

A probabilistic finite element method based on random meshes: A posteriori error estimators and Bayesian inverse problems

Assyr Abdulle*, Giacomo Garegnani

Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Switzerland

Received 12 March 2021; received in revised form 21 May 2021; accepted 24 May 2021

Available online xxxx

Abstract

We present a novel probabilistic finite element method (FEM) for the solution and uncertainty quantification of elliptic partial differential equations based on random meshes, which we call random mesh FEM (RM-FEM). Our methodology allows to introduce a probability measure on classical FEMs to quantify the uncertainty due to numerical errors either in the context of a-posteriori error quantification or for FE based Bayesian inverse problems. The new approach involves only a perturbation of the mesh and an interpolation that are very simple to implement. We present a posteriori error estimators and a rigorous a posteriori error analysis based uniquely on probabilistic information for standard piecewise linear FEM. A series of numerical experiments illustrates the potential of the RM-FEM for error estimation and validates our analysis. We furthermore demonstrate how employing the RM-FEM enhances the quality of the solution of Bayesian inverse problems, thus allowing a better quantification of numerical errors in pipelines of computations.

© 2021 Elsevier B.V. All rights reserved.

MSC: 62F15; 65N21; 65N30; 65N50; 65N75

Keywords: Probabilistic methods for PDEs; Random meshes; Uncertainty quantification; A posteriori error estimators; Bayesian inverse problems

1. Introduction

In recent years, there has been a growing interest in developing and analyzing probabilistic counterparts of traditional numerical methods spanning most areas of computational mathematics. This gave rise to the field of Probabilistic Numerics (PN), whose founding principles and aims are summarized in the review papers [1–3]. All methods belonging to the field of PN share the idea of introducing a probability measure on the solution of traditional numerical methods. The underlying rationale is to quantify the uncertainty due to numerical errors in a probabilistic manner, rather than with standard error estimates. Indeed, a probability measure over approximate solutions can be readily pushed through a pipeline of computations, thus justifying the need of probabilistic methods especially when the solution of the problem at hand is employed as the input of a subsequent analysis. A typical example of computational pipelines for which probabilistic methods are successfully employed is given by Bayesian inverse problems, where introducing a probability measure on the forward model allows for a better quantification of the uncertainty in the inversion procedure.

* Corresponding author.

E-mail addresses: assy.abdulle@epfl.ch (A. Abdulle), giacomo.garegnani@epfl.ch (G. Garegnani).

1.1. Literature review

Several contributions to the field of PN concern differential equations. For ordinary differential equations (ODEs), the methodologies can be roughly split in two different areas. In [4–10] the authors present a series of schemes which rely in different measure on Bayesian filtering techniques. These methodologies proceed by updating Gaussian measures over the numerical solution with filtering formulas and evaluations of the right-hand side of the ODE, which are interpreted as observations. While being not involved computationally, analyzing the convergence properties of this class of methods is not always possible, and one can only marginally rely on standard techniques for this purpose. A valuable effort in this sense can be found in [6], where the authors show rates of convergence of the mean of the Gaussian measure towards the exact solution. A different approach is presented in the series of works [11–15], where the authors propose probabilistic schemes which are based on perturbing randomly the approximate solution and on letting evolve these perturbations through the dynamics of the ODE. In this manner, it is possible to obtain empirical probability measures over the otherwise deterministic numerical solution. A random perturbation can be applied directly to the state, as it was presented and analyzed for one-step methods in [11,12], with a particular focus on implicit schemes in [14] and for multistep methods in [15]. Another approach, which was presented in [13], consists in perturbing the scheme itself by randomizing the time steps of a Runge–Kutta method. This allows to maintain certain geometric properties of the deterministic scheme in its probabilistic counterpart, such as the conservation of invariants or the symplecticity.

There has been a keen interest from the PN community on developing probabilistic numerical solvers for partial differential equations (PDEs), too [9,11,16–24]. In [17], the authors present a meshless Bayesian method for PDEs, which they then apply to inverse problems in [16], and in particular to a challenging time-dependent instance drawn from an engineering application in [18]. Their methodology consists of placing a Gaussian prior on the space of solutions, thus updating it with evaluations of the right-hand side, which are interpreted as noisy observations. A similar idea has been presented in [9], where the main focus are time-dependent problems, and in [22,23], where the method is recast in the framework of machine learning algorithms. In [20,21], a probabilistic approach involving gambles is applied to the solution of PDEs with rough coefficients and by multigrid schemes, with a particular interest to reducing the complexity of implicit algorithms for time-dependent problems [21]. Moreover, in [19] the author presents a Bayesian reinterpretation of the theory of homogenization for PDEs, which can be seen as a contribution to the field of PN. To our knowledge, the only perturbation-based finite element (FE) probabilistic scheme for PDEs is presented in [11], where the authors randomize FE bases by adding random fields endowed with appropriate boundary conditions, thus obtaining an empirical measure over the space of solutions. By tuning the covariance of these random fields, they obtain a consistent characterization of the numerical error, which can then be employed to solve Bayesian inverse problems and to quantify the uncertainty over their numerical solution.

1.2. Our contributions

In this work we present a probabilistic finite element method (FEM) which is based on a randomization of the mesh, and which we call RM-FEM. To put our method into the context of the vast literature on FEM and a posteriori error analysis let us briefly motivate the RM-FEM. First, in the limit $h \rightarrow 0$, the random meshes used in RM-FEM tend to coincide and therefore the RM-FEM solution tends to be equal to the standard FEM solution. Our methodology, as other probabilistic numerical methods, is particularly relevant and helpful in the non-asymptotic limit, when one wishes to quantify the uncertainty due to *numerical approximation* in a statistical manner in order for the numerical error to be in the same range than errors arising from data or model uncertainties. Second, the methodology employed in the RM-FEM can be applied to most of the numerous FEM solvers that have been successfully developed over the years. As such it is a new tool that can build on existing methods when relevant but in any case replace them.

A direct engineering motivation for our paper is then provided by the quantification via RM-FEM of the numerical error introduced by FEM discretization in the forward map when solving inverse problems. Inverse problems such as the one we considered in our work are in fact ubiquitous in engineering and natural sciences, and improving their solution qualitatively is clearly of the utmost importance. Again in the limit $h \rightarrow 0$ there will be no difference between FEM and RM-FEM but this limit is of course not reasonable, specially due to the high computational cost in solving inverse problems. Balancing the errors in order for the numerical discretization to be in the same range of

error as the observational error is crucial for efficient computation. Having a statistical description of the numerical error allows to achieve such a computational balance.

A second motivation is given by the novel probabilistic interpretation of a posteriori error estimators. Valid (i.e., reliable and efficient) a posteriori error estimators often involve tedious calculations and complex implementations in order to be computed. We believe that our method, which only involves a perturbation of the mesh and an interpolation, can be employed with success in mesh adaptation and be useful for ranges of problems as illustrated by our numerical experiments.

The idea underlying our method stems from both [11], where the authors propose a probabilistic FEM based on random perturbations, and from [13], in which the first instance of randomizing the discretization instead of the solution itself is presented. In the context of ODEs, a careful randomization of the time step in Runge–Kutta methods allows to maintain certain convergence and geometric properties, either path-wise or in the mean-square sense. In a similar fashion, creating a probability measure on the space of solutions by randomizing the mesh has the advantage that each sample is a FEM solution itself, and therefore a projection of the exact solution on some random finite-dimensional space. Keeping in mind the fundamental goal of PN, we consider the problem of employing probabilistic methods to quantify numerical errors in the context of PDEs. Indeed, in [6,11,13] and other works concerning ODEs, the authors show that the probabilistic solution converges to the true solution with the same rate as the deterministic method, which represents a consistency result. No work so far shows that PN methods can be readily employed for an a posteriori estimation of the error. Some forms of adaptivity for nonlinear ODEs based on probabilistic information can be found in [8,25,26], where the arguments are based on heuristics but are not rigorously analyzed. In this work, we construct and present a posteriori error estimators which can be readily employed for mesh adaptation in elliptic PDEs. Our estimators are entirely based on probabilistic information, are simple to compute and do not entail considerable computational cost. We present an analysis in the one-dimensional case that shows that our error estimators based on the RM-FEM are equivalent to a classical estimator by Babuška and Rheinboldt [27], which employs the jumps of the derivative of the solution at the nodes to quantify the numerical errors. Our one-dimensional theoretical analysis is complemented by a series of numerical experiments confirming the validity of our theory in higher dimensions.

As stated above, probabilistic numerical methods are especially appealing when employed in pipeline of computations such as Bayesian inverse problems. In particular, employing deterministic methods for approximating forward maps leads to overly confident posterior measures, which can be corrected by appropriate probabilistic approximations. Similarly to [11,13], we show in this paper how the RM-FEM can be employed to construct empirical distributions over the forward problem and compute a random posterior measure, solution to the inverse problem in the Bayesian sense. The solution is consistent asymptotically with respect to the mesh spacing, but its quality is enhanced if the latter is relatively large, i.e., if the forward model is approximated cheaply.

1.3. Outline

The outline of the paper is as follows. In Section 2 we state the problem of interest, introduce the RM-FEM and the main assumptions and notation required by our analysis. We then present the two main applications of the RM-FEM, i.e., a posteriori error estimators and Bayesian inverse problems, in Sections 3 and 4, respectively. For both applications, a series of numerical experiments in the one and two-dimensional cases illustrate the usefulness and efficiency of the RM-FEM. In Section 5 we present a rigorous a priori and a posteriori error analysis. Finally, in Section 6 we draw our conclusions.

2. Random mesh finite element method

2.1. Notation

Let $d = 1, 2, 3$ and $D \subset \mathbb{R}^d$ be an open bounded domain with sufficiently smooth boundary ∂D . For $v \in \mathbb{R}^d$, we denote by $\|v\|_2$ the Euclidean norm on \mathbb{R}^d . We denote by $L^2(D)$ the space of square integrable functions, by (\cdot, \cdot) the natural $L^2(D)$ inner product, and by $H^p(D)$ the Sobolev space of functions with p weak derivatives in $L^2(D)$. Moreover, we denote by $H_0^1(D)$ the space of functions in $H^1(D)$ vanishing on ∂D in the sense of traces, by $H^{-1}(D)$ the dual of $H_0^1(D)$ and by $\langle \cdot, \cdot \rangle$ the natural pairing between $H^{-1}(D)$ and $H_0^1(D)$. We equip the space $H_0^1(D)$ with the norm $\|v\|_{H_0^1(D)} = \|\nabla v\|_{L^2(D)}$, i.e. the $H^1(D)$ seminorm.

For an event space Ω , with a σ -algebra \mathcal{A} and a probability measure P , we let the triple (Ω, \mathcal{A}, P) denote a probability space. For an event $A \in \mathcal{A}$, we say that A occurs almost surely (a.s.) if $P(A) = 1$. For $n \in \mathbb{N}$ we call random variables the measurable functions $X: \Omega \rightarrow \mathbb{R}^n$, and denote by $L^2(\Omega)$ the space of square integrable random variables, with associated inner product. Denoting by $\mathcal{B}(\mathbb{R}^n)$ the Borel σ -algebra on \mathbb{R}^n , we say that a probability measure μ_X on the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ satisfying $\mu_X(B) = P(X^{-1}(B))$ for all $B \in \mathcal{B}(\mathbb{R}^n)$ is the measure induced by X , or equivalently the distribution of X . For a set of random variables $\{X_i\}_{i=1}^n$ which are independent and identically distributed, we say they are i.i.d., and denoting by μ their common induced measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, we write $\{X_i\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} \mu$.

