. Moreover, let us denote by x_1, x_2, x_3 the vertices of K and by $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ the corresponding vertices of $\widetilde{K}$. We denote by $\widehat{K}$ the reference element, i.e., the triangle with vertices $\hat{x}_1 = (0,0)^\top$, $\hat{x}_2 = (1,0)^\top$, $\hat{x}_3 = (0,1)^\top$. We consider the affine maps $F_K \colon \widehat{K} \to K$ and $\widetilde{F}_K \colon \widehat{K} \to \widetilde{K}$ defined for all $\widehat{x} \in \widehat{K}$ as

$$F_K(\hat{x}) = B_K\hat{x} + b_K, \quad \widetilde{F}_K(\hat{x}) = \widetilde{B}_K\hat{x} + \tilde{b}_K,$$

where $b_K = x_1$, $\tilde{b}_K = \tilde{x}_1$ and the matrices $B_K, \widetilde{B}_K \in \mathbb{R}^{2\times 2}$ are defined as

$$B_K = \begin{pmatrix} x_2 - x_1 \mid x_3 - x_1 \end{pmatrix}, \quad B_{\widetilde{K}} = \begin{pmatrix} \tilde{x}_2 - \tilde{x}_1 \mid \tilde{x}_3 - \tilde{x}_1 \end{pmatrix},$$

so that $F_K(\widehat{x}_i) = x_i$ and $\widetilde{F}_K(\widehat{x}_i) = \tilde{x}_i$ for $i = 1, 2, 3$. The "closeness" of K to $\widetilde{K}$ can be studied in terms of the matrices B_K and $\widetilde{B}_K$, and the relation linking these two matrices is given in the following Lemma.

Lemma 3.5. *Let Assumption 2.3 hold. If $K \in \mathcal{T}_h$ and $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ is the corresponding element in the perturbed mesh, then*

$$B_{\widetilde{K}}^{-\top} - B_K^{-\top} = \Gamma,$$

where the matrix $\Gamma \in \mathbb{R}^{2\times 2}$ satisfies $|\Gamma| \leq C$ a.s. for a constant $C > 0$ independent of h and is given by the infinite sum

$$\Gamma = \sum_{j=1}^{\infty} (-1)^j \left(h^p \Lambda^\top B_K^{-\top}\right)^j, \quad \textit{with } \Lambda = \begin{pmatrix} \alpha_2 - \alpha_1 \mid \alpha_3 - \alpha_1 \end{pmatrix}. \tag{3.3}$$

Proof. From the definition of $\widetilde{K}$ and under Assumption 2.3, it is clear that $\widetilde{B}_K = B_K + h^p\Lambda$. Moreover, we can rewrite $B_{\widetilde{K}}^{-\top}$ with algebraic operations as

$$B_{\widetilde{K}}^{-\top} = (B_K + h^p\Lambda)^{-\top} = B_K^{-\top}(I + h^p B_K^{-1}\Lambda)^{-\top}.$$

In order for the final result to hold, we need that

$$(I + h^p B_K^{-1}\Lambda)^{-\top} = I + \Gamma, \tag{3.4}$$

with Γ given in (3.3). This is true whenever $h^p|B_K^{-1}\Lambda| < 1$. Denoting by $\widehat{h}$ the exterior diameter of the reference triangle $\widehat{K}$ and applying [4, Theorem 3.1.3] we have

$$|B_K^{-1}\Lambda| \leq |B_K^{-1}||\Lambda| \leq \frac{\widehat{h}}{\rho_K}|\Lambda|.$$

Let us now remark that due to Assumption 2.3, it holds

$$|\Lambda|^2 \leq 2|\alpha_1|^2 + |\alpha_2|^2 + |\alpha_3|^2 \leq \left(\frac{\rho_K}{h^p}\right)^2, \quad \text{a.s.,}$$

which implies $h^p|B_K^{-1}\Lambda| \leq \widehat{h}$ a.s. Now $\widehat{h} = 1/\sqrt{2}$, so we have that $h^p|B_K^{-1}\Lambda| < 1$ a.s. and therefore the series (3.3) converges. □

We now proceed to bound the difference between the gradient of v_h and of its interpolant $\widetilde{\Pi}_h v_h$ on a single element.

Lemma 3.6. *With the notation of Definition 2.1, Definition 3.1 and Lemma 3.5, it holds for all $v_h \in V_h^{2,\infty}$*

$$|\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}}| \leq C h^p |\log h| |v|_{W^{2,\infty}(D)},$$

where p is given in Definition 2.1.

Proof. Let us define $\widehat{v}_h \colon \widehat{K} \to \mathbb{R}$ as $\widehat{v}_h \coloneqq v_h \circ F_K$ and $\widetilde{\Pi}_h\widehat{v}_h \colon \widehat{K} \to \mathbb{R}$ as $\widetilde{\Pi}_h\widehat{v}_h \coloneqq \widetilde{\Pi}_h v_h \circ \widetilde{F}_K$. Then, the chain rule yields

$$\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}} = B_K^{-\top}\widehat{\nabla}\widehat{v}_h - B_{\widetilde{K}}^{-\top}\widehat{\nabla}\widetilde{\Pi}_h\widehat{v}_h, \tag{3.5}$$

where $\widehat{\nabla}$ is the gradient with respect to $\hat{x}$. Moreover, the gradient on the reference triangle of the interpolated function satisfies

$$\widehat{\nabla}\widetilde{\Pi}_h\widehat{v}_h = \begin{pmatrix} \widetilde{\Pi}_h\widehat{v}_h(\hat{x}_2) - \widetilde{\Pi}_h\widehat{v}_h(\hat{x}_1) \\ \widetilde{\Pi}_h\widehat{v}_h(\hat{x}_3) - \widetilde{\Pi}_h\widehat{v}_h(\hat{x}_1) \end{pmatrix} = \begin{pmatrix} \widetilde{\Pi}_h v_h(\tilde{x}_2) - \widetilde{\Pi}_h v_h(\tilde{x}_1) \\ \widetilde{\Pi}_h v_h(\tilde{x}_3) - \widetilde{\Pi}_h v_h(\tilde{x}_1) \end{pmatrix}.$$

Since the interpolation is exact on the nodes of the mesh $\widetilde{\mathcal{T}}_h$ and due to Lemma 3.2, this yields

$$\widehat{\nabla}\widetilde{\Pi}_h\widehat{v}_h = \begin{pmatrix} v_h(\tilde{x}_2) - v_h(\tilde{x}_1) \\ v_h(\tilde{x}_3) - v_h(\tilde{x}_1) \end{pmatrix} = \begin{pmatrix} v_h(x_2) - v_h(x_1) \\ v_h(x_3) - v_h(x_1) \end{pmatrix} + h^p\gamma = \widehat{\nabla}\widehat{v}_h + h^p\gamma,$$

where

$$\gamma = \begin{pmatrix} \alpha_2^\top \nabla v_h(\tilde{x}_2) - \alpha_1^\top \nabla v_h(\tilde{x}_1) \\ \alpha_3^\top \nabla v_h(\tilde{x}_3) - \alpha_1^\top \nabla v_h(\tilde{x}_1) \end{pmatrix}.$$

Employing Lemma 3.5, we can replace the expression for $B_{\widetilde{K}}$ in (3.5) and deduce

$$\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}} = -B_K^{-\top}\left(h^p(I+\Gamma)\gamma + \Gamma\widehat{\nabla}\widehat{v}_h\right). \tag{3.6}$$

Let us now remark that from (3.4) Γ satisfies

$$\Gamma = -(I+\Gamma)h^p\Lambda^\top B_K^{-\top},$$

which together with (3.6) yields

$$\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}} = h^p B_K^{-\top}(I+\Gamma)\left(\lambda^\top B_K^{-\top}\widehat{\nabla}\widehat{v}_h - \gamma\right). \tag{3.7}$$

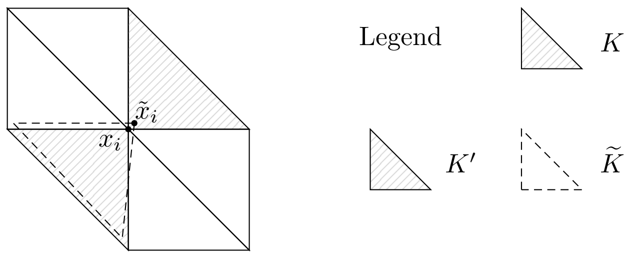

Figure 1: Scheme for the proof of Lemma 3.7. The triangle with solid grey lines on the background is K, the triangle with dashed grey lines on the background is K' and the triangle with dashed borders is $\widetilde{K}$.