2.2. Problem and method presentation

Let $\kappa \in L^\infty(D, \mathbb{R}^{d \times d})$, $f \in H^{-1}(D)$ and u be the weak solution of the partial differential equation (PDE)

$$\begin{aligned} -\nabla \cdot (\kappa \nabla u) &= f, \quad \text{in } D, \\ u &= 0, \quad \text{on } \partial D, \end{aligned} \quad (2.1)$$

i.e., the function $u \in V \equiv H_0^1(D)$ satisfying

$$a(u, v) = F(v), \quad a(u, v) := \int_D \kappa \nabla u \cdot \nabla v \, dx, \quad F(v) := \langle f, v \rangle, \quad (2.2)$$

for all functions $v \in V$. We assume there exist positive constants $\underline{\kappa}$ and $\bar{\kappa}$ such that for all $\xi \in \mathbb{R}^d$

$$\underline{\kappa} \|\xi\|_2^2 \leq \kappa \xi \cdot \xi \leq \bar{\kappa} \|\xi\|_2^2,$$

where $\|\cdot\|_2$ is the Euclidean norm on \mathbb{R}^d , so that there exist constants $m, M > 0$ such that for all $u, v \in V$ it holds

$$|a(u, v)| \leq M \|u\|_V \|v\|_V, \quad |a(u, u)| \geq m \|u\|_V^2.$$

The Lax–Milgram theorem then guarantees that the problem (2.2) is well-posed.

Let N be a positive integer and let $\mathcal{T}_h = \bigcup_{i=1}^N K_i$ be a partition of D , where for all $i = 1, \dots, N$, the element $K_i \subset D$ is a segment, triangle or tetrahedron for $d = 1, 2, 3$ respectively. We denote by $h_i = \text{diam}(K_i)$ the radius of the smallest ball containing K_i , and by $h = \max_i h_i$ the maximum radius, indexing the mesh \mathcal{T}_h . We denote by \mathcal{V}_h the set of all vertices of the elements of \mathcal{T}_h , and in particular as $\mathcal{V}_h^I \subset \mathcal{V}_h$ the set of vertices which do not lie on the boundary of D , and by $\mathcal{V}_h^B = \mathcal{V}_h \setminus \mathcal{V}_h^I$. Moreover, we denote by N_I the number of internal vertices, i.e., $N_I = |\mathcal{V}_h^I|$. We assume the partition to be conforming, i.e., if two elements have non-empty intersection, then the latter consists of a point (for $d = 1$), of either a vertex or a side (for $d = 2$), and of either a vertex, a segment or a face (for $d = 3$). We then denote by $V_h \subset V$, $\dim(V_h) < \infty$ the space of continuous piecewise linear finite elements on \mathcal{T}_h , i.e.,

$$V_h := \{v \in V : v|_K \in \mathbb{P}_1, \forall K \in \mathcal{T}_h\},$$

where \mathbb{P}_1 is the space of linear functions. Let us remark that imposing $u_h = 0$ on ∂D yields $\dim(V_h) = N_I$. The FEM proceeds by finding $u_h \in V_h$ such that

$$a(u_h, v_h) = F(v_h), \quad (2.3)$$

for all $v_h \in V_h$, which is equivalent to solving the linear system $A\mathbf{u} = \mathbf{f}$, where

$$\mathbf{u}_j = u_h(x_j), \quad x_j \in \mathcal{V}_h^I, \quad A_{ij} = a(\varphi_j, \varphi_i), \quad \mathbf{f}_j = F(\varphi_j), \quad i, j = 1, \dots, N_I,$$

and where $\{\varphi_j\}_{j=1}^{N_I}$ are the Lagrange basis functions defined on the internal vertices of \mathcal{T}_h . The assumptions on κ guarantee that A is symmetric positive definite, and in turn that \mathbf{u} is uniquely defined and the problem (2.3) is well-posed.

We now introduce the random-mesh finite element method (RM-FEM), which is based on a random perturbation of the mesh \mathcal{T}_h obtained by moving the internal vertices. First, we here detail how we build perturbed meshes and which kind of random perturbations we consider to be admissible. Let $p \geq 1$, $\alpha := \{\alpha_i : \Omega \rightarrow \mathbb{R}^d\}_{i=1}^{N_I}$ be a sequence of random variables and let us define the set of internal points $\tilde{\mathcal{V}}_h^I = \{\tilde{x}_i\}_{i=1}^{N_I}$ where

$$\tilde{x}_i := x_i + h^p \alpha_i. \quad (2.4)$$

We then define the set of perturbed vertices as $\tilde{\mathcal{V}}_h = \tilde{\mathcal{V}}_h^I \cup \mathcal{V}_h^B$, i.e., the vertices on the boundary are left unchanged. The perturbed mesh is then simply $\tilde{\mathcal{T}}_h = \bigcup_{i=1}^N \tilde{K}_i$, where each element \tilde{K}_i has the same vertices as its corresponding element K_i in the original mesh, modulo the random perturbation (2.4). In other words, we compute the internal points of the perturbed mesh following (2.4), and keep the connectivity structure of the original mesh \mathcal{T}_h . Clearly, the mesh so defined is not conforming for any sequence of random variable α , for which we therefore introduce an assumption.

Assumption 2.1. The sequence of random variables α is such that

- (i) its components α_i admit densities F_{α_i} with respect to the Lebesgue measure on \mathbb{R}^d , which satisfy $\text{supp}(F_{\alpha_i}) \subset B_{r_i}$, where $B_{r_i} \subset \mathbb{R}^d$ is the ball centered in the origin and of radius $r_i > 0$, and which are radial, i.e., $F_{\alpha_i}(x) = F_{\alpha_i}(\|x\|_2)$,
- (ii) the perturbed mesh $\tilde{\mathcal{T}}_h$ is conforming a.s.

Let us remark that the assumption (i) actually implies for all $p \geq 1$ the assumption (ii) a.s., provided the radii r_i are chosen small enough. We assume in (i) the densities F_{α_i} to be radial functions so that the random perturbations do not have a privileged direction.

Example 2.2. In the one-dimensional case, let $0 = x_0 < x_1 < \dots < x_N = 1$ so that we have $N_I = N - 1$. Denoting $K_i = (x_i, x_{i+1})$ we call \tilde{h}_i the minimum element size for the two intervals sharing the point x_i as a vertex, i.e., $\tilde{h}_i := \min\{h_i, h_{i+1}\}$. Then, a choice of random variables satisfying Assumption 2.1 is given by

$$\alpha_i = (h^{-1}\tilde{h}_i)^p \tilde{\alpha}_i, \quad i = 1, \dots, N-1, \quad \{\tilde{\alpha}_i\}_{i=1}^{N-1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}\left(\left(-\frac{1}{2}, \frac{1}{2}\right)\right),$$

where for a set $D \in \mathbb{R}^d$ we denote by $\mathcal{U}(D)$ the uniform distribution over D . With this choice, indeed, we have that $\tilde{x}_{i-1} < \tilde{x}_i$ a.s., and therefore the perturbed mesh is conforming. In the two-dimensional case, we introduce for $i = 1, \dots, N_I$ the notation

$$\Delta_i = \{K \in \mathcal{T}_h : K \text{ has } x_i \text{ as a vertex}\}.$$

Analogously to the one-dimensional case, we write $\tilde{h}_i := \min_{j:K_j \in \Delta_i} h_j$. In this case, it is possible to verify that choosing for all $i = 1, \dots, N_I$

$$\alpha_i = (h^{-1}\tilde{h}_i)^p \tilde{\alpha}_i, \quad i = 1, \dots, N_I, \quad \{\tilde{\alpha}_i\}_{i=1}^{N_I} \stackrel{\text{i.i.d.}}{\sim} \mathcal{U}(B_{1/2}),$$

then α satisfies Assumption 2.1. We verify this graphically in Fig. 1, where we show a realization of the perturbed mesh based on a generic Delaunay mesh and on a structured mesh on $D = (0, 1)^2$ along with the sets where the perturbed points are constrained to belong a.s. We notice that for $p > 1$ the magnitude of the perturbations clearly tends to vanish. Finally, we remark that similar admissible perturbations can be introduced in higher dimensions.

Having defined the perturbed mesh, we now proceed with describing the RM-FEM. Let \tilde{V}_h be the space of continuous piecewise linear finite elements on $\tilde{\mathcal{T}}_h$. Let moreover $\{\tilde{\varphi}_i\}_{i=1}^{N_I}$ be the Lagrange basis functions defined on the internal vertices of $\tilde{\mathcal{T}}_h$ and $\tilde{\mathcal{I}}: \mathcal{C}^0(D) \cap V \rightarrow \tilde{V}_h$ be the Lagrange interpolation operator onto \tilde{V}_h , i.e., for a function $v \in \mathcal{C}^0(D) \cap V$ and for $x \in D$ we define

$$\tilde{\mathcal{I}}v(x) := \sum_{i=1}^{N_I} v(x_i) \tilde{\varphi}_i(x).$$

We are then interested in the two functions belonging to the finite element space \tilde{V}_h whose definition we give below.

Definition 2.3. Let $u_h \in V_h$ be defined in (2.3). We define the RM-FEM interpolant as the random function $\tilde{\mathcal{I}}u_h \in \tilde{V}_h$, where $\tilde{\mathcal{I}}$ is the Lagrange interpolant onto \tilde{V}_h .

Definition 2.4. Given the random finite element space \tilde{V}_h , we define the RM-FEM solution as the unique random function $\tilde{u}_h \in \tilde{V}_h$ such that

$$a(\tilde{u}_h, \tilde{v}_h) = F(\tilde{v}_h),$$

for all $\tilde{v}_h \in \tilde{V}_h$.

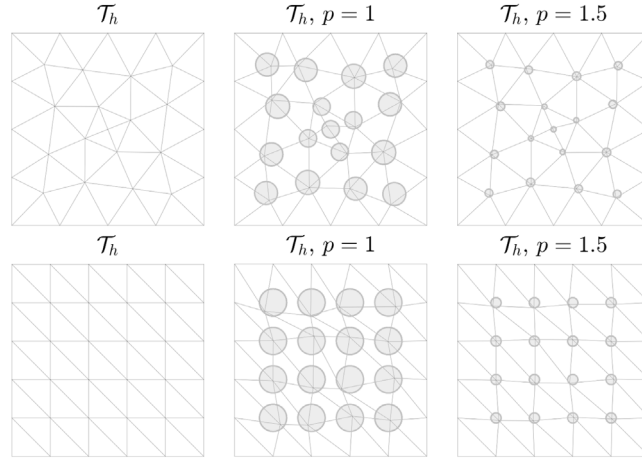


Fig. 1. A realization of \tilde{T}_h for $p = \{1, 1.5\}$ based on two meshes T_h of $D = (0, 1)^2$. On the first line, a Delaunay mesh. On the second line, a structured mesh. The regions where the perturbed points are included a.s. are depicted by light gray circles.

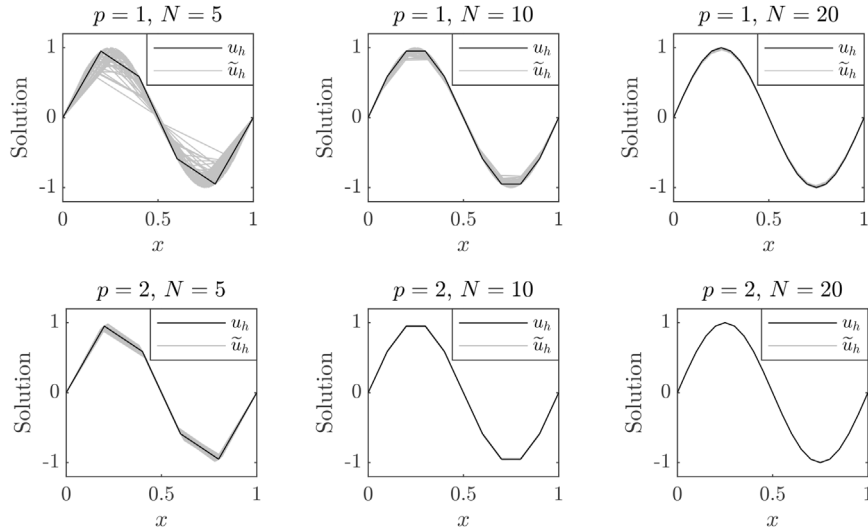


Fig. 2. Comparison between the RM-FEM and the FEM solutions. We display the solution u_h and 50 realizations of \tilde{u}_h , by row respectively for $p = \{1, 2\}$ and by column for $N = \{5, 10, 20\}$.

Remark 2.5. Clearly, either for any fixed $p \geq 1$ and $h \rightarrow 0$ or for any fixed $h < 1$ and $p \rightarrow \infty$, the functions u_h , \tilde{u}_h and \tilde{u}_h tend to coincide. We visualize this for u_h and \tilde{u}_h in Fig. 2, where we simply fix $\kappa = 1$ and the right-hand side f such that $u = \sin(2\pi x)$ in (2.1), and consider the effects of increasing p and decreasing h . For this simple problem, we notice that for $p = 2$ and $N = 20$ the FEM solution u_h and the RM-FEM solution \tilde{u}_h are almost indistinguishable.

Remark 2.6. All the quantities distinguished by a tilde (e.g., \tilde{T}_h , \tilde{V}_h , \tilde{L}) are random variables with values in appropriate spaces. For example \tilde{u}_h is a random function $\tilde{u}_h: \Omega \times D \rightarrow \mathbb{R}$, such that $\Omega \times D \ni (\omega, x) \mapsto \tilde{u}_h(\omega, x)$. For economy of notation, in the following we drop the argument ω from all random variables.

Remark 2.7. The coefficient p in (2.4) has the same role as the coefficient identified by the same symbol in both [11,13], i.e., it controls the variability of the probabilistic solutions by tuning the variability of the noise which is applied to the method.

Remark 2.8. The RM-FEM interpolant $\tilde{\mathcal{I}}u_h$ is well-defined even allowing the vertices of \mathcal{T}_h which lay on the boundary ∂D to be perturbed, as far as the perturbation moves them inside the domain D . The random RM-FEM interpolant $\tilde{\mathcal{I}}u_h$ does not in this case belong to the space V since it is not defined on the whole domain D and does not satisfy boundary conditions. For practical applications, one can nevertheless employ the RM-FEM interpolant defined on a smaller domain, which results from a perturbation of all vertices of \mathcal{T}_h , including those on the boundaries.

Before proceeding with the two main applications of the RM-FEM, i.e., a posteriori error estimators and Bayesian inverse problems, we state an a priori error estimate, which suggests how to balance the sources of error due to numerical discretization and to the randomization of the method, respectively.

Theorem 2.9. *Let the solution u of (2.1) be $u \in H^2(D)$. Then, it holds*

$$\|\tilde{u}_h - u\|_V \leq Ch, \quad \text{a.s.},$$

for a constant $C > 0$ independent of h . Moreover, if $p = 1$ in (2.4) the numerical and discretization errors are balanced with respect to h , i.e., it holds

$$\|u_h - \tilde{u}_h\|_V = \mathcal{O}(h) = \mathcal{O}(\|u - u_h\|_V), \quad \text{a.s.}$$

This result indicates that one should fix $p = 1$ in (2.4) in order to obtain a family of probabilistic solutions whose statistical properties should reflect the true error. This is crucial when the RM-FEM is employed in a pipeline of computations such as Bayesian inverse problems, which will be presented in detail in Section 4. The proof of Theorem 2.9 is elementary and discussed in Section 5.1.

3. A posteriori error estimators based on the RM-FEM

The first and foremost application of the RM-FEM is deriving a posteriori error estimators which are entirely based on the statistical information carried on by the mesh perturbation. We say that a quantity \mathcal{E}_h is an a posteriori error estimator if it gives an error estimate on the numerical approximation and is computable only by knowledge of the numerical solution. Moreover, if there exist constants C_{up} and C_{low} independent of h and of u such that

$$C_{\text{low}}\mathcal{E}_h \leq \|u - u_h\|_V \leq C_{\text{up}}\mathcal{E}_h, \quad (3.1)$$

we say that the a posteriori error estimator is reliable and efficient, respectively. Indeed, the upper bound above guarantees that when the estimator is small, so is the numerical error. The lower bound, instead, gives an insurance on the quality of the estimator, as it shows that the estimation of the error is not exceedingly pessimistic. There exist in the literature a huge number of a posteriori error estimators, and we refer the reader to the surveys given e.g. in [28,29]. Most a posteriori error estimators are expressed in the form

$$\mathcal{E}_h = \left(\sum_{K \in \mathcal{T}_h} \eta_K^2 \right)^{1/2},$$

where the η_K are local quantities depending on the solution and the data on the element K and its neighbors. For example, in the two-dimensional case a valid a posteriori error estimator is given by the expression of its local components

$$\eta_K^2 = h_K^2 \|f\|_{L^2(K)}^2 + h_K \|\llbracket \nabla u_h \cdot \nu_K \rrbracket\|_{L^2(\partial K)}^2,$$

where $\llbracket \cdot \rrbracket$ is the jump operator and ν_K denotes the unitary vector normal to the boundary of K (see e.g. [30, Section 3] or [29, Chapter 2]). Other a posteriori error estimators are based on recovered gradients, which are employed as surrogates of the gradient of the exact solution to estimate the error. A notable member of these methodologies is the Zienkiewicz–Zhu (ZZ) patch recovery technique [31,32], which is proved to be superconvergent on special meshes, and which is in practice widely employed on any mesh.

It has been heuristically noted for ODEs in [8,25,26] that information on the variability of a probabilistic solution can be employed to estimate the error and thus adapt the numerical discretization. Indeed, building probabilistic solution to otherwise deterministic problems should pursue the goal of quantifying numerical errors through uncertainty. Guided by this observation, we now introduce two probabilistic error estimators for elliptic PDEs.

Definition 3.1. Let $\tilde{\mathcal{I}}_{u_h}$ be the RM-FEM interpolant defined in Definition 2.3 and for each $K \in \mathcal{T}_h$, let us denote by $\tilde{K} \in \tilde{\mathcal{T}}_h$ its corresponding element in $\tilde{\mathcal{T}}_h$. We define the first RM-FEM a posteriori error estimator as

$$\tilde{\mathcal{E}}_{h,1} := \left(\sum_{K \in \mathcal{T}_h} \tilde{\eta}_{K,1}^2 \right)^{1/2}, \quad \text{with} \quad \tilde{\eta}_{K,1}^2 = h_K^{-(p-1)} \mathbb{E} \left[\|\nabla(u_h - \tilde{\mathcal{I}}_{u_h})\|_{L^2(\tilde{K})}^2 \right].$$

Moreover, we define the second RM-FEM a posteriori error estimator as

$$\tilde{\mathcal{E}}_{h,2} := \left(\sum_{K \in \mathcal{T}_h} \tilde{\eta}_{K,2}^2 \right)^{1/2}, \quad \text{with} \quad \tilde{\eta}_{K,2}^2 = h_K^{-(2p-2)} |K| \mathbb{E} \left[\|\nabla u_h|_K - \nabla \tilde{\mathcal{I}}_{u_h}|_{\tilde{K}}\|^2 \right].$$

Remark 3.2. The scaling factors $h_K^{-(p-1)}$ and $h_K^{-(2p-2)}$ in the definition of $\tilde{\eta}_{K,1}$ and $\tilde{\eta}_{K,2}$ are necessary to obtain well-calibrated error estimators. This is made clearer in the one-dimensional case by the analysis presented in Section 5.2. For higher dimensions, they can be partially explained with the ansatz (5.2), especially for the first estimator $\tilde{\mathcal{E}}_{h,1}$, and they appear in practice to be the correct scaling.

Remark 3.3. It is computationally more involved to compute the estimator $\tilde{\mathcal{E}}_{h,1}$ than the estimator $\tilde{\mathcal{E}}_{h,2}$. Indeed, for the latter it is sufficient to compute the interpolant $\tilde{\mathcal{I}}_{u_h}$ and the gradients over each element of u_h and of the interpolant. For $\tilde{\mathcal{E}}_{h,1}$, instead, one has to compute on each element \tilde{K} the quantity

$$\|\nabla(u_h - \tilde{\mathcal{I}}_{u_h})\|_{L^2(\tilde{K})}.$$

By construction, each element \tilde{K} overlaps with the elements corresponding to its neighbors in the original mesh in a non-trivial manner, and if $d > 1$ one has to rely to the construction of a “super-mesh” (see e.g. [33,34]) such that on each of its elements the quantity $\nabla(u_h - \tilde{\mathcal{I}}_{u_h})$ is constant. A super-mesh has to be built for each realization of the perturbed mesh $\tilde{\mathcal{T}}_h$, which could therefore be expensive.

In this article, we show in the one-dimensional case that the estimators given in Definition 3.1 are reliable and efficient in the sense of (3.1). In the statement of our theoretical result, which is given below, we make use of a quantity $\Lambda \in \mathbb{R}$ which is of higher order in most practical scenarios and which is defined as

$$\Lambda^2 := h^\zeta \sum_{j=1}^N \int_{K_j} (f(x) + C_j)^2 dx, \quad (3.2)$$

where $\zeta \in (0, 1)$ is arbitrary and where for each K_j the real constant C_j will be specified in the analysis of Section 5.2 (see e.g. [27, Equation (8.7)]). Moreover, we consider one-dimensional meshes which are λ -quasi-uniform, i.e., we assume there exists a constant $\lambda \in (1, \infty)$ such that it holds

$$\frac{1}{\lambda} \leq \frac{h_j}{h_{j-1}} \leq \lambda, \quad j = 2, \dots, N,$$

uniformly in h . Finally, we consider perturbations satisfying

$$\alpha_i = (h^{-1} \bar{h}_i)^p \bar{\alpha}_i, \quad i = 1, \dots, N-1,$$

where $\bar{h}_i = \min\{h_i, h_{i+1}\}$ and for an i.i.d. sequence of random variables $\{\bar{\alpha}_i\}_{i=1}^{N-1}$ such that $|\bar{\alpha}_1| \leq 1/2$ a.s. These perturbations are indeed the same as the ones presented in Example 2.2, but without the assumption of $\{\bar{\alpha}_i\}_{i=1}^N$ to be uniformly distributed, which is not necessary in the following. In practice, a uniform distribution is nevertheless advisable, as it is still general enough and satisfies the radial assumption of Assumption 2.1(i). We moreover introduce the following technical assumption on the perturbation.

Assumption 3.4. Let the family of meshes \mathcal{T}_h be λ -quasi-uniform, let p be the coefficient introduced in (2.4) and assume that for all h and p there exists $C > 0$ such that

$$4h^{p-1} \frac{\mathbb{E} |\bar{\alpha}_1|^2}{\mathbb{E} |\bar{\alpha}_1|} + C < 1 + \lambda^{-(p-1)}.$$

Remark 3.5. We note that [Assumption 3.4](#) holds for $p > 1$ and h sufficiently small, and is therefore not restrictive in practice.

We can now state the main result involving a posteriori error estimators that will be proved in [Section 5.2](#).