We can now compute explicitly

$$
\begin{aligned}
\gamma - \Lambda^\top B_K^{-\top} \widehat{\nabla} \widehat{v}_h &= \gamma - \Lambda^\top \nabla v_h\big|_K \\
&= \begin{pmatrix} \alpha_2^\top \nabla v_h(\tilde{x}_2) - \alpha_1^\top \nabla v_h(\tilde{x}_1) \\ \alpha_3^\top \nabla v_h(\tilde{x}_3) - \alpha_1^\top \nabla v_h(\tilde{x}_1) \end{pmatrix} - \begin{pmatrix} (\alpha_2^\top - \alpha_1^\top) \nabla v_h\big|_K \\ (\alpha_3^\top - \alpha_1^\top) \nabla v_h\big|_K \end{pmatrix} \\
&= \begin{pmatrix} \alpha_2^\top \left(\nabla v_h(\tilde{x}_2) - \nabla v_h\big|_K\right) + \alpha_1^\top \left(\nabla v_h\big|_K - \nabla v_h(\tilde{x}_1)\right) \\ \alpha_3^\top \left(\nabla v_h(\tilde{x}_3) - \nabla v_h\big|_K\right) + \alpha_1^\top \left(\nabla v_h\big|_K - \nabla v_h(\tilde{x}_1)\right) \end{pmatrix}.
\end{aligned}
$$

Due to Lemma 3.4 and Assumption 2.3 we have therefore

$$
|\gamma - \Lambda^\top B_K^{-\top} \widehat{\nabla} \widehat{v}_h| \leq C h |\log h| |v|_{W^{2,\infty}(D)},
$$

almost surely, which, together with (3.7), implies

$$
|\nabla v_h\big|_K - \nabla \widetilde{\Pi}_h v_h\big|_{\widetilde{K}}| \leq C h^{p+1} |\log h| |v|_{W^{2,\infty}(D)} |B_K^{-\top}| |I + \Gamma|
$$

Finally, since $|B_K^{-1}| \leq C h^{-1}$ and due to Lemma 3.5, we obtain

$$
|\nabla v_h\big|_K - \nabla \widetilde{\Pi}_h v_h\big|_{\widetilde{K}}| \leq C h^{p} |\log h| |v|_{W^{2,\infty}(D)},
$$

which is the desired result. □

We can now prove an interpolation result in $L^\infty(D)$

Lemma 3.7. *With the notation of Definition 3.3, let $v_h \in V_h^{2,\infty}$. Then, with the notation of Definition 3.1, it holds*

$$
\|v_h - \widetilde{\Pi}_h v_h\|_{L^\infty(D)} \leq C h^{p+1} |\log h|,
$$

where $C > 0$ is a constant independent of h.

Proof. Let us denote $e_h = \widetilde{\Pi}_h v_h - v_h$ and let us consider $x_i \in \mathcal{N}^I$. By definition $e_h(\tilde{x}_i) = 0$ for all $i = 0, \ldots, N$ and due to Lemma 3.2

$$
e_h(x_i) = h^p \alpha_i \big(\nabla v_h(\tilde{x}_i) - \nabla \widetilde{\Pi}_h v_h(x_i)\big) =: h^p \alpha_i \varepsilon_i. \tag{3.8}
$$

Let us denote by K the element of $\mathcal{T}_h$ such that the corresponding element $\widetilde{K} \in \widetilde{\mathcal{T}}_h$ contains x_i. Furthermore, let us denote by K' the element in the original mesh containing $\tilde{x}_i$. We refer to Fig. 1 for a schematic representation of these elements. With this notation, we have

$$
\nabla v_h(\tilde{x}_i) = \nabla v_h\big|_{K'}, \quad \nabla \widetilde{\Pi}_h v_h(x_i) = \nabla \widetilde{\Pi}_h v_h\big|_{\widetilde{K}},
$$

and we can then decompose ε_i as $\varepsilon_i = \varepsilon_{i,1} + \varepsilon_{i,2}$ with

$$
\varepsilon_{i,1} = \nabla v_h\big|_{K'} - \nabla v_h\big|_K, \quad \varepsilon_{i,2} = \nabla v_h\big|_K - \nabla \widetilde{\Pi}_h v_h\big|_{\widetilde{K}}.
$$

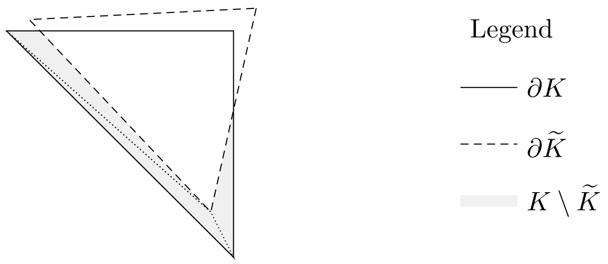

Figure 2: Scheme for the proof of Lemma 3.8. The triangle with solid border is $K \in \mathcal{T}_h$, the one with dashed border is $\widetilde{K} \in \widetilde{\mathcal{T}}_h$. The area filled in grey is $K \setminus \widetilde{K}$, and the dotted lines give one of the possible subdivision in triangles of $K \setminus \widetilde{K}$.

Due to Lemma 3.4, we have

$$|\varepsilon_{i,1}| \le Ch|\log h||v|_{W^{2,\infty}(D)},$$

Moreover, due to Lemma 3.6, we have

$$|\varepsilon_{i,2}| \le Ch^p|\log h||v|_{W^{2,\infty}(D)}$$

Since $p \ge 1$, the triangular inequality yields $\varepsilon_i \le Ch|\log h||v|_{W^{2,\infty}(D)}$ for each node. Replacing this bound in (3.8), we get for each $x_i \in \mathcal{N}_h^I$

$$|e_h(x_i)| \le Ch^{p+1}\alpha_i|\log h||v|_{W^{2,\infty}(D)}.$$

Let us now remark that since by definition $e_h(\tilde{x}_i) = 0$ for all modified nodes, and since e_h is linear on D, the maximum of e_h has to be realised on one of the nodes of the original mesh. Hence

$$\|e_h\|_{L^\infty(D)} = \max_{x_i \in \mathcal{N}_h^I} |e_h(x_i)| \le Ch^{p+1}|\log h||v|_{W^{2,\infty}(D)},$$

which implies the desired result. □

We now consider the interpolation error in $H^1(D)$.

Lemma 3.8. *With the notation of Definition 3.3, let $v_h \in V_h^{2,\infty}$. Then, with the notation of Definition 3.1, it holds*

$$\|v_h - \widetilde{\Pi}_h v_h\|_{H^1(D)} \le Ch^{(p+1)/2}|\log h|,$$

where $C > 0$ is a constant independent of h.

Proof. First, let us recall that for any triangle of sides of length a, b, c and of area A it holds [8]

$$A \le \frac{4\sqrt{3}}{9}\frac{abc}{a+b+c}. \tag{3.9}$$

Let us now consider an element $K \in \mathcal{T}_h$ and the corresponding element $\widetilde{K} \in \widetilde{\mathcal{T}}_h$. It is clear (see e.g. Fig. 2) that it is possible to subdivide $K \setminus \widetilde{K}$ into a bounded number of triangles for which the length one side is bounded by Ch^p and the length of the two other side is bounded by Ch. Therefore, due to (3.9) we have

$$|K \setminus \widetilde{K}| \le C\frac{h^{p+2}}{h+h+h^p} \le Ch^{p+1}. \tag{3.10}$$

Moreover, we remark that

$$|K \cap \widetilde{K}| \le |K| \le Ch^2. \tag{3.11}$$

Let us now denote by N_K the number of triangles in which the set $K \setminus \widetilde{K}$ is divided and by $K_{\text{diff}}^{(i)}$, $i = 1, \dots, N_K$ these triangles. We have

$$\int_K |\nabla e_h|^2 \,\mathrm{d}x = \int_{K\cap\widetilde{K}} |\nabla e_h|^2 \,\mathrm{d}x + \sum_{i=1}^{N_K} \int_{K_{\text{diff}}^{(i)}} |\nabla e_h|^2 \,\mathrm{d}x.$$

Now, Lemma 3.6 and (3.11) yield

$$\int_{K\cap\widetilde{K}} |\nabla e_h|^2 \,\mathrm{d}x = \int_{K\cap\widetilde{K}} |\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}}|^2 \,\mathrm{d}x \le Ch^{2p+2}|\log h|^2 |v|^2_{W^{2,\infty}(D)}. \tag{3.12}$$

Let us now consider the second term. Each triangle $K_{\text{diff}}^{(i)}$ intersects a finite number $N_K^{(i)}$ of triangles in the mesh $\widetilde{\mathcal{T}}_h$. We denote by $\widetilde{K}^{(i,j)}$ for $j = 1, \dots, N_K^{(i)}$ these triangles and by $\widetilde{K}_{\text{diff}}^{(i,j)} = \widetilde{K}^{(i,j)} \cap \widetilde{K}_{\text{diff}}^{(i)}$, for which we remark that it holds

$$|\widetilde{K}_{\text{diff}}^{(i,j)}| \le |K_{\text{diff}}^{(i)}| \le |K \setminus \widetilde{K}| \le Ch^{p+1}.$$

Finally, for each i, j we denote by $K^{(i,j)}$ the element of $\mathcal{T}_h$ corresponding to $\widetilde{K}^{(i,j)} \in \mathcal{T}_h$ and remark that it is a neighbour of K. Therefore

$$\int_{K_{\text{diff}}^{(i)}} |e_h|^2 \,\mathrm{d}x = \sum_{j=1}^{N_K^{(i)}} \int_{\widetilde{K}_{\text{diff}}^{(i,j)}} |e_h|^2 \,\mathrm{d}x,$$

where due to Young's inequality we have

$$\begin{aligned}\int_{\widetilde{K}_{\text{diff}}^{(i,j)}} |\nabla e_h|^2 \,\mathrm{d}x &= \int_{\widetilde{K}_{\text{diff}}^{(i,j)}} |\nabla v_h\big|_K - \nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}^{(i,j)}}|^2 \,\mathrm{d}x \\ &\le 2\left(|\nabla v_h\big|_K - \nabla v_h\big|_{K^{(i,j)}}|^2 + |\nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}^{(i,j)}} - \nabla v_h\big|_{K^{(i,j)}}|^2\right) |\widetilde{K}_{\text{diff}}^{(i,j)}|.\end{aligned} \tag{3.13}$$

Due to Lemma 3.4, we have first

$$|\nabla v_h\big|_K - \nabla v_h\big|_{K^{(i,j)}}|^2 \le Ch^2|\log h|^2|v|^2_{W^{2,\infty}(D)},$$

and due to Lemma 3.6, we obtain

$$|\nabla\widetilde{\Pi}_h v_h\big|_{\widetilde{K}^{(i,j)}} - \nabla v_h\big|_{K^{(i,j)}}|^2 \le Ch^{2p}|\log h|^2|v|^2_{W^{2,\infty}(D)}.$$

Therefore, replacing these two inequalities and (3.10) in (3.13) and since $p \ge 1$ and $h < 1$, we obtain for a constant $C > 0$

$$\int_{\widetilde{K}_{\text{diff}}^{(i,j)}} |\nabla e_h|^2 \,\mathrm{d}x \le Ch^{p+3}|\log h|^2|v|^2_{W^{2,\infty}(D)}.$$

Hence, we get

$$\int_{K\setminus\widetilde{K}} |\nabla e_h|^2 \,\mathrm{d}x \le CN_K \left(\sum_{i=1}^{N_K} N_K^{(i)}\right) h^{p+3} \left(|\log h| + 1\right)^2 |v|^2_{W^{2,\infty}(D)}. \tag{3.14}$$

Combining (3.12) and (3.14) and since $2p + 2 \ge p + 3$ for $p \ge 1$, we conclude that there exists a constant $C > 0$ independent of h but dependent on v such that

$$\int_K |\nabla e_h|^2 \,\mathrm{d}x \le Ch^{p+3}|\log h|^2.$$

Finally, due to assumption on the mesh, we have that $N \le Ch^2$ and therefore

$$|e_h|^2_{H^1(D)} = \sum_{K\in\mathcal{T}_h} \int_K |\nabla e_h|^2 \,\mathrm{d}x \le Ch^{p+1}|\log h|^2,$$

which implies the desired result. $\square$

3.2 The sum space

In order to prove convergence of the probabilistic solution, and moreover the closeness of u_h to $\widetilde{u}_h$ in the sense of (3.1), we first need to define a convenient function space which is finite dimensional but which contains both V_h and $\widetilde{V}_h$.

Definition 3.9. Let us denote by $V_h^+ \subset V$ the space of functions that can be written as the sum of a function in V_h and a function in $\widetilde{V}_h$, i.e., for any function $v_h^+ \in V_h^+$ there exists functions $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$ such that $v_h^+ = v_h + \widetilde{v}_h$.

Remark 3.10. Let us remark that in our setting of homogeneous boundary conditions $V_h \cap \widetilde{V}_h = \{0\}$ almost surely. Therefore, the space V_h^+ is given by the direct sum $V_h^+ = V_h \oplus \widetilde{V}_h$ and the decomposition of $v_h^+ \in V_h^+$ is unique. Moreover, since $\dim(V_h) = \dim(\widetilde{V}_h) = N_I$, we have $\dim(V_h^+) = 2N_I$. Moreover, let us remark that we are not building a so-called supermesh as in [7, 11, 12]

The following result characterizes the distance of the finite elements solutions on the spaces V_h and $\widetilde{V}_h$.

Lemma 3.11. *Let u_h and $\widetilde{u}_h$ be the solutions of*

$$a(u_h, v_h) = F(v_h), \quad a(\widetilde{u}_h, \widetilde{v}_h) = F(\widetilde{v}_h),$$

for all $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$. Then, it holds for all $v_h, w_h \in V_h$ and for all $\widetilde{v}_h, \widetilde{w}_h \in \widetilde{V}_h$

$$\|u_h - \widetilde{u}_h\|_V^2 \leq C \left(\|u_h^+ - w_h\|_V \|\widetilde{u}_h - v_h\|_V + \|u_h^+ - \widetilde{w}_h\|_V \|u_h - \widetilde{v}_h\|_V \right),$$

where $C > 0$ is a constant independent of h and where $u_h^+ \in V_h^+$ is the solution of

$$a(u_h^+, v_h^+) = F(v_h^+), \tag{3.15}$$

for all $v_h^+ \in V_h^+$.

Proof. Since V_h and $\widetilde{V}_h$ are both subspaces of V_h^+, we have due to Galerkin's orthogonality

$$\begin{aligned} a(u_h^+ - u_h, v_h) &= 0, \quad \forall v_h \in V_h, \\ a(u_h^+ - \widetilde{u}_h, \widetilde{v}_h) &= 0, \quad \forall \widetilde{v}_h \in \widetilde{V}_h, \end{aligned} \tag{3.16}$$

which means that u_h and $\widetilde{u}_h$ are the elliptic projection of u_h^+ onto V_h and $\widetilde{V}_h$ respectively. Hence, due to Cea's lemma

$$\|u_h^+ - u_h\|_V \leq C \|u_h^+ - w_h\|_V, \quad \|u_h^+ - \widetilde{u}_h\|_V \leq C \|u_h^+ - \widetilde{w}_h\|_V, \tag{3.17}$$

for all $w_h \in V_h$ and $\widetilde{w}_h \in \widetilde{V}_h$, where $C = M/\alpha$. Using the coercivity on V of $a(\cdot, \cdot)$, adding and subtracting $a(u_h^+, u_h - \widetilde{u}_h)$ and due to (3.16) we have for all $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$

$$\begin{aligned} \alpha \|u_h - \widetilde{u}_h\|_V^2 &\leq a(u_h - \widetilde{u}_h, u_h - \widetilde{u}_h) \\ &= a(u_h - u_h^+, u_h - \widetilde{u}_h) + a(u_h^+ - \widetilde{u}_h, u_h - \widetilde{u}_h) \\ &= a(u_h - u_h^+, v_h - \widetilde{u}_h) + a(u_h^+ - \widetilde{u}_h, u_h - \widetilde{v}_h). \end{aligned}$$

Due to the continuity of the bilinear form we then have for all $w_h \in V_h$ and $\widetilde{w}_h \in \widetilde{V}_h$

$$\begin{aligned} \alpha \|u_h - \widetilde{u}_h\|_V^2 &\leq M \Big(\|u_h^+ - u_h\|_V \|\widetilde{u}_h - v_h\|_V + \|u_h^+ - \widetilde{u}_h\|_V \|u_h - \widetilde{v}_h\|_V \Big) \\ &\leq \frac{M^2}{\alpha} \Big(\|u_h^+ - w_h\|_V \|\widetilde{u}_h - v_h\|_V + \|u_h^+ - \widetilde{w}_h\|_V \|u_h - \widetilde{v}_h\|_V \Big), \end{aligned}$$

which is the desired result. □

Let us remark that the Lemma above holds true for any choice of V_h and $\widetilde{V}_h$, not necessarily disjoint, and for any space V_h^+ such that $V_h \cup \widetilde{V}_h \subseteq V_h^+$. For the next result, we instead consider the setting in which V_h and $\widetilde{V}_h$ are the fixed and randomly perturbed finite element spaces of Definition 2.1.

Lemma 3.12. *Let $u_h^+ \in V_h^+$ be the solution of* (3.15)*, and let us denote by z_h and $\widetilde{z}_h$ its unique components in V_h and $\widetilde{V}_h$, respectively, i.e., $u_h^+ = z_h + \widetilde{z}_h$. Then, it holds*

$$\|z_h - \widetilde{z}_h\|_V \leq Ch^r.$$

Proof. □

Corollary 3.13. *With the notation of Lemma 3.12 and of Definition 3.3, we have $z_h \in V_h^{2,\infty}$ and $\widetilde{z}_h \in \widetilde{V}_h^{2,\infty}$.*

Proof. Let us consider without loss of generality z_h. Due to (3.17), we have

$$\|u - u_h^+\|_V \leq \|u - u_h\|_V,$$

□

We now introduce a result of interpolation with the Legendre interpolants defined in Definition 3.1.

Lemma 3.14. *Let Π_h and $\widetilde{\Pi}_h$ be defined in Definition 3.1. Then, for all $v_h^+ \in V_h^+$ it holds*

$$\Pi_h v_h^+ - v_h^+ = \Pi_h \widetilde{v}_h - \widetilde{v}_h, \quad \widetilde{\Pi}_h v_h^+ - v_h^+ = \Pi_h v_h - v_h,$$

where $v_h \in V_h$, $\widetilde{v}_h \in \widetilde{V}_h$ and $v_h^+ = v_h + \widetilde{v}_h$.

Proof. The result is implied by the linearity of Π_h and $\widetilde{\Pi}_h$ and since the restriction of Π_h on V_h is the identity function (respectively, $\widetilde{\Pi}_h$ on $\widetilde{V}_h$). □

3.3 Convergence result

We now present here a classic convergence result for the finite elements method [4, Theorem 3.3.7], which allows to control the supremum of the error under smoothness assumptions on the solution.