Theorem 3.6. *Let the dimension $d = 1$, let $p > 1$ in [\(2.4\)](#) and let [Assumption 2.1](#) hold. Moreover, let $\tilde{\mathcal{E}}_{h,1}$, $\tilde{\mathcal{E}}_{h,2}$ and Λ be given in [Definition 3.1](#) and [\(3.2\)](#) respectively and let the family of meshes \mathcal{T}_h be λ -quasi-uniform. Then, there exists $C > 0$ independent of h and of the solution u such that it holds for $k \in \{1, 2\}$*

$$\|u - u_h\|_V \leq \tilde{C}(\tilde{\mathcal{E}}_{h,k}^2 + \Lambda^2)^{1/2},$$

up to higher order terms in h and under [Assumption 3.4](#) for $k = 1$. If additionally $\kappa \in C^2(D)$ and $f \in C^1(D)$, then there exist constants \tilde{C}_{low} and \tilde{C}_{up} independent of h and of the solution u such that for $k \in \{1, 2\}$ it holds

$$\tilde{C}_{\text{low}}\tilde{\mathcal{E}}_{h,k} \leq \|u - u_h\|_V \leq \tilde{C}_{\text{up}}\tilde{\mathcal{E}}_{h,k},$$

up to higher order terms in h and under [Assumption 3.4](#) for $k = 1$.

Let us notice that the estimators given in [Definition 3.1](#) involve the computation of an expectation with respect to the random perturbations of the mesh, and therefore a Monte Carlo simulation is needed in practice. Let N_{MC} be a positive integer, $k \in \{1, 2\}$ and $\{\tilde{E}_{h,k}^{(i)}\}_{i=1}^{N_{\text{MC}}}$ be i.i.d. realizations of the estimator $\tilde{\mathcal{E}}_{h,k}$, obtained with independent perturbations of the mesh. Then, in practice we compute

$$\tilde{E}_{h,k} := \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \tilde{E}_{h,k}^{(i)}. \quad (3.3)$$

Remark 3.7. It could be suggested that the application of Monte Carlo techniques increases dramatically the simulation time. We argue that in practice the computational overhead is not relevant, mainly for three reasons. First, it has been proved in [\[13\]](#) that the variance of Monte Carlo estimators drawn from probabilistic numerical methods decreases with respect to the discretization size h . Hence, the number of simulations M does not need to be large, nor increasing if $h \rightarrow 0$, to guarantee a good quality of the estimator. The same arguments hold for the RM-FEM, too. Second, the Monte Carlo estimation is completely parallelizable, thus reducing the cost by a factor equal to the number of available computing units, but still with the drawback of increasing the overall computational effort. Finally, the computation of the RM-FEM interpolant $\tilde{\mathcal{I}}_h$ is not computationally involved, neither when it is repeated N_{MC} times.

3.1. Numerical experiments

We now present numerical experiments on one and two-dimensional test cases to demonstrate the validity of our a posteriori error estimators. In particular, we are interested in determining whether the probabilistic error estimators introduced in [Definition 3.1](#) are indeed reliable estimators for the numerical error in the FEM, and in employing these estimators for local refinements of the mesh. Setting a tolerance $\gamma > 0$, our goal is building a mesh \mathcal{T}_h such that

$$\frac{\|u - u_h\|_V}{\|u_h\|_V} \leq \gamma. \quad (3.4)$$

Replacing the numerator with $\tilde{\mathcal{E}}_{h,k}$, $k \in \{1, 2\}$, we notice that the condition [\(3.4\)](#) is satisfied if it holds for all $K \in \mathcal{T}_h$

$$\tilde{\eta}_{K,k} \leq \frac{\gamma \|u_h\|_V}{\tilde{C}_{\text{up}} \sqrt{N}} =: \gamma_{\text{loc}}. \quad (3.5)$$

Indeed, in this case

$$\|u - u_h\|_V^2 \leq \tilde{C}_{\text{up}}^2 \tilde{\mathcal{E}}_{h,k}^2 = \tilde{C}_{\text{up}}^2 \sum_{K \in \mathcal{T}_h} \tilde{\eta}_{K,k}^2 \leq \gamma^2 \|u_h\|_V^2,$$

and thus (3.4) holds. Let us remark that \tilde{C}_{up}^2 is not known a priori in practice, and therefore we just decide to employ the condition (3.5) fixing $\tilde{C}_{\text{up}} = 1$ in our experiments. We therefore adapt the mesh by computing the local contributions and comparing them with γ_{loc} , thus locally refining the mesh if the condition (3.5) is not met, and coarsening if the local estimators are excessively small with respect to γ_{loc} .

In the following we employ for both the one and the two-dimensional cases the uniform distributions given in Example 2.2 for the random perturbations of the points. In light of Lemmas 5.3 and 5.4, we decide to correct the estimators by normalizing them with respect to the random perturbations. In particular, in the following, the estimators are normalized as $\tilde{\mathcal{E}}_{h,1} \leftarrow \tilde{\mathcal{E}}_{h,1}/\mathbb{E} \|\tilde{\alpha}_1\|$ and $\tilde{\mathcal{E}}_{h,2} \leftarrow \tilde{\mathcal{E}}_{h,2}/\mathbb{E} \|\tilde{\alpha}_1\|^2$.

3.1.1. One-dimensional case

We first consider $d = 1$ and the two-point boundary value problem (5.4) with κ and the exact solution u given by

$$\kappa(x) = 1 + x^3, \quad u(x) = x^3 \sin(a\pi x) \exp(-b(x - 0.5)^2),$$

where we fix $a = 15$ and $b = 50$, and where we choose the right-hand side f so that u is indeed the solution. As a goal, we set the tolerance $\gamma = 10^{-2}$ in (3.4) and stop the algorithm when condition (3.5) is met by all elements of the mesh. We consider the RM-FEM implemented with uniform random variables as in Example 2.2 and fix $p = 3$ in (2.4). Moreover, we consider $N_{\text{MC}} = 20$ realizations of the probabilistic mesh to approximate the error estimator as in (3.3). We then compute both the error estimators given in Definition 3.1 and employ $\tilde{\mathcal{E}}_{h,1}$ for adapting the mesh by refinement and coarsening, guided by the condition (3.5). The adaptivity algorithm is initialized with a mesh \mathcal{T}_h built on $N = 30$ elements of equal size and proceeds by refinement and coarsening. Results, given in Fig. 3, confirm the validity of our probabilistic error estimators. In particular, we remark that the local error estimators succeed in identifying the regions where the mesh has to be refined, thus getting a solution with an approximately equal distribution of the error over the domain. Both probabilistic estimators, moreover, succeed in bounding the global error until the tolerance is reached, with the estimator $\tilde{\mathcal{E}}_{h,2}$ which appears to be more efficient than $\tilde{\mathcal{E}}_{h,1}$.

3.1.2. Two-dimensional case

We now present two numerical experiments conducted in the two-dimensional case. In particular, for both experiments we only focus on the computation of $\tilde{\mathcal{E}}_{h,2}$ in Definition 3.1, since in view of Remark 3.3 this second estimator is computationally easier to implement than $\tilde{\mathcal{E}}_{h,1}$ for $d > 1$. To account for errors on the boundary elements, we decide for these experiments to perturb all points, including those on the boundaries, following Remark 2.8. In order for $\tilde{\mathcal{I}}_{u_h}$, and thus $\tilde{\mathcal{E}}_{h,2}$ to be well-defined, we reflect the perturbed boundary points symmetrically to the boundary ∂D in case they are outside the domain. For both experiments, we implement the RM-FEM with a uniform distribution for the random perturbations, as described in Example 2.2. Moreover, we fix $p = 3$ and compute the Monte Carlo approximation (3.3) on $N_{\text{MC}} = 500$ realizations of the random mesh. For the adaptivity algorithm, we start from a coarse mesh and apply regular local refinements if the condition (3.5) is not met by the local error estimator $\tilde{\eta}_{K,2}$. In the two-dimensional case we do not apply coarsening to the mesh.

We first consider $D = (0, 1)^2$, the conductivity $\kappa = 1$, so that (2.1) reduces to $-\Delta u = f$ with homogeneous Dirichlet boundary conditions. Moreover, we choose the right-hand side f such that

$$u_1(x, y) = -x(1-x)y(1-y) \arctan\left(\beta \left(\frac{x+y}{\sqrt{2}} - \frac{4}{5}\right)\right),$$

where $\beta > 0$. The solution has a steep transition around the line $\{y = 4\sqrt{2}/5 - x\}$, whose steepness is proportional to the parameter β . In Fig. 4, we show the exact solution for $\beta = 20$, which we fix for this experiment. We initialize the adaptivity procedure with a mesh with maximum element size $h = 1/5$ and proceed with adaptation until a tolerance $\gamma = 0.1$. In Fig. 5 we show the convergence of $\tilde{\mathcal{E}}_{h,2}$ with respect to the convergence of the true error, as well as the effectivity index for this experiment. We can see that the estimator indeed captures the error globally. In Fig. 6, we show the behavior of the local contributions $\tilde{\eta}_{K,2}$ with respect to the true error on each element, as well as the mesh adaptation. We can see that the error estimator succeeds in identifying the region where gradients are the steepest and proposes a mesh which appears adapted to this problem.

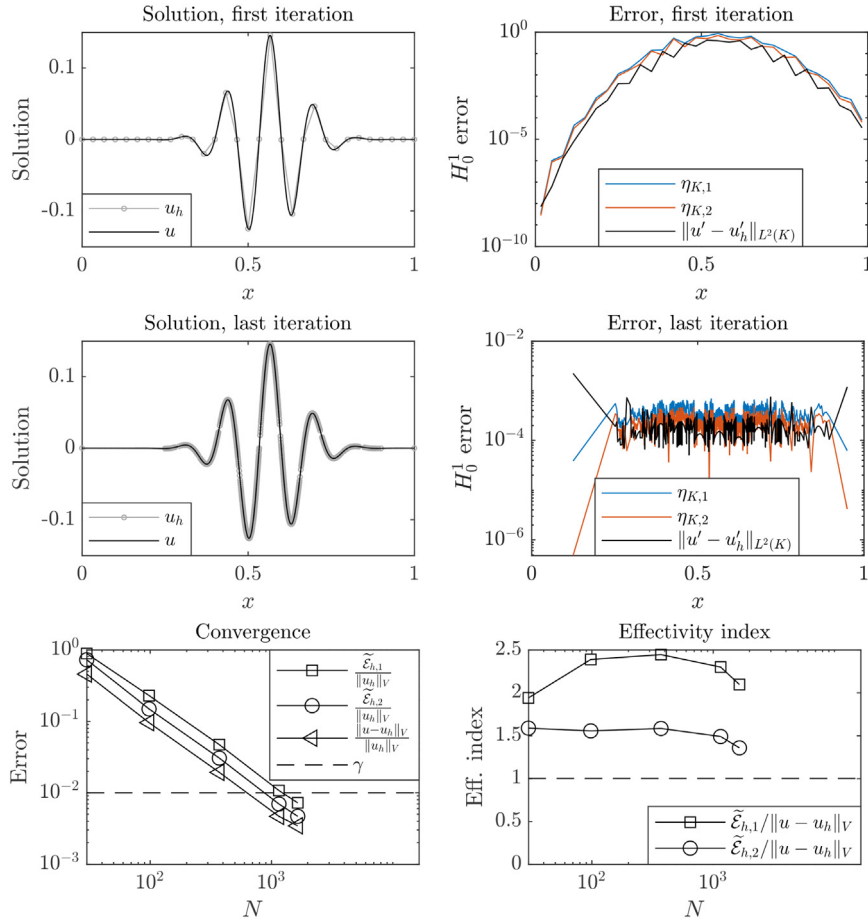


Fig. 3. Results for the one-dimensional experiment of Section 3.1. First and second rows: numerical and exact solutions u_h and u on the left, local contributions to the error estimators of Definition 3.1 compared with the true error $\|u - u_h\|_K$ on the right, at initialization and termination of the adaptivity procedure. Third row: on the left convergence of the global error $\|u - u_h\|_V$ and of the estimator \mathcal{E}_h until the tolerance γ , on the right the effectivity index.