Theorem 3.15. *Let u be the solution of* (2.1) *and $u_h \in V_h$ be the solution of* (2.2)*. Then, if $u \in W^{2,\infty}(D)$, it holds*

$$\begin{aligned}
\|u - u_h\|_{L^\infty(D)} &\leq Ch^2 |\log h|^{3/2} |u|_{W^{2,\infty}(D)}, \\
|u - u_h|_{W^{1,\infty}(D)} &\leq Ch |\log h| |u|_{W^{2,\infty}(D)},
\end{aligned}$$

where $C > 0$ is a constant independent of h. In particular, with the notation of Definition 3.3, we have that $u_h \in V_h^{2,\infty}$.

We can now prove the main result of a priori convergence for the probabilistic solution.

Theorem 3.16. *Let u be the solution of* (2.1) *and let u_h and $\widetilde{u}_h$ be the solutions of*

$$a(u_h, v_h) = F(v_h), \quad a(\widetilde{u}_h, \widetilde{v}_h) = F(\widetilde{v}_h),$$

for all $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$. Then, if $u \in W^{2,\infty}(D)$, it holds for $V = H_0^1(D)$

$$\|u_h - \widetilde{u}_h\|_V \leq \quad a.s.,$$

and moreover, it holds

$$\|\widetilde{u}_h - u\| \leq \quad a.s.$$

Proof. Considering Lemma 3.11 with $v_h = \Pi_h \widetilde{u}_h$, $w_h = \Pi_h u_h^+$, $\widetilde{v}_h = \widetilde{\Pi}_h u_h$ and $\widetilde{w}_h = \widetilde{\Pi}_h u_h^+$, we get

$$\|u_h - \widetilde{u}_h\|_V^2 \leq C \left(\|u_h^+ - w_h\|_V \|\widetilde{u}_h - v_h\|_V + \|u_h^+ - \widetilde{w}_h\|_V \|u_h - \widetilde{v}_h\|_V \right),$$

□

4 A posteriori error analysis

Several techniques exist for obtaining a posteriori error estimators in the framework of the FEM (see [20] for an overview), with the twofold goal of controlling the quality of numerical solutions and hence improve the meshing procedure to maximise efficiency. The main purpose of probabilistic numerical methods is to quantify the uncertainty introduced by approximate computations [13]. For the reasons above, we believe that deriving an error estimator from a family of numerical solutions fits perfectly in the probabilistic framework. In this section we present such a procedure for a probabilistic error estimation.

Assumption 4.1. Let $u_h^+ \in V_h^+$ be defined in (3.15). Then we assume there exists $0 \leq \beta < 1$ such that

$$\|u - u_h^+\|_a \leq \beta \|u - u_h\|_a,$$

where $\|u\|_a^2 = a(u,u)$. Moreover, there exists a constant $\gamma > 0$ such that

$$\|u_h - u_h^+\|_a \leq \gamma \|u_h - \widetilde{u}_h\|_a, \tag{4.1}$$

almost surely, where $\widetilde{u}_h$ is the probabilistic solution.

Let us remark that since $V_h \subset V_h^+$, we have $\beta \leq 1$ for the best approximation property of the Galerkin method and that Assumption 4.1 is often denoted in literature as the saturation assumption.

Lemma 4.2. *Let us denote by $z_h \in V_h$ the function $z_h = w_h - u_h/2$. Then*

$$\|z_h - \widetilde{\Pi}_h z_h\|_a \leq \dots$$

Proof.

$$\|z_h\|_a \leq \frac{1}{2}\|w_h - \widetilde{w}_h\|_a.$$

□

Lemma 4.3. *Under …, there exists $\gamma > 0$ independent of h and p such that*

$$\|u_h - u_h^+\|_a \leq \gamma \|u_h - \widetilde{u}_h\|_a,$$

almost surely in Ω.

Proof. Let us write $u_h^+ = w_h + \widetilde{w}_h$, where w_h and $\widetilde{w}_h$ are the two components of u_h^+ in V_h and $\widetilde{V}_h$ respectively. For any $v_h^+ \in V_h^+$, $v_h^+ = v_h + \widetilde{v}_h$, with $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$, by Galerkin orthogonality

$$\begin{aligned} a(u_h^+ - u_h, v_h^+) &= a(u_h^+ - u_h, \widetilde{v}_h) - a(u_h^+ - \widetilde{u}_h, \widetilde{v}_h) \\ &= a(\widetilde{u}_h - u_h, \widetilde{v}_h). \end{aligned}$$

Choosing $v_h^+ = u_h^+ - u_h$, we have $\widetilde{v}_h = \widetilde{w}_h$ and

$$\|u_h^+ - u_h\|_a^2 = a(\widetilde{u}_h - u_h, \widetilde{w}_h).$$

The same procedure applied to $u_h^+ - \widetilde{u}_h$ yields

$$\|u_h^+ - \widetilde{u}_h\|_a^2 = a(u_h - \widetilde{u}_h, w_h).$$

Hence

$$\|u_h^+ - u_h\|_a^2 + \|u_h^+ - \widetilde{u}_h\|_a^2 = a(u_h - \widetilde{u}_h, w_h - \widetilde{w}_h).$$

Let us introduce the functions $z_h = w_h - u_h/2 \in V_h$ and $\tilde{z}_h = \widetilde{w}_h - \widetilde{u}_h/2 \in \widetilde{V}_h$. Then

$$\begin{aligned} \|u_h^+ - u_h\|_a^2 + \|u_h^+ - \widetilde{u}_h\|_a^2 &= \frac{1}{2} a(u_h - \widetilde{u}_h, u_h - \widetilde{u}_h) + a(u_h - \widetilde{u}_h, w_h - \frac{u_h}{2} - (\widetilde{w}_h - \frac{\widetilde{u}_h}{2})) \\ &= \frac{1}{2}\|u_h - \widetilde{u}_h\|_a^2 + a(u_h - \widetilde{u}_h, z_h - \tilde{z}_h). \end{aligned}$$

Consider now the second term in the sum. Adding and subtracting $a(u_h^+, z_h - \tilde{z}_h)$ and considering Galerkin orthogonality we obtain

$$a(u_h - \widetilde{u}_h, z_h - \tilde{z}_h) = a(u_h - u_h^+, v_h - \tilde{z}_h) + a(u_h^+ - \widetilde{u}_h, z_h - \widetilde{v}_h),$$

for all $v_h \in V_h$ and $\widetilde{v}_h \in \widetilde{V}_h$. Hence, applying Cauchy–Schwarz and Young's inequalities we obtain

$$\|u_h^+ - u_h\|_a^2 + \|u_h^+ - \widetilde{u}_h\|_a^2 \leq \|u_h - \widetilde{u}_h\|_a^2 + \inf_{v_h \in V_h} \|\tilde{z}_h - v_h\|_a^2 + \inf_{\widetilde{v}_h \in \widetilde{V}_h} \|z_h - \widetilde{v}_h\|_a^2.$$

□

Moreover, since the perturbed mesh and the original mesh could switch their roles by changing the sign to the random perturbations, the same assumption as (4.1) should be imposed for the probabilistic solution, i.e.

$$\|\widetilde{u}_h - u_h^+\|_a \leq \widetilde{\gamma} \|u_h - \widetilde{u}_h\|_a.$$

Applying the triangular inequality, we get

$$\begin{aligned}(\gamma + \widetilde{\gamma})\|u_h - \widetilde{u}_h\|_a &\geq \|\widetilde{u}_h - u_h^+\|_a + \|u_h - u_h^+\|_a \\ &\geq \|u_h - \widetilde{u}_h\|_a,\end{aligned}$$

which implies that $(\gamma + \widetilde{\gamma}) \geq 1$. The duality in the roles of deterministic and probabilistic meshes implies that γ and $\widetilde{\gamma}$ should be in general approximately equal, at least asymptotically. Hence, the lower bound above guarantees that neither γ nor $\widetilde{\gamma}$ should tend to zero with $h \to 0$.

It is known [2] that under Assumption 4.1 the estimate

$$\|u_h - u_h^+\|_a \leq \|u - u_h\|_a \leq \frac{1}{1-\beta}\|u_h - u_h^+\|_a,$$

holds almost surely. The quantity $\|u_h - u_h^+\|_a$ thus serves as an a posteriori error estimator for the error. However, computations involving the sum space V_h^+ are often intractable if the dimension $d > 1$. Hence, we further expand the upper bound thanks to (4.1) as

$$\|u - u_h\|_a \leq \frac{\gamma}{1-\beta}\|u_h - \widetilde{u}_h\|_a, \tag{4.2}$$

which means that the difference between the deterministic and the probabilistic solutions can be employed as an a posteriori upper bound for the error.

Remark 4.4. Let us remark that the value of β is influenced by the choice of p in Assumption 2.3. Let us consider the limit case of $p \to \infty$. In this case, the spaces V_h and $\tilde{V}_h$ coincide, and in turn coincide both with V_h^+. Hence, the space V_h^+ is in the limit not wider than V_h and one expects $\beta \to 1$. We hence postulate that $\beta = \beta(h,p)$ takes the form

$$\beta(h,p) = 1 - \beta_1 h^{\beta_2(p-1)},$$

for some $0 < \beta_1 \leq 1$ and $\beta_2 > 0$. This is motivated by the fact that the two terms in (4.2) converge with the same rate $\mathcal{O}(h)$ in case $p = 1$ due to a priori error results. Hence, in this case, $\beta(h,1)$ is independent of h and equals a constant value β. Conversely, if $p > 1$, one gets on the right hand side a term of order $\mathcal{O}(h^{\beta_2(1-p)}h^{(p+1)/2})$, bounding a term of order $\mathcal{O}(h)$ on the left hand side. Hence, we impose

$$\beta_2(1-p) + \frac{p+1}{2} \leq 1,$$

which, since $p > 1$, gives $\beta_2 \geq 1/2$. Numerical experiments confirm the qualitative behaviour of the function $\beta(h,p)$ explained above, and lead to the good working practice of fixing $p = 1$.