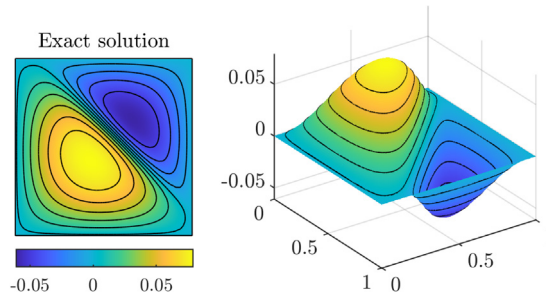


Fig. 4. Exact solution u_1 for the experiment of Section 3.1.2. Both the contour and the three-dimensional view highlight the steep gradient that features u_1 .

We then consider the L-shaped domain with the re-entrant corner on the origin, i.e. $D = (-1, 1)^2 \setminus (-1, 0)^2$. We set $\kappa = 1$, $f = 0$ and fix an inhomogeneous Dirichlet boundary conditions $u = g$ on ∂D , with g chosen such that the exact solution satisfies

$$u_2(r, \vartheta) = r^{2/3} \sin\left(\frac{2}{3}\left(\vartheta + \frac{\pi}{2}\right)\right),$$

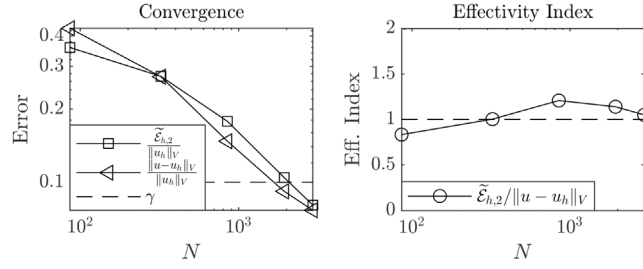


Fig. 5. Error convergence and effectivity index for the first experiment (function u_1) of Section 3.1.2.

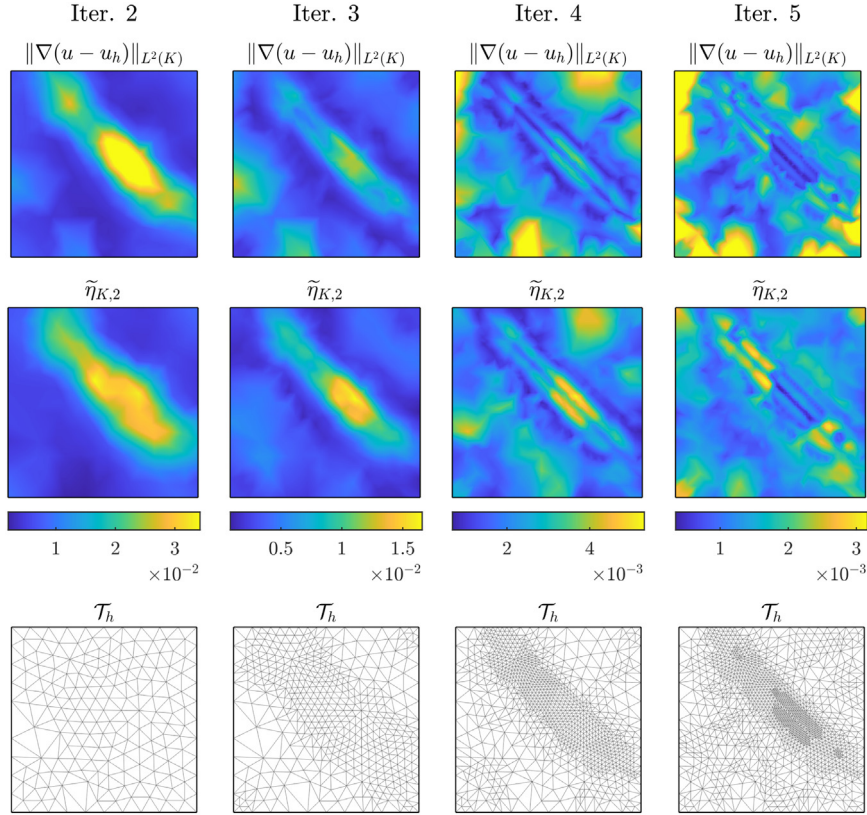


Fig. 6. Per row: True local error, local error estimator $\tilde{\eta}_{K,2}$ and mesh \mathcal{T}_h at each iteration of the adaptivity algorithm for the function u_1 of Section 3.1.2. The color bar is shared by the first and the second rows.

where $(r, \vartheta) \in \mathbb{R}^+ \times (0, 2\pi]$ are the polar coordinates in \mathbb{R}^2 . The exact solution of this problem is given in Fig. 7. Let us remark that the gradient of the exact solution is singular at the re-entrant corner, and we expect the mesh to be refined consequently at the singularity. For this experiment, we fix the tolerance $\gamma = 0.03$, and initialize the mesh to have a maximum element size of $h = 1/3$. Results, given in Figs. 8 and 9, show on the one hand that the estimator reproduces well the behavior of the global error during adaptation, and on the other hand that the mesh is progressively refined at the singularity as expected.

4. The RM-FEM for Bayesian inverse problems

Probabilistic numerical methods have been demonstrated to be particularly effective in the context of Bayesian inverse problems [9,11,13,16–18,35]. We consider the framework of [36, Section 3.4] and introduce the parameterized

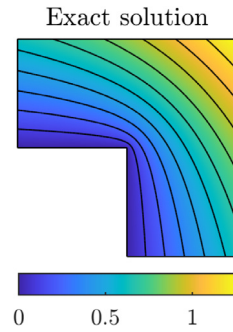


Fig. 7. Exact solution u_2 for the experiment of Section 3.1.2.

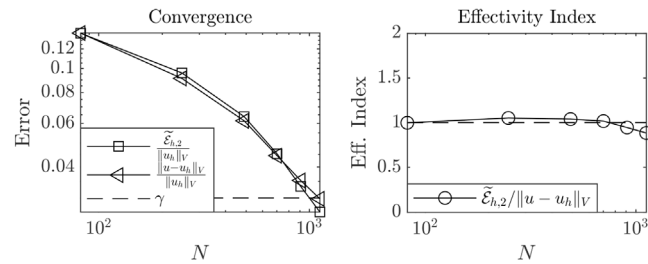


Fig. 8. Error convergence and effectivity index for the second experiment (function u_2) of Section 3.1.2.

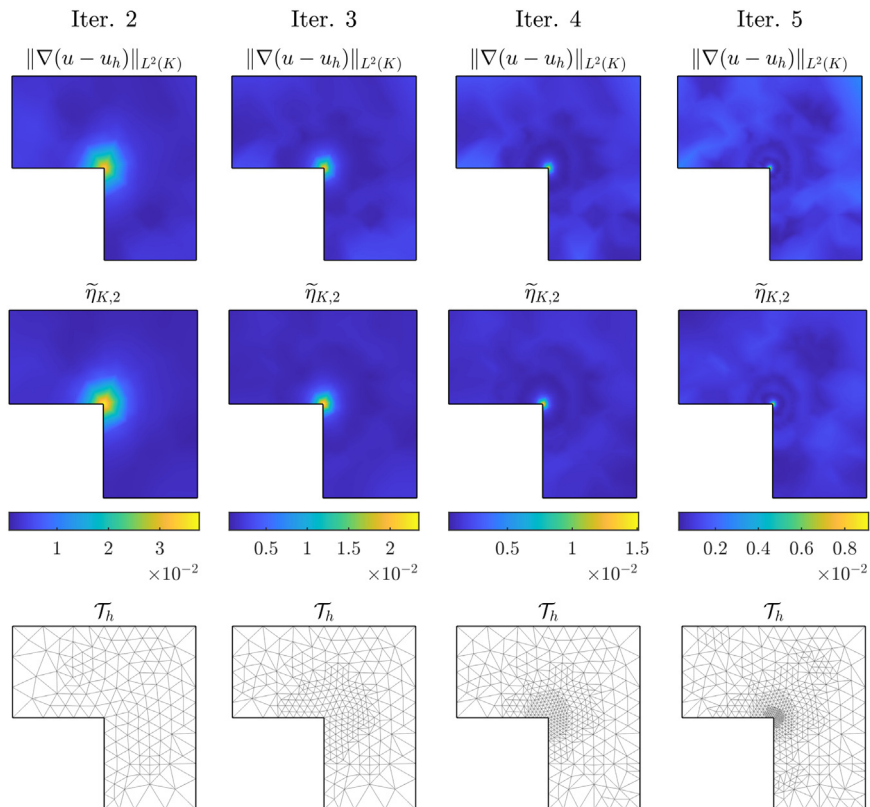


Fig. 9. Per row: True local error, local error estimator $\tilde{\eta}_{K,2}$ and mesh \mathcal{T}_h at each iteration of the adaptivity algorithm for the function u_2 of Section 3.1.2. The color bar is shared by the first and the second rows.

PDE

$$\begin{aligned} -\nabla \cdot (\exp(\vartheta) \nabla u) &= f, \quad \text{in } D, \\ u &= 0, \quad \text{on } \partial D, \end{aligned} \quad (4.1)$$

where D is an open bounded set of \mathbb{R}^d and $\vartheta: D \rightarrow \mathbb{R}$ is a scalar function. In particular, we let ϑ be such that problem (4.1) is well-posed, i.e., $\kappa = \exp(\vartheta) \in L^\infty(D)$ and $\kappa \geq \underline{\kappa} > 0$, and we denote by X the space of admissible values for ϑ . We introduce the solution operator $\mathcal{S}: X \rightarrow V$ such that $\mathcal{S}: \vartheta \mapsto u$, and the observation operator $\mathcal{O}: V \rightarrow \mathbb{R}^m$, which maps the solution of the PDE to point evaluations inside the domain on points $x^* = x_1^*, \dots, x_m^*$, i.e. $\mathcal{O}: u \mapsto y := (u(x_1^*), \dots, u(x_m^*))^\top$. Moreover, we denote by $\mathcal{G} = \mathcal{O} \circ \mathcal{S}$, $\mathcal{G}: X \rightarrow \mathbb{R}^m$, the so-called forward operator, which maps the parameter to the observations. We then have the Gaussian observation model

$$y = \mathcal{G}(\vartheta) + \beta, \quad \beta \sim \mathcal{N}(0, \Sigma),$$

where $\Sigma \in \mathbb{R}^{m \times m}$ is a non-singular covariance matrix on \mathbb{R}^m . Given an observation $y^* = (u(x_1^*), \dots, u(x_m^*))^\top$ associated to an unknown value $\vartheta^* \in X$ and corrupted by observational noise $\beta \in \mathbb{R}^m$ the inverse problem can then be stated as:

$$\text{find } \vartheta^* \in X \text{ given observations } y^* = \mathcal{G}(\vartheta^*) + \beta. \quad (4.2)$$

The randomness and the mismatch between the dimensionality of the unknown and of the observation make problem (4.2) ill-posed. Regularization can be achieved in the Bayesian framework (see e.g. [36,37]) by introducing probability measures on the unknown, which summarize prior knowledge and the information provided by data. We briefly introduce the Bayesian paradigm in the remainder of this section. First, we restrict ourselves to the space $\mathcal{H} = \mathcal{C}^0(\overline{D}) \cap V$, which is a valid subspace of admissible values for ϑ , i.e., $\mathcal{H} \subset X$. We then introduce a prior measure μ_0 on \mathcal{H} , encoding all knowledge on the unknown ϑ before observations are obtained. In particular, we consider a Gaussian prior measure $\mu_0 = \mathcal{N}(0, \Gamma_0)$ on \mathcal{H} , where Γ_0 is a positive semi-definite covariance operator on \mathcal{H} , and such that $\mu_0(\mathcal{H}) = 1$, so that any sample from μ_0 is in \mathcal{H} , a.s. We set the mean of the prior measure to zero without loss of generality. A broader class of prior measures could be employed, such as Besov or heavy-tailed measures [36,38], but we restrict ourselves to the Gaussian case for simplicity. Finally, we can obtain a measure μ^y on X encoding all the knowledge on ϑ given the prior and the observations. We call this measure the posterior, and formally compute with Bayes' formula its Radon–Nikodym derivative with respect to the prior as

$$\frac{d\mu^y}{d\mu_0}(\vartheta) = \frac{1}{Z^y} \exp(-\Phi(\vartheta; y)), \quad (4.3)$$

where Z^y is the normalization constant

$$Z^y = \int_{\mathcal{H}} \exp(-\Phi(\vartheta; y)) d\mu_0(\vartheta), \quad (4.4)$$

and where for any $y \in \mathbb{R}^m$ the potential $\Phi(\cdot; y): X \rightarrow \mathbb{R}$ is given due to the Gaussian assumption on the noise $\beta \sim \mathcal{N}(0, \Sigma)$ by

$$\Phi(\vartheta; y) = \frac{1}{2} \left\| \Sigma^{-1/2} (\mathcal{G}(\vartheta) - y) \right\|_2^2.$$

For economy of notation, in the following we drop the dependence of μ^y and Z^y on the data, and simply denote these quantities μ and Z . In order for (4.3) to be well-defined, the posterior measure needs to be absolutely continuous with respect to the prior. This is ensured under some conditions on the forward map \mathcal{G} , in particular, Lipschitz continuity and some controlled growth condition. Precisely, \mathcal{G} can be shown to satisfy [37, Assumption 2.7]. We then choose the covariance $\Gamma_0 = -\Delta^{-\alpha}$ with $\alpha > d/2$ and where we equip the Laplacian with homogeneous boundary conditions. Fractional powers of the Laplacian should be understood as per [37, Section 2]. With this choice, we have that $\mu_0(\mathcal{H}) = 1$. Together with the properties of \mathcal{G} we can conclude by [37, Theorem 3.4] that the posterior is indeed given by (4.3) and that the Bayesian inverse problem is well-posed, meaning that the dependence of the posterior on the observations is absolutely continuous.

In practice to evaluate the solution operator \mathcal{S} we recur to the FEM and consider for an $h > 0$ the forward operator $\mathcal{S}_h: X \rightarrow V_h$ which maps $\mathcal{S}_h: \vartheta \mapsto u_h$, where u_h is defined in (2.3). Moreover, we denote by $\mathcal{G}_h = \mathcal{O} \circ \mathcal{S}_h$ the resulting approximate forward operator. Maintaining the same observation model and the same prior on \mathcal{H} for

the parameter as above, we consider the approximate posterior μ_h on \mathcal{H} whose Radon–Nikodym derivative with respect to the prior is given by

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

where the potential Φ_h is given by

$$\Phi_h(\vartheta; y) = \frac{1}{2} \left\| \Sigma^{-1/2} (\mathcal{G}_h(\vartheta) - y) \right\|_2^2, \quad (4.5)$$

and where the normalization constant Z_h is defined equivalently to (4.4). A natural question arising from this setting is whether the approximate posterior μ_h converges to the true posterior μ in the limit $h \rightarrow 0$. This is indeed true, as it holds $d_H(\mu, \mu_h) \rightarrow 0$ for $h \rightarrow 0$, where the Hellinger distance $d_H(\cdot, \cdot)$ is defined as

$$d_H(\mu, \mu_h) := \sqrt{\frac{1}{2} \int_{\mathcal{H}} \left(\sqrt{\frac{d\mu}{d\mu_0}} - \sqrt{\frac{d\mu_h}{d\mu_0}} \right)^2 d\mu_0}.$$

For a proof of this result, see [37, Theorem 4.6], where the statement above is made more precise by the introduction of convergence rates.

It has been demonstrated heuristically that the approximate posterior measure μ_h can be overly confident on the parameter if $h > 0$ is a finite value and if the observation model is precise, i.e., when the covariance Σ of the observational noise is small [9,11,13,16–18,35]. In particular, this is accentuated when h is big relatively to the forward problem, or in other words when we employ a poor approximation of the forward map \mathcal{G} . It is therefore useful in applications to have a cheap surrogate which can be evaluated quickly, without renouncing to a complete uncertainty quantification of the solution to the inverse problem. Probabilistic numerical methods can be employed for this purpose. Let $\tilde{S}_h: X \rightarrow \tilde{V}_h$ be the solution operator mapping $\tilde{S}_h: \vartheta \mapsto \tilde{u}_h$, where \tilde{u}_h is the RM-FEM solution of Definition 2.4. In particular, in this context it is advisable to fix $p = 1$ in (2.4), so that the random deviations of the probabilistic numerical solution are of the same order of magnitude than the error itself by Theorem 2.9. Coherently to the definitions above, we define the random forward map $\tilde{\mathcal{G}}_h: X \rightarrow \mathbb{R}^m$ as $\tilde{\mathcal{G}}_h = \mathcal{O} \circ \tilde{S}_h$ and the approximate random posterior measure $\tilde{\mu}_h$ on X as

$$\frac{d\tilde{\mu}_h}{d\mu_0}(\vartheta) = \frac{\exp(-\tilde{\Phi}_h(\vartheta; y))}{\tilde{Z}_h},$$

where the potential $\tilde{\Phi}_h$ and the normalization constant \tilde{Z}_h are defined as above. Let us remark that the posterior $\tilde{\mu}_h$ is a random measure, as it depends on the random variable $\alpha: \Omega \rightarrow \mathbb{R}^{N_I}$ governing the random perturbations of the mesh. To be more precise, the posterior $\tilde{\mu}_h$ is a random variable $\tilde{\mu}_h: \Omega \rightarrow \mathcal{M}(\mathcal{H})$, where $\mathcal{M}(\mathcal{H})$ is the space of probability measures over the space \mathcal{H} . Employing the tools of [35] and due to Theorem 2.9, it is possible to prove convergence results for $\tilde{\mu}_h$ towards the true posterior μ for $h \rightarrow 0$ in the Hellinger metrics.

Remark 4.1. There exist other approaches to factor the effects of discretization into Bayesian inverse problems. In particular, numerical errors can be treated as modeling discrepancies. Under the assumption that the error is independent of the observational noise and of the inferred parameter ϑ , a viable alternative to probabilistic methods is employing the techniques introduced in [39,40], and further applied and analyzed e.g. in [41,42]. We argue that while assuming numerical errors to be independent of the observational noise is reasonable, their independence from the inferred parameter itself is not, at least for the inverse problem (4.2).

4.1. Implementation details

We now detail how one can solve the inverse problems above in practice. Given a smooth functional Ψ on X , we are interested in approximating the quantities $\mathbb{E}_{\mu_h}[\Psi(\cdot)]$ and $\mathbb{E}_{\tilde{\mu}_h}[\Psi(\cdot)]$, where \mathbb{E}_{μ_h} denotes expectation with respect to the measure μ_h (respectively $\tilde{\mu}_h$). To be more precise, in the probabilistic case we are interested to the quantity $\mathbb{E}[\mathbb{E}_{\tilde{\mu}_h}[\Psi(\cdot)]]$, where the outer expectation is taken with respect to the random perturbations intrinsic to the RM-FEM, and where for a sufficiently smooth functional Ψ the expectations can be exchanged by means of Fubini's theorem. Due to the high-dimensionality of the problem, Monte Carlo techniques are a natural choice. Let

N_{MC} be a positive integer and let us assume that we have samples $\{\vartheta^{(i)}\}_{i=0}^{N_{\text{MC}}} \sim \mu_h$, not necessarily independent. Then, defining

$$\widehat{\Psi}_{\mu_h} := \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \Psi(\vartheta^{(i)}). \quad (4.6)$$

we have $\widehat{\Psi}_{\mu_h} \approx \mathbb{E}_{\mu_h}[\Psi(\cdot)]$. For the probabilistic case, let $\widetilde{N}_{\text{MC}}^{\text{out}}$ and $\widetilde{N}_{\text{MC}}^{\text{in}}$ be positive integers, let $\{\widetilde{\mu}_h^{(j)}\}_{j=1}^{\widetilde{N}_{\text{MC}}^{\text{out}}}$ be a sequence of realizations of the measure $\widetilde{\mu}_h$, obtained with a corresponding series of random perturbations of the mesh, and let $\{\widetilde{\vartheta}^{(j,i)}\}_{i=1}^{\widetilde{N}_{\text{MC}}^{\text{in}}} \sim \widetilde{\mu}_h^{(j)}$. Then, we define

$$\widehat{\Psi}_{\widetilde{\mu}_h} := \frac{1}{\widetilde{N}_{\text{MC}}^{\text{out}} \widetilde{N}_{\text{MC}}^{\text{in}}} \sum_{i=1}^{\widetilde{N}_{\text{MC}}^{\text{out}}} \sum_{j=1}^{\widetilde{N}_{\text{MC}}^{\text{in}}} \Psi(\vartheta^{(j,i)}), \quad (4.7)$$

and we have $\widehat{\Psi}_{\widetilde{\mu}_h} \approx \mathbb{E}[\mathbb{E}_{\widetilde{\mu}_h}[\Psi(\cdot)]]$. Still, the random variable ϑ is infinite-dimensional, and we need to define a finite-dimensional approximation in order to obtain a practical procedure to generate the above samples and thus solve the inverse problem. We recur to the Karhunen–Loève expansion (KL). Denoting by $\{(\lambda_i, \varphi_i)\}_{i \geq 1}$ the ordered eigenvalues/eigenfunctions of the prior covariance Γ_0 , a function $\vartheta \sim \mu_0$ is given by the convergent sum

$$\vartheta = \sum_{i \geq 1} \sqrt{\lambda_i} \varphi_i \xi_i,$$

where $\{\xi_i\}_{i \geq 1} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. We then let N_{KL} be a positive integer and truncate the sum above as

$$\vartheta = \sum_{i=1}^{N_{\text{KL}}} \sqrt{\lambda_i} \varphi_i \xi_i, \quad (4.8)$$

thus obtaining a function $\vartheta \in \mathcal{H}$ which is approximately sampled from μ_0 . Due to the super-quadratic decay of the eigenvalues of $\Gamma_0 = -\Delta^{-\alpha}$ for $\alpha > d/2$, disregarding the tail of the sum causes a negligible error in case N_{KL} is chosen appropriately large. Our inversion problem is therefore shifted to computing the posterior distribution on a finite-dimensional parameter, comprising the coefficients of the expansion (4.8). We define the mapping $\mathcal{K}: \mathbb{R}^{N_{\text{KL}}} \rightarrow \mathcal{H}$, $\mathcal{K}: \xi \mapsto \vartheta$ by (4.8). The prior measure on ξ is $\mu_0 = \mathcal{N}(0, I)$, with I being the identity matrix of dimension $N_{\text{KL}} \times N_{\text{KL}}$, and we denote by π_0 the density of μ_0 with respect to the Lebesgue measure. The density π_h of the posterior on the coefficients ξ given the observations is then

$$\pi_h(\xi) = \frac{1}{Z_h} \pi_0(\xi) \exp(-\Phi_h(\mathcal{K}(\xi); y)),$$

where Φ_h is defined in (4.5) and Z_h is the normalization constant

$$Z_h = \int_{\mathbb{R}^{N_{\text{KL}}}} \exp(-\Phi_h(\mathcal{K}(\xi); y)) \pi_0(\xi) d\xi.$$

The same procedure can be applied seamlessly to the probabilistic case, thus obtaining a random posterior density $\widetilde{\pi}_h$ over $\mathbb{R}^{N_{\text{KL}}}$ for the coefficient ξ .