A more robust estimator could be obtained by averaging a family of M probabilistic solutions $\widetilde{u}_h^{(i)}$, $i = 1, \dots, M$, obtained by M i.i.d. random perturbations of the original mesh. In particular, we have

$$\|u - u_h\|_a^2 \leq C\,\mathbb{E}\|u_h - \widetilde{u}_h\|_a^2 =: C\eta_h^2,$$

where we approximate the estimator η_h via Monte Carlo sampling as

$$\eta_h \approx \sqrt{\frac{1}{M}\sum_{i=1}^{M} \|u_h - \widetilde{u}_h^{(i)}\|_a^2}.$$

Taking the expectation over several realisations should in practice provide a sharper error estimator, as in case $p = 1$ a good portion of the domain D is explored by the vertices of several realisations of the random mesh. Let us consider for simplicity the case $\kappa \equiv 1$, so that $\|u\|_a = \|\nabla u\|_{L^2(D)}$ for all $u \in H_0^1(D)$. In this case, we have

$$\begin{aligned}\eta_h &= \int_K \mathbb{E}|\nabla(u_h - \widetilde{u}_h)|^2 \,\mathrm{d}x \\ &\approx \int_K \mathbb{E}|\mathbb{E}(\nabla u_h) - \nabla \widetilde{u}_h|^2 \,\mathrm{d}x \\ &= \int_K \operatorname{tr}(\operatorname{Var} \nabla u_h) \,\mathrm{d}x.\end{aligned}$$

Hence, following the probabilistic numerics canon, it is possible to interpret the error estimator as an integral measure of the statistical dispersion of numerical solutions over the domain.

We now consider the task of adapting the mesh. Given the error estimator derived above and a prescribed tolerance, we apply a standard technique for generating a sequence of meshes, which we briefly summarise in the following. Let us first split the estimator over the elements of the original mesh as

$$\begin{aligned}\eta_h^2 &= \sum_{K \in \mathcal{T}_h} \mathbb{E} \int_K \kappa \nabla(u_h - \widetilde{u}_h) \cdot \nabla(u_h - \widetilde{u}_h) \,\mathrm{d}x \\ &= \sum_{K \in \mathcal{T}_h} \eta_K^2,\end{aligned}$$

where we consider η_K to be an indicator of the error at a local level. If we impose a tolerance level ϵ for the error, i.e.,

$$\|u - u_h\|_a \leq \epsilon \|u_h\|_a,$$

we obtain that a sufficient condition is given by

$$\eta_K \leq \frac{\epsilon \|u_h\|_a}{\sqrt{N}},$$

where N is the number of elements in $\mathcal{T}_h$. Hence, we proceed iteratively by refining the mesh around elements which do not fulfil the error requirement until the required tolerance is attained. Coarsening of elements where the error indicator is small could be as well employed for saving computational power. The algorithm for mesh adaptation is given in Algorithm 1, where safety factors fac_1 and fac_2 are introduced.

5 Inverse problems

Probabilistic numerical methods are particularly helpful when inserted in the framework of Bayesian inverse problems (BIPs) involving differential equations, as studied in [1,6] for ODEs, and in [3,5] for PDEs. Furthermore, in [16] a theoretical basis is laid for ensuring the well-posedness of probabilistic solutions to BIPs.

We consider the framework introduced in [18] and expanded in [9]. With the notation of (1.1), we consider the PDE

$$\begin{aligned}-\nabla \cdot (e^{\vartheta} \nabla u) &= f, \quad \text{in } D, \\ u &= 0, \quad \text{on } \partial D,\end{aligned} \tag{5.1}$$

Algorithm 1: Probabilistic mesh adaptivity.

Data: $\mathcal{T}_h^{(0)}$, tolerance ε, safety factors fac_1, fac_2, $M \in \mathbb{N}$.

Set $i = 0$;
while $\eta_h > \epsilon \|u_h\|_a$ **do**
  Compute u_h and $\|u_h\|_a$;
  Draw M random meshes and compute $\widetilde{u}_h^{(j)}$ for $j = 1, \dots, M$;
  for $K \in \mathcal{T}_h^{(i)}$ **do**
    Compute η_K ;
    if $\eta_K > \text{fac}_1\, \epsilon \|u_h\|_a / \sqrt{N}$ **then**
      Mark element K for refinement ;
    else if $\eta_K < \text{fac}_2\, \epsilon \|u_h\|_a / \sqrt{N}$ **then**
      Mark element K for coarsening ;
  Build $\mathcal{T}^{(i+1)}$;
  Set $i \leftarrow i + 1$;

where the conductivity field κ is transformed through an exponential function $\kappa = \exp(\vartheta)$ in order to ensure positivity and hence well-posedness of the solution. Moreover, we suppose that $u \in W^{2,\infty}(D)$ and we let $\mathcal{U} = \text{addspace}$ be the space of admissible log-conductivity fields ϑ. The BIP consists in retrieving the true value $\vartheta^\dagger$ of the field ϑ given prior information and corrupted observations $z \in \mathbb{R}^m$ given by

$$z = \mathcal{G}(\vartheta^\dagger) + \varepsilon,$$

where we assume that $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$ is a Gaussian source of additive noise and $\mathcal{G} \colon \mathcal{U} \to \mathbb{R}^m$ is the forward operator. In particular, we can write $\mathcal{G} = \mathcal{O} \circ \mathcal{S}$, where $\mathcal{S} \colon \mathcal{U} \to W^{2,\infty}(D)$ is the solution operator, mapping any value of the field ϑ to the solution u of (5.1), and $\mathcal{O} \colon W^{2,\infty}(D) \to \mathbb{R}^m$ is the observation operator. In this work, we simply consider $\mathcal{O}$ to be defined by point-wise evaluations of the solution, i.e.,

$$\mathcal{O} \colon \vartheta \mapsto \begin{pmatrix} u(x_1) & u(x_2) & \dots & u(x_m) \end{pmatrix}^\top .$$

If the prior information is encoded by a prior measure μ_0 over the space $\mathcal{U}$, then the solution of the BIP is given by the posterior distribution μ such that its Radon–Nikodym derivative satisfies

$$\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(\vartheta; z) = \frac{1}{Z} \exp(-\Phi(\vartheta; z)),$$

where $\Phi \colon (L^\infty)^d \times \mathbb{R}^m \to \mathbb{R}$ is referred to as the potential function and Z is a normalisation constant. Under the Gaussian assumption for the noise, we have

$$\Phi(\vartheta; z) = \frac{1}{2} \|z - \mathcal{G}(\vartheta)\|_{\Sigma_\varepsilon}^2,$$

where the norm $\|\cdot\|_{\Sigma_\varepsilon}$ is defined as

$$\|y\|_{\Sigma_\varepsilon} = \|\Sigma_\varepsilon^{-1/2} y\|_{\mathbb{R}^m}.$$

In the following, we will consider the observations z to be fixed and hence denote $\mu(\mathrm{d}\vartheta) = \mu(\mathrm{d}\vartheta; z)$ as well as $\Phi(\vartheta) = \Phi(\vartheta; z)$. Let us denote by $\mathcal{G}_h \colon \mathcal{U} \to \mathbb{R}^m$ the forward model obtained as $\mathcal{G}_h = \mathcal{O} \circ \mathcal{S}_h$, where $S_h \colon \mathcal{U} \to V_h$ is the solution operator given by the linear FEM and we still denote by $\mathcal{O}$ the restriction of $\mathcal{O}$ to V_h. Denoting by Φ_h the approximate potential, given by

$$\Phi_h(\vartheta) = \frac{1}{2} \|z - \mathcal{G}_h(\vartheta)\|_{\Sigma_\varepsilon}^2, \tag{5.2}$$

we obtain the approximate posterior measure μ_h as

$$\frac{\mathrm{d}\mu_h}{\mathrm{d}\mu_0}(\vartheta) = \frac{1}{Z_h} \exp(-\Phi_h(\vartheta)),$$

where Z_h is the normalisation constant. Stuart proved [18, Theorem 4.6] that under suitable assumptions $d_{\mathrm{H}}(\mu_h, \mu) \to 0$ for $h \to 0$, where $d_{\mathrm{H}}(\cdot, \cdot)$ is the Hellinger distance for probability measures. Hence, assuming an infinite computational budget is available it is possible to compute the posterior measure via approximate computations. This result has then been extended to more general priors than Gaussian [10, 19].