The last detail missing is how to produce samples in order to obtain the approximations (4.6) and (4.7). Being the normalizations constant unknown, we employ Markov chain Monte Carlo techniques (MCMC) (see e.g. [43, Chapter 3] or [44, Chapter 6]), which proceed by generating an ergodic Markov chain whose invariant density is the desired posterior. Successive states of the aforementioned Markov chain then serve as samples from the posterior density. We choose to employ the Metropolis–Hastings (MH) algorithm, which we here briefly detail. The Markov chain is built employing a symmetric proposal $q: \mathbb{R}^{N_{\text{KL}}} \times \mathbb{R}^{N_{\text{KL}}} \rightarrow \mathbb{R}$ satisfying $q(x, y) = q(y, x)$ for all $x, y \in \mathbb{R}^{N_{\text{KL}}}$ and such that for any fixed $x \in \mathbb{R}^{N_{\text{KL}}}$ the function $q(x, \cdot)$ is a probability density, and with an acceptance–rejection strategy. In particular, given an initial guess $\xi^{(1)}$, the algorithm proceeds for $i = 2, \dots, N_{\text{MC}}$ as

- (i) Sample $\bar{\xi}^{(i)} \sim q(\xi^{(i-1)}, \cdot)$;

(ii) Set $\xi^{(i)} = \bar{\xi}^{(i)}$ with probability α , and $\xi^{(i)} = \xi^{(i-1)}$ with probability $1 - \alpha$, where

$$\alpha = \min \left\{ \frac{\pi_h(\bar{\xi}^{(i)})}{\pi_h(\xi^{(i-1)})}, 1 \right\}.$$

Let us remark that the normalization constant Z_h does not need to be known to run the algorithm, since we only compute ratios of posterior densities. Moreover, we note that the proposal distribution is the only tunable element of the MH algorithm. The easiest choice, at least for implementation, would be to fix $q(x, \cdot) = \mathcal{N}(x, \sigma^2 I)$ for some user-prescribed variance σ^2 . Unfortunately, the quality of the resulting Markov chain is not robust with respect to σ . In particular, if σ is too small, the probability to accept is too large and the Markov chain fails to effectively explore the posterior. At the other end of the spectrum, if σ is too large the probability of accepting a new sample reduces drastically, and the Markov chain presents a sticky behavior. We therefore decide to employ the robust adaptive Metropolis algorithm (RAM) (see [45] for details), in which the proposal is $q(x, \cdot) = \mathcal{N}(x, \Sigma_q)$, where the covariance Σ_q is adapted on the fly to obtain a user-specified final acceptance ratio, i.e. the ratio between the accepted and the total number samples, which should be roughly 25% (see e.g. [45]). Another viable option for the implementation of MCMC could have been the preconditioned Crank–Nicolson MCMC (pCN-MCMC) of [46,47], which is tailored for high-dimensional inverse problems.

For the probabilistic case, we perform a run of the MH, implemented with RAM proposal, with $\tilde{N}_{MC}^{\text{in}}$ iterations for each one of the $\tilde{N}_{MC}^{\text{out}}$ realizations of the random mesh, thus obtaining the approximation (4.7).

Remark 4.2. The total number of samples is given by in the probabilistic case by $\tilde{N}_{MC}^{\text{in}} \cdot \tilde{N}_{MC}^{\text{out}}$. One could argue that, for a fair comparison between the probabilistic and the deterministic case in terms of computational cost, one should choose $N_{MC} \approx \tilde{N}_{MC}^{\text{in}} \cdot \tilde{N}_{MC}^{\text{out}}$. In fact, since the “outer” Monte Carlo simulation can be performed in parallel, the correct scaling is $N_{MC} \approx \tilde{N}_{MC}^{\text{in}}$. Moreover, due to Remark 3.7, the number of random meshes does not need to be chosen excessively large.

4.2. Numerical experiments

In this section we present numerical experiments highlighting the beneficial effects of adopting the probabilistic framework of RM-FEM in the context of Bayesian inverse problems.

4.2.1. One-dimensional case

We first consider $D = (0, 1)$ and solve the inverse problem presented above for two different true diffusion fields κ^* . In both cases, we consider the prior on \mathcal{H} to be given by $\mathcal{N}(0, \Gamma_0)$, with $\Gamma_0^{-1} = -d^2/dx^2$ with homogeneous boundary conditions, so that the Bayesian inverse problem is well-posed. First, we consider $\kappa_1^* = \exp(\vartheta_1^*)$, where the log-conductivity $\vartheta_1^* \in \mathcal{H}$ is given by

$$\vartheta_1^*(x) = \sum_{j=1}^4 \xi_j \sqrt{\lambda_j} \varphi_j(x),$$

with $\xi_1 = \xi_2 = 1$, $\xi_3 = \xi_4 = 1/4$, and where $\{(\lambda_i, \varphi_i)\}_{i=1}^4$ are the first four ordered eigenpairs of Γ_0 . Second, we consider $\vartheta_2^* \in X \cap \mathcal{H}^C$, so that the true conductivity does not belong to the domain in which we solve the inverse problem, but it is still admissible for (4.1) to be well-posed. In particular, we consider the discontinuous conductivity

$$\kappa_2^*(x) = \begin{cases} 1.5, & 0.2 < x < 0.6, \\ 0.5, & 0.6 < x < 0.8, \\ 1, & \text{otherwise,} \end{cases}$$

and infer $\vartheta_2^* = \log(\kappa_2^*)$. For both problems, we choose the right-hand side in (4.1) as $f(x) = \sin(2\pi x)$. Synthetic observations are obtained as point evaluations of a reference solution on points $x_i^* = i/10$, for $i = 1, \dots, 9$, corrupted by Gaussian noise $\mathcal{N}(0, 10^{-8}I)$. The forward map is approximated with FEM and RM-FEM. The mesh \mathcal{T}_h for the FEM is equally spaced, and we vary the number of elements $N = \{10, 20, 40\}$. For the RM-FEM, we consider $p = 1$ in (2.4) as per Theorem 2.9 and implement the random perturbations with a uniform distribution as in Example 2.2.

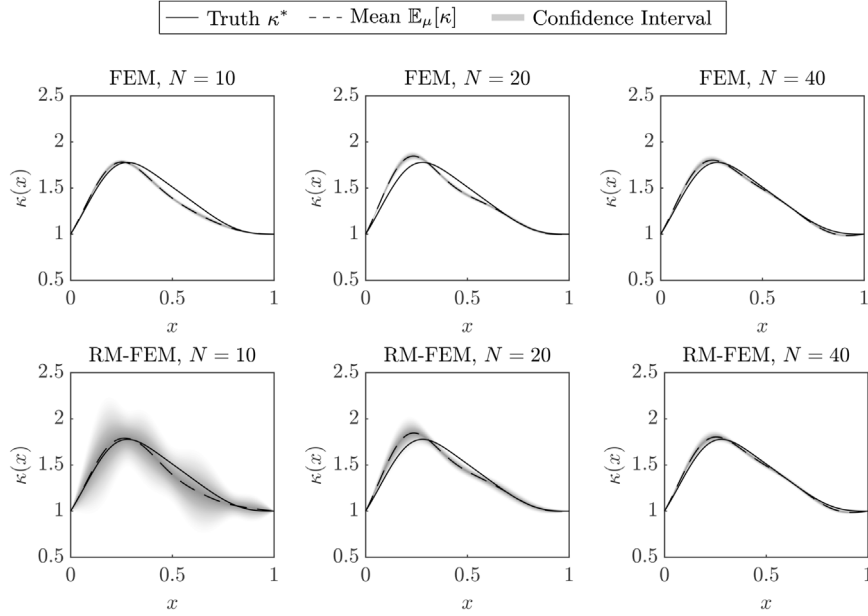


Fig. 10. Numerical results for κ_1^* in Section 4.2.1. In all plots, the solid line represents the true conductivity, the dashed line is the posterior mean, and the shaded gray area is a confidence interval. In the first row, results are obtained by approximating the forward map with FEM, and in the second with the RM-FEM.

We sample with the MH algorithm from the posterior distributions μ_h and $\tilde{\mu}_h$, with $N_{MC} = 2 \cdot 10^5$ for μ_h and with $\tilde{N}_{MC}^{\text{out}} = 50$ and $\tilde{N}_{MC}^{\text{in}} = 2 \cdot 10^5$ for $\tilde{\mu}_h$. Knowing for the first conductivity κ_1^* that the true conductivity is fully determined by four coefficients, we fix the truncation index $N_{KL} = 4$ in the Karhunen–Loève expansion (4.8). For the second conductivity κ_2^* , we fix $N_{KL} = 9$. We then approximate the mean and pointwise standard deviation with (4.6) and (4.7) for the deterministic and probabilistic posteriors, respectively. Moreover, we arbitrarily fix a pointwise confidence interval at twice the standard deviation away from the mean. Numerical results are given in Figs. 10 and 11. Results highlight that for a coarse approximation, specifically for $N = 10$, the posterior distribution μ_h is overly confident on the result. Indeed, the posterior mean fails to capture precisely the true conductivity in both the continuous and discontinuous cases, and the confidence interval is extremely sharply concentrated around the mean. Conversely, the distribution $\tilde{\mu}_h$ based on the probabilistic forward model accounts better for the uncertainty due to numerical discretization. Increasing the number of elements N , the mean computed under μ_h and $\tilde{\mu}_h$ tends to approximate better the true conductivity field. In particular, for $N = 40$ the posteriors μ_h and $\tilde{\mu}_h$ are already practically undistinguishable and are close to the true field. Moreover, let us remark that while the width of the confidence interval seems independent of N for μ_h , it shrinks coherently to the discretization for $\tilde{\mu}_h$. Finally, we note that for κ_2^* even for larger values of N the posterior $\tilde{\mu}_h$ seems to capture with its uncertainty local errors in the solution of the inverse problem. Indeed, the posterior mean is particularly off the true field on the left side of the domain, where the confidence interval is wider with respect to areas where the solution is more accurate.

4.2.2. Two-dimensional case

We consider now a two dimensional example on the domain $D = (0, 1)^2$. We fix a Gaussian prior μ_0 on \mathcal{H} for the log-conductivity ϑ chosen as $\mu_0 = \mathcal{N}(0, \Gamma_0)$, where $\Gamma_0 = -\Delta^{-1.3}$ with homogeneous boundary conditions, so that the inverse problem is well-posed. We fix $N_{KL} = 6$ and let the true conductivity $\kappa^* = \exp(\vartheta^*)$ in (4.1) be given by

$$\vartheta^* = \sum_{i=1}^6 \sqrt{\lambda_i} \varphi_i \xi_i^*,$$

where $\{(\lambda_i, \varphi_i)\}_{i=1}^6$ are the first six ordered eigenpairs of Γ_0 , and where $\xi_i^* = (-1)^{i+1} \cdot 10$ for $i = 1, 2, \dots, 6$. Let us remark that $\vartheta^* \in \mathcal{H}$. The right-hand side in (4.1) is chosen as $f(x, y) = 8\pi^2 \sin(2\pi x) \sin(2\pi y)$. Synthetic