It has been shown empirically that under a fixed computational budget, employing a standard numerical method for the approximation of the solution operator $\mathcal{S}$ can lead to inaccurate results [1, 5, 6]. In particular, in case the variance Σ_ε of the observational noise is small with respect to the discretisation error, the posterior measure μ_h will be overconfident and peaked away from the true value of the unknown field. Probabilistic numerical methods can efficiently tackle this overconfidence issue thanks to the uncertainty quantification of numerical errors they naturally introduce. Given the probability space Ω on which the random variables defining the probabilistic scheme $\alpha_i \colon \Omega \to \mathbb{R}^d$ introduced in Assumption 2.3 are defined, let us denote by $\widetilde{\mathcal{G}}_h \colon \Omega \times \mathcal{U} \to \mathbb{R}^m$ the random forward model obtained as $\widetilde{\mathcal{G}}_h = \mathcal{O} \circ \widetilde{\mathcal{S}}_h$, where $\widetilde{\mathcal{S}}_h \colon \Omega \times \mathcal{U} \to \widetilde{V}_h$ is the solution operator corresponding to the random FEM introduced in this work. Replacing $\mathcal{G}_h$ with $\widetilde{\mathcal{G}}_h$ in (5.2) we get a random potential $\widetilde{\Phi}_h$ and eventually a random posterior measure $\tilde{\mu}_h$ defined by

$$\frac{\mathrm{d}\tilde{\mu}_h}{\mathrm{d}\mu_0}(\vartheta) = \frac{1}{\widetilde{Z}_h} \exp(-\widetilde{\Phi}_h(\vartheta)),$$

where $\widetilde{Z}_h$ is the normalisation constant. In order to obtain an approximation of μ through $\tilde{\mu}_h$, we need to take the expectation of the random probabilistic solution, which is viable in two different manners as explained in [16]. The first approach is to define the measure $\tilde{\mu}_h^{\mathrm{fix}} = \mathbb{E}\, \tilde{\mu}_h$. Otherwise, one could define a measure $\tilde{\mu}_h^{\mathrm{var}}$ through

$$\frac{\mathrm{d}\tilde{\mu}_h^{\mathrm{var}}}{\mathrm{d}\mu_0}(\vartheta) = \frac{1}{\mathbb{E}\, \widetilde{Z}_h} \mathbb{E} \exp(-\widetilde{\Phi}_h(\vartheta)),$$

which is already a deterministic measure. The choice of the names of these two approximation comes from their computation, which is in spirit slightly different. In the case of $\tilde{\mu}_h^{\mathrm{fix}}$, for each event ω one evaluates the forward model and computes the value of the posterior. The expectation is then taken with the respect to the posterior itself, in practice via averaging techniques. Hence, for each ω we fix a perturbed mesh $\widetilde{\mathcal{T}}_h(\omega)$ and compute the posterior for several values of ϑ. Conversely, in the case of the measure $\tilde{\mu}_h^{\mathrm{var}}$ the field ϑ is first fixed, and then the posterior is in practice obtained evaluating the forward model on several (variable) realisations of the random probabilistic solution.

We now need to prove the convergence of the posterior distributions $\tilde{\mu}_h$ and $\tilde{\mu}_h^{\mathrm{var}}$ towards the true posterior μ with respect to the mesh size, which is granted by the following result under three regularity assumptions.

Theorem 5.1 (Theorem 3.9 of [16]). *With the notation above, if*

(i) there exists $q > 0$ such that $\exp(\Phi) \in L^q_{\mu_0}(\mathcal{U})$,

(ii) there exists a constant $C > 0$ such that

$$\mathbb{E}_{\mu_0}[\widetilde{\Phi}_N] \leq C, \quad \textit{almost surely in } \Omega,$$

(iii) it holds

$$\lim_{h \to 0} \left\| \left(\mathbb{E} \| \widetilde{\mathcal{G}}_h - \mathcal{G} \|^2 \right)^{1/2} \right\|_{L^s_{\mu_0}(\mathcal{U})} = 0,$$

where $s = 2q/(q-1)$ and q is given in (i),

then

$$\mathbb{E} \left[d_{\mathrm{H}}(\mu, \tilde{\mu}_h)^2 \right]^{1/2} \leq C \left\| \left(\mathbb{E} \| \widetilde{\mathcal{G}}_h - \mathcal{G} \|^4_{\mathbb{R}^m} \right)^{1/2} \right\|^{1/2}_{L^2_{\mu_0}(\mathcal{U})},$$
$$d_{\mathrm{H}}(\mu, \tilde{\mu}_h^{\mathrm{var}}) \leq C \left\| \mathbb{E} \| \widetilde{\mathcal{G}}_h - \mathcal{G} \|^2_{\mathbb{R}^m} \right\|^{1/2}_{L^s_{\mu_0}(\mathcal{U})}.$$

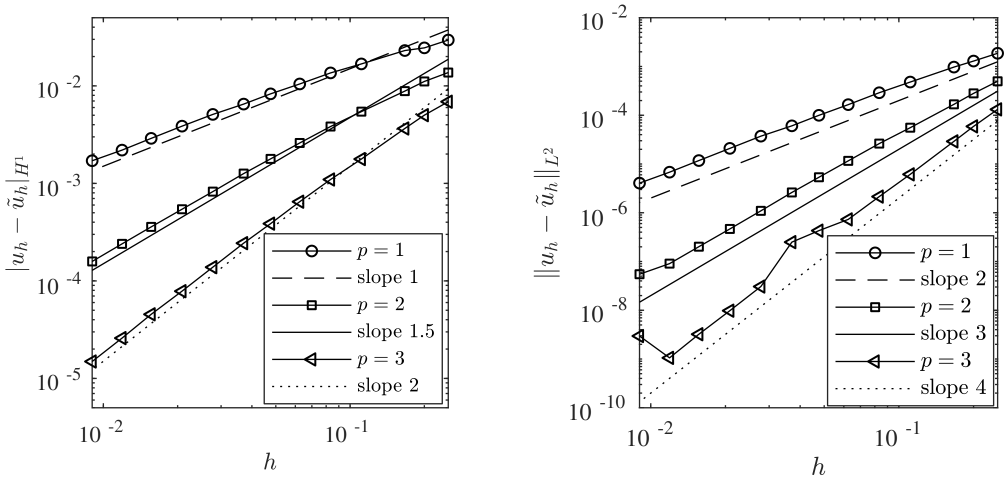

Figure 3: Convergence rates in the H^1 semi-norm and the L^2 norm for the one-dimensional Poisson equation.

Let us remark that for a measure μ the spaces $L^q_\mu(\mathcal{U})$ are defined as

$$L^q_\mu(\mathcal{U}) = \left\{ f \colon \mathcal{U} \to \mathbb{R} : \int_{\mathcal{U}} f(\vartheta)^q \, \mu(\,\mathrm{d}\vartheta) < \infty \right\},$$

with norm

$$\|f\|_{L^q_\mu(\mathcal{U})} = \left(\int_{\mathcal{U}} f(\vartheta)^q \, \mu(\,\mathrm{d}\vartheta) \right)^{1/q}.$$

Theorem 5.1 gives in a general framework the convergence of posterior measures defined through approximate random forward models The following result now guarantees the convergence of the posterior distributions

6 Numerical experiments

6.1 Convergence

One-dimensional case

We consider (1.1) with $\kappa \equiv 1$ on $D = (0, 1)$ and $f(x) = (x - 1/2)\chi_{(1/2,1)}(x)$, so that the solution u satisfies **??**. We verify the result of **??** by choosing $p \in \{1, 2, 3\}$ and by varying the mesh size h in the range $[9 \cdot 10^{-3}, 0.25]$. Moreover, we compute only one realisation of the random mesh for each couple $\{p, h\}$ as our bound holds almost surely. Results, shown in Fig. 3, confirm the validity of the convergence estimates.

6.2 Error estimators

One-dimensional example

Consider

$$\begin{aligned} -u'' &= f, \quad \text{in } (0, 1), \\ u(0) &= u(1) = 0, \end{aligned}$$

with f chosen such that $u(x) = -\sin(12\pi x)\exp(-100(x - 1/2)^2)$ is the true solution. We consider the error estimations of presented in Section 4, both in a local and global manner. Results, displayed

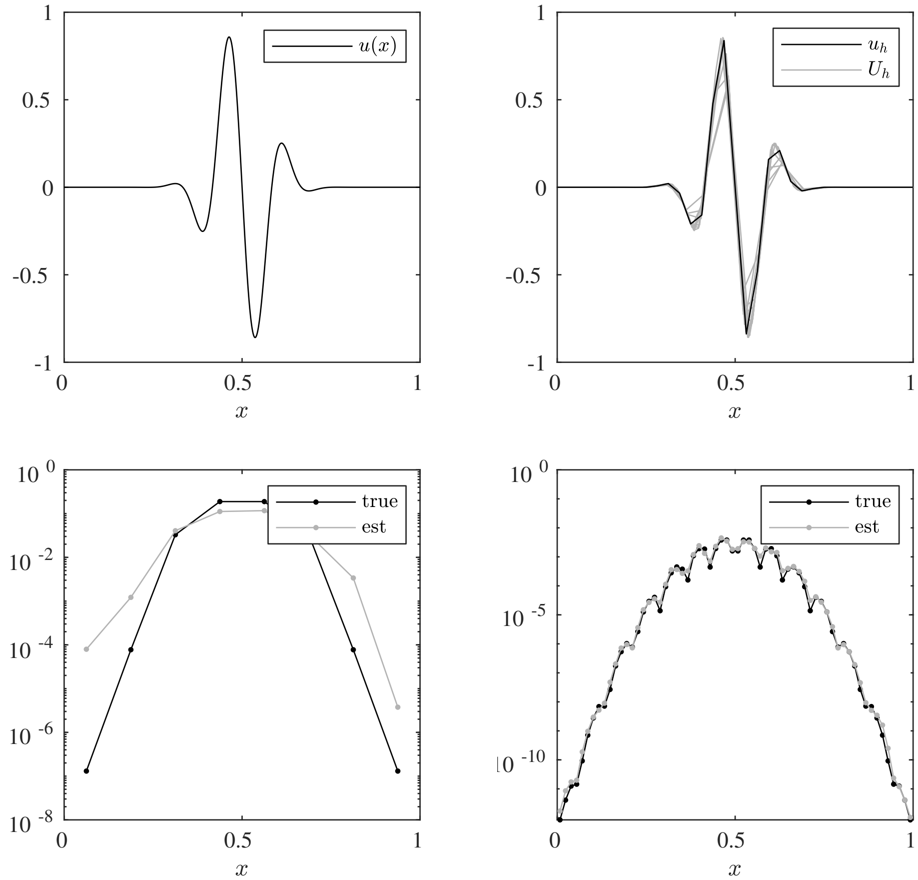

Figure 4: Error estimation for the 1D problem with two different values of h – error in each element.

in Figs. 4 and 5 show that the estimates hold in practice for this case. In particular, in Fig. 5 we can remark that the overall effectivity index $\eta_{\mathcal{X}}$, defined as

$$\eta_{\mathcal{X}} = \frac{\mathbb{E}\|u_h - \widetilde{u}_h\|_{\mathcal{X}}}{\|u_h - u\|_{\mathcal{X}}},$$

with $\mathcal{X} = H_0^1, L^2$, is in this case close to one for both norms. Errors are estimated employing $M = 10$ realisations of the probabilistic solution and with a Monte Carlo simulation.

Two-dimensional case

TO DO

6.3 Mesh adaptivity

Two-dimensional case

See results Fig. 6.

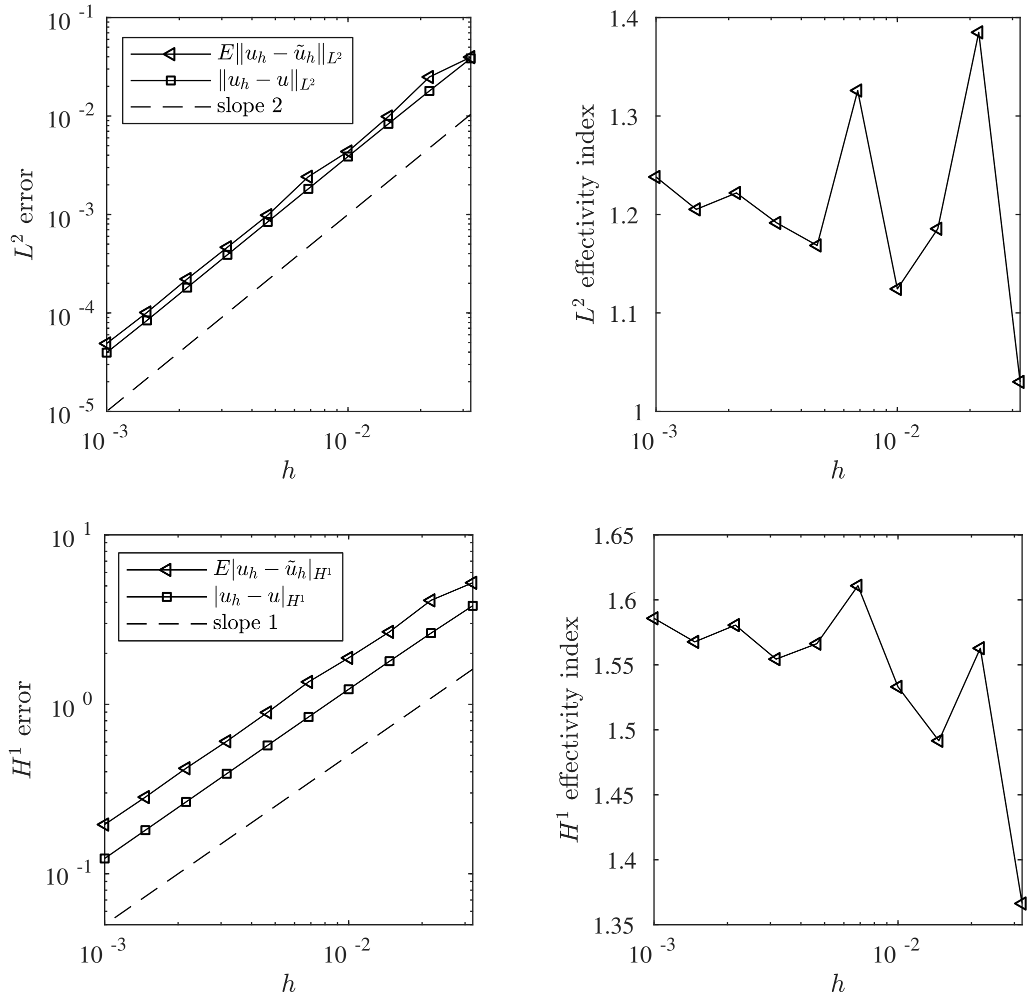

Figure 5: Error estimation for the 1D problem with two different values of h – convergence of error estimators and effectivity indices

6.4 Bayesian inverse problems

Let us consider the following one-dimensional elliptic equation

$$
\begin{aligned}
-\frac{\mathrm{d}}{\mathrm{d}x}\big(e^{\kappa}\frac{\mathrm{d}u}{\mathrm{d}x}\big) &= f, \quad \text{in } (0,1),\\
u &= 0, \quad \text{on } \{0,1\},
\end{aligned}
$$

and the inverse problem of retrieving the field $\kappa \in L^2(0,1)$ given synthetic noisy observations of the solution u corresponding to a true field κ^*. First, we consider a case where information on κ is available beforehand. In particular, we assume that κ has the form

$$
\kappa(x) = \begin{cases} \log(1+\kappa_1), & \text{if } x \in I_1,\\ \log(1+\kappa_2), & \text{if } x \in I_2,\\ 0 & \text{otherwise,} \end{cases} \tag{6.1}
$$

where κ_1, κ_2 are real scalars and I_1, I_2 are the intervals $(0.2, 0.4)$ and $(0.6, 0.8)$ respectively. Fixing a standard Gaussian prior on both parameters κ_1 and κ_2 we are able to compute the posterior distribution corresponding to both the deterministic and probabilistic forward models. In particular,

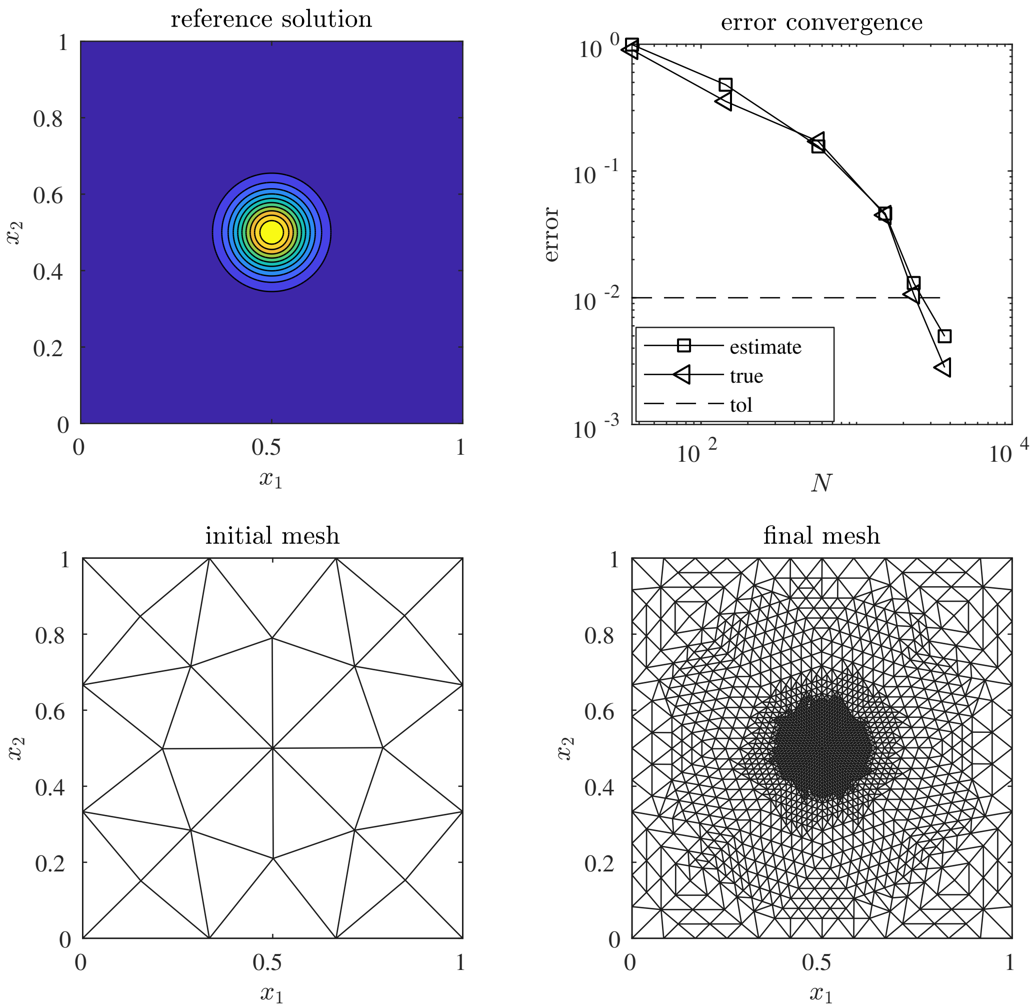

Figure 6: Mesh adaptivity – two-dimensional case

we vary the number of elements N in the set $\{20, 40, 80, 160\}$, thus studying the effects of numerical errors on the numerical posterior distribution. Observations are obtained from a reference solution evaluated at four equispaced points in the interior of $(0, 1)$ each corrupted by an additive source of noise $\varepsilon \sim \mathcal{N}(0, 10^{-4})$. The posterior distributions are obtained with Metropolis–Hastings initialised near the true value of (κ_1, κ_2) and ran as explained in Section 5, with 240 parallel chains employed for the probabilistic forward model. Results are shown in Fig. 8, where TO DO

In a second experiment, we consider the same exact field κ^* and observation model, but without the additional information encoded in (6.1).

References

[1] A. Abdulle and G. Garegnani, *Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration*, Stat. Comput., (2020).

[2] R. E. Bank and R. K. Smith, *A posteriori error estimates based on hierarchical bases*, SIAM J. Numer. Anal., 30 (1993), pp. 921–935.

[3] O. A. Chkrebtii, D. A. Campbell, B. Calderhead, and M. A. Girolami, *Bayesian solution uncertainty quantification for differential equations*, Bayesian Anal., 11 (2016), pp. 1239–1267.

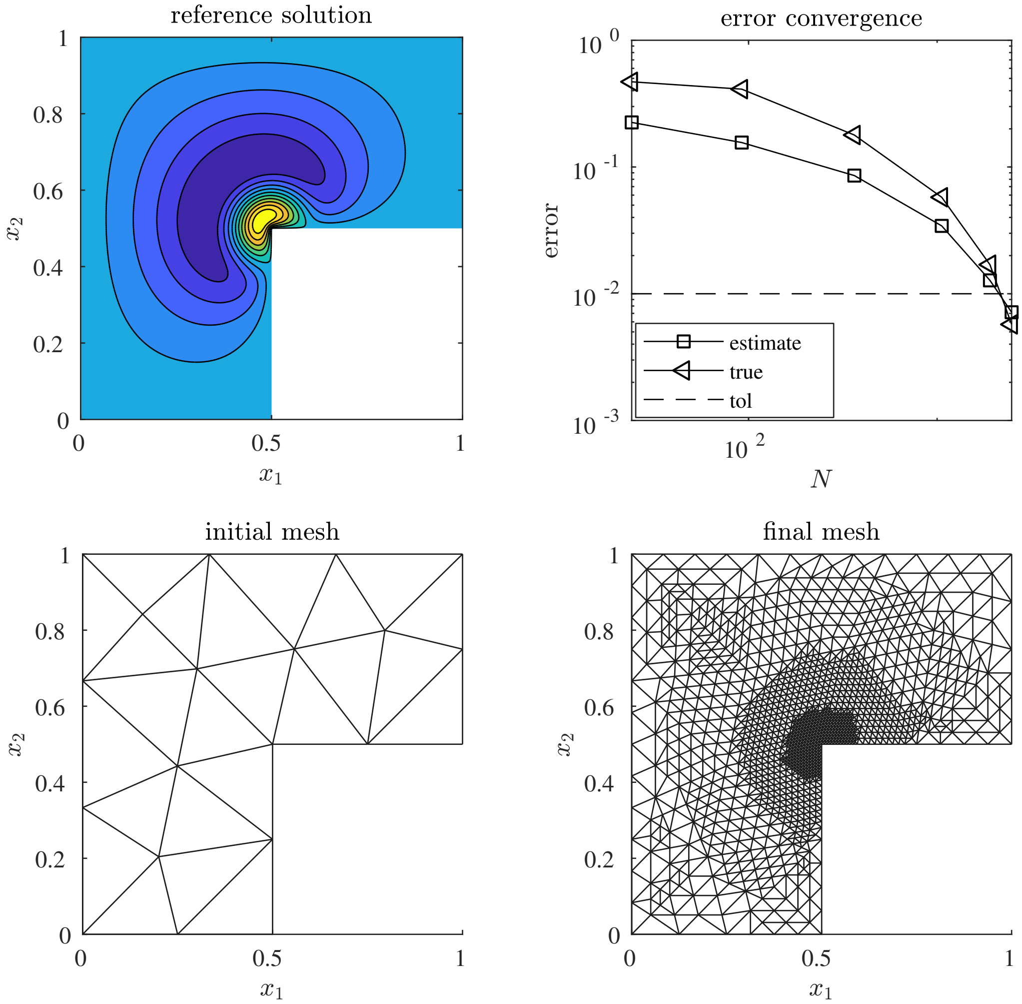

Figure 7: Mesh adaptivity – two-dimensional case

[4] P. G. Ciarlet, *The finite element method for elliptic problems.*, vol. 40 of Classics Appl. Math., SIAM, Philadelphia, 2002.

[5] J. Cockayne, C. J. Oates, T. J. Sullivan, and M. Girolami, *Probabilistic numerical methods for PDE-constrained Bayesian inverse problems*, AIP Conference Proceedings, 1853 (2017), p. 060001.

[6] P. R. Conrad, M. Girolami, S. Särkkä, A. Stuart, and K. Zygalakis, *Statistical analysis of differential equations: introducing probability measures on numerical solutions*, Stat. Comput., 27 (2017), pp. 1065–1082.

[7] M. Croci, M. B. Giles, M. E. Rognes, and P. E. Farrell, *Efficient white noise sampling and coupling for multilevel Monte Carlo with nonnested meshes*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1630–1655.

[8] T. R. Curry and L. Bankoff, *Problems and Solutions: Solutions of Elementary Problems: E1861*, Amer. Math. Monthly, 74 (1967), pp. 724–725.

[9] M. Dashti and A. M. Stuart, *Uncertainty quantification and weak approximation of an elliptic inverse problem*, SIAM J. Numer. Anal., 49 (2011), pp. 2524–2542.

[10] ——, *The Bayesian Approach to Inverse Problems*, in Handbook of Uncertainty Quantification, Springer, 2016, pp. 1–118.

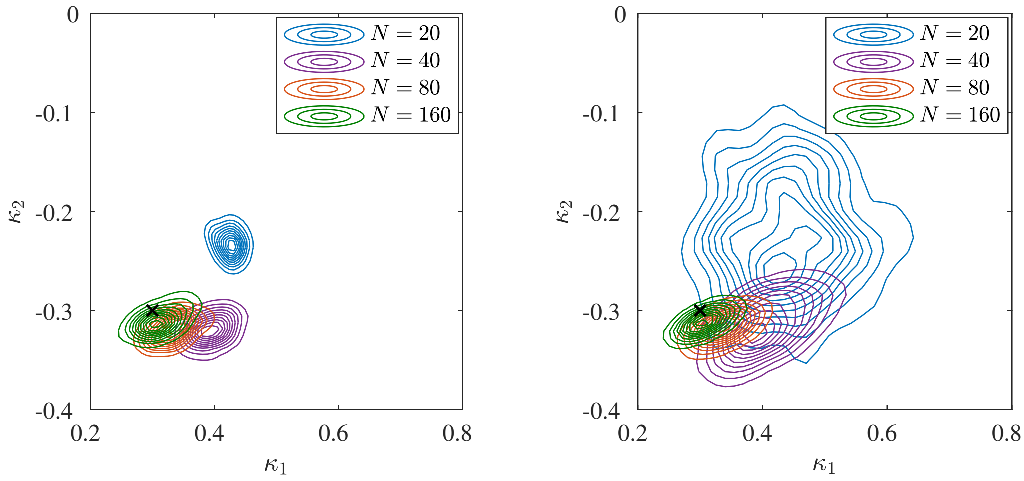

Figure 8: Bayesian inverse problem – finite dimensional case.

[11] P. E. Farrell and J. R. Maddison, *Conservative interpolation between volume meshes by local Galerkin projection*, Comput. Methods Appl. Mech. Engrg., 200 (2011), pp. 89–100.

[12] P. E. Farrell, M. D. Piggott, C. C. Pain, G. J. Gorman, and C. R. Wilson, *Conservative interpolation between unstructured meshes via supermesh construction*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 2632–2642.

[13] P. Hennig, M. A. Osborne, and M. Girolami, *Probabilistic numerics and uncertainty in computations*, Proc. A., 471 (2015), pp. 20150142, 17.

[14] H. Kersting and P. Hennig, *Active uncertainty calibration in Bayesian ODE solvers*, in Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI 2016), AUAI Press, 2016, pp. 309–318.

[15] H. Kersting, T. J. Sullivan, and P. Hennig, *Convergence rates of Gaussian ODE filters*. arXiv preprint arXiv:1807.09737, 2018.

[16] H. C. Lie, T. J. Sullivan, and A. L. Teckentrup, *Random Forward Models and Log-Likelihoods in Bayesian Inverse Problems*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 1600–1629.

[17] A. Quarteroni, *Numerical Models for Differential Problems*, vol. 2 of Modeling, Simulation & Applications, Springer, 2009.

[18] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.

[19] T. J. Sullivan, *Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors*, Inverse Probl. Imaging, 11 (2017), pp. 857–874.

[20] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013.