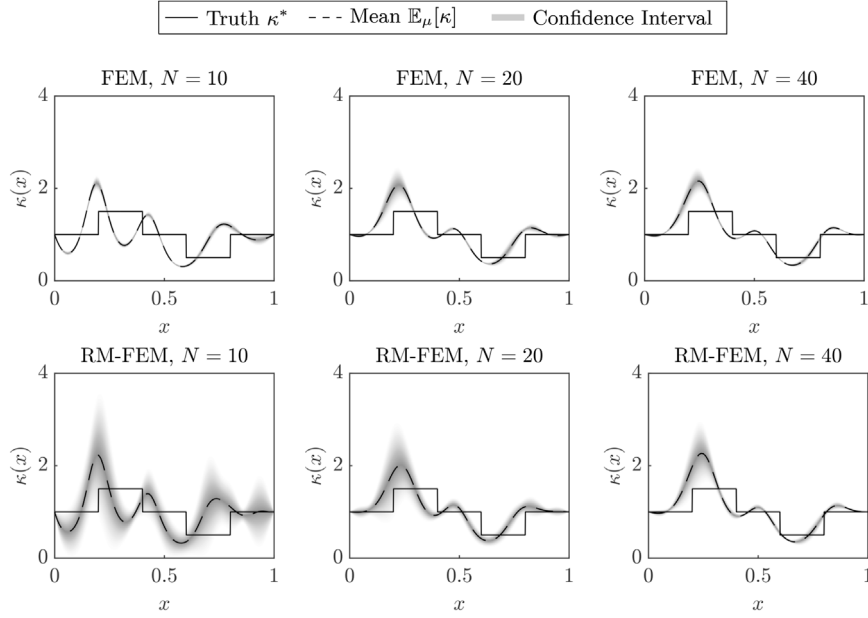


Fig. 11. Numerical results for κ_2^* in Section 4.2.1 In all plots, the solid line represents the true conductivity, the dashed line is the posterior mean, and the shaded gray area is a confidence interval. In the first row, results are obtained by approximating the forward map with FEM, and in the second with the RM-FEM.

observations are obtained by evaluating a reference solution on $m = 50$ random locations sampled from $\mathcal{U}(D)$ and then corrupted by an observational noise distributed as $\mathcal{N}(0, 10^{-6}I)$. We then approximate the forward map in the inverse problem with the FEM and the RM-FEM. We choose a structured mesh \mathcal{T}_h as the one in Example 2.2 (or the second row of Fig. 1). In particular, in this case we let h denote the constant length of the short side of the triangular elements, i.e., the inverse of the number of subdivisions of each side of D . In particular, we consider $h = 0.1 \cdot 2^{-i}$, $i = 0, 1, \dots, 3$. The RM-FEM is implemented with $p = 1$ in (2.4) as per Theorem 2.9, and with a uniform choice for the random perturbations as the one described in Example 2.2.

Employing the notation introduced in Section 4.1, we then sample from the posterior distributions μ_h and $\tilde{\mu}_h$ employing the RAM method, considering only the first $N_{\text{KL}} = 6$ coefficients in the KL expansion (4.8). In particular, we consider $N_{\text{MC}} = 10^5$ samples for the deterministic case, and for the probabilistic case we generate $\tilde{N}_{\text{MC}}^{\text{in}} = 10^5$ samples for $\tilde{N}_{\text{MC}}^{\text{out}} = 24$ parallel chains, each corresponding to a realization of the random mesh in the RM-FEM. We then compute for each value of h the mean and standard deviation of the field computed under μ_h (resp. $\tilde{\mu}_h$) and denote their Monte Carlo approximations as $\hat{\kappa}_{\mu_h}$ and $\hat{\sigma}_{\mu_h}^{\kappa}$ (resp. $\hat{\kappa}_{\tilde{\mu}_h}$, $\hat{\sigma}_{\tilde{\mu}_h}^{\kappa}$). Moreover, we consider the statistics of the 6-dimensional coefficient σ of the KL expansion, and denote by $\hat{\xi}_{\mu_h}$ and $\hat{\sigma}_{\mu_h}^{\xi}$ the Monte Carlo approximation of mean and standard deviation computed under μ_h (resp. $\hat{\xi}_{\tilde{\mu}_h}$, $\hat{\sigma}_{\tilde{\mu}_h}^{\xi}$). We show in Fig. 12 the posterior mean $\hat{\kappa}_{\mu_h}$ for three values of h , compared to the truth κ^* , and remark that the mean approximation is sensibly better for smaller values of h . The mean value under the probabilistic posterior $\tilde{\mu}_h$ is not shown, as it is essentially equal to the deterministic case. The beneficial effect of employing the RM-FEM-based posterior distribution $\tilde{\mu}_h$, with respect to the FEM-based posterior μ_h , consists of the approximate equalities

$$\left\| \hat{\sigma}_{\mu_h}^{\kappa} \right\|_{L^2(D)} = \mathcal{O} \left(\left\| \kappa^* - \hat{\kappa}_{\tilde{\mu}_h} \right\|_{L^2(D)} \right), \quad \left\| \hat{\sigma}_{\tilde{\mu}_h}^{\xi} \right\| = \mathcal{O} \left(\left\| \xi^* - \hat{\xi}_{\tilde{\mu}_h} \right\| \right),$$

which indicate that the error on the conductivity field, or on the coefficients of its KL expansion, are well represented by the uncertainty in the posterior distribution. This is shown in Fig. 12, where we notice that under μ_h the standard deviation is practically independent of h and small with respect to the error on the solution of the inverse problem. Conversely, under $\tilde{\mu}_h$ we have that the posterior standard deviation converges accordingly to the error, both for the L^2 -norm of the error on the mean and for the coefficients of the KL expansion.

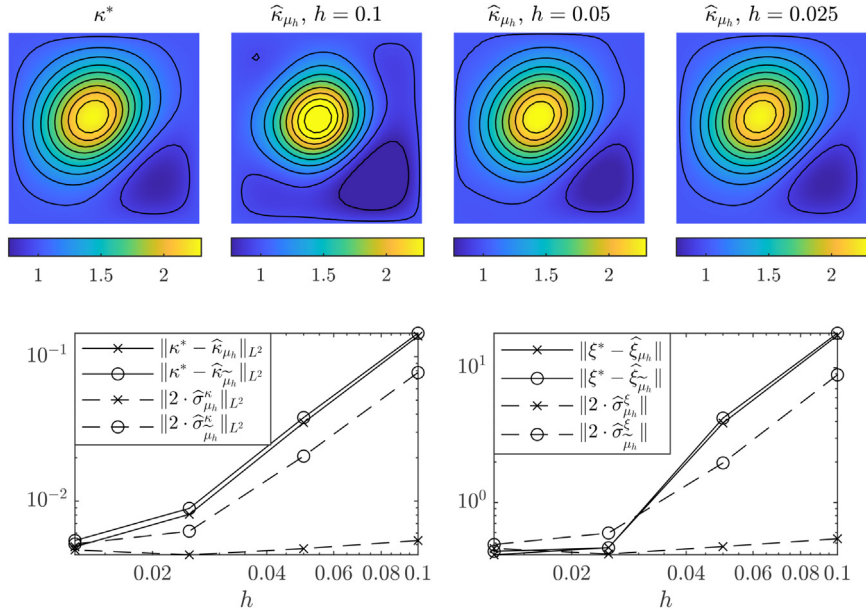


Fig. 12. Numerical results for Section 4.2.2. First row: True conductivity field κ^* and posterior mean field $\hat{\kappa}_{\mu_h}$ estimated with MCMC and different values of h . Second row: Mean error vs standard deviation under μ_h and $\tilde{\mu}_h$. On the left, $L^2(D)$ error on the mean field vs $L^2(D)$ norm of the punctual standard deviation under μ_h and $\tilde{\mu}_h$. On the right, error with respect to exact KL coefficients, and standard deviations under μ_h and $\tilde{\mu}_h$.

5. Error analysis for the RM-FEM

In this section, we present our a priori and our a posteriori error analysis for the RM-FEM. Let us remark that while the a priori error analysis is carried on for a general space dimension d and the adaptive algorithm has been shown to be efficient in higher dimensions (see Section 3.1), we present a rigorous a posteriori error analysis only in case $d = 1$. Conversely, in the a priori analysis we fix the coefficient $p = 1$ in (2.4), whereas in the a posteriori analysis we consider general perturbations, i.e., general coefficients $p \geq 1$ in the same equality.

5.1. A priori error estimates

We first prove the a priori error estimate given in Theorem 2.9. The convergence properties of the FEM for the elliptic problem (2.1) are well-established. In particular, without any additional assumptions on the exact solution, i.e., when $u \in V$, it holds $\|u - u_h\|_V \rightarrow 0$ for $h \rightarrow 0$. Under the more restrictive assumption $u \in H^2(D) \cap V$, we have a linear convergence rate, i.e.

$$\|u - u_h\|_V \leq Ch |u|_{H^2(D)}, \quad (5.1)$$

for a constant $C > 0$, which is independent of h and u [48–50]. It is desirable that the RM-FEM is endowed with the same property. Moreover, we wish the error due to randomization to be balanced with the error due to the FEM discretization, which is shown in the proof of Theorem 2.9.

Proof of Theorem 2.9. Since (5.1) holds independently of the mesh, we have

$$\|u - \tilde{u}_h\|_V \leq \tilde{C}h |u|_{H^2(D)}, \quad \text{a.s.},$$

for a constant \tilde{C} independent of h and u and of the coefficient p in (2.4). Hence, by the triangle inequality we have for $p = 1$

$$\|u_h - \tilde{u}_h\|_V \leq \|u - u_h\|_V + \|u - \tilde{u}_h\|_V \leq (C + \tilde{C})h |u|_{H^2(D)}, \quad \text{a.s.},$$

i.e., we have $\mathcal{O}(\|u_h - \tilde{u}_h\|_V) = \mathcal{O}(\|u - u_h\|_V)$, which shows the desired result. \square

Let us remark that we have shown above that the probabilistic solution converges with the same rate with respect to h in case $p = 1$, but we have not considered the case $p > 1$, for which the probabilistic term may be of higher order. Indeed, a preliminary theoretical and numerical investigation leads us to conjecture that

$$(\mathbb{E} \|u_h - \tilde{u}_h\|_V^2)^{1/2} \leq Ch^{(p+1)/2}, \quad (5.2)$$

so that, at least in the mean-square sense, the error due to randomization should converge faster than the error due to discretization if $p > 1$.

5.2. A posteriori error analysis in the one-dimensional case

In this section we prove our main result for the a posteriori error estimator of the RM-FEM given in [Definition 3.1](#), namely [Theorem 3.6](#). Our goal is to prove in the one-dimensional case that the probabilistic a posteriori error estimators are reliable and efficient, i.e., that there exist positive constants \tilde{C}_{low} and \tilde{C}_{up} independent of h and u such that

$$\tilde{C}_{\text{low}} \tilde{\mathcal{E}}_{h,k} \leq \|u - u_h\|_V \leq \tilde{C}_{\text{low}} \tilde{\mathcal{E}}_{h,k}. \quad (5.3)$$

for $k = \{1, 2\}$. Consider the elliptic two-point boundary value problem

$$\begin{aligned} -(\kappa u')' &= f, \quad \text{in } D, \\ u(0) &= u(1) = 0, \end{aligned} \quad (5.4)$$

where $\kappa \in L^\infty(D)$ satisfies $\kappa(x) \geq \underline{\kappa}$ almost everywhere in D , and where we assume $f \in L^1(D)$. We recall that the notation for one-dimensional problems has been introduced and discussed in [Example 2.2](#) and at the end of [Section 3](#). Additionally, we introduce here for a function w which is piecewise constant on each $K_i \in \mathcal{T}_h$ the jump operator

$$[w]_{x_i} := w|_{K_i} - w|_{K_{i+1}}, \quad i = 1, \dots, N-1, \quad [w]_{x_0} = [w]_{x_N} = 0.$$

Our strategy for proving that the error estimator introduced in [Definition 3.1](#) satisfies (5.3) relies on showing it is equivalent to known valid estimators. In particular, we consider the following estimator, defined in [\[27, Definition 6.3\]](#).

Definition 5.1. Let κ be the diffusion coefficient of (5.4) satisfy $\kappa \in C^0(D)$ and $\kappa \geq \underline{\kappa} > 0$. We define the error estimator

$$\mathcal{E}_h^2 := \sum_{j=1}^N \eta_j^2, \quad \eta_j := \|\kappa^{-1} \ell_j\|_{L^2(K_j)},$$

with $\ell_j: K_j \rightarrow \mathbb{R}$ the linear function defined by $\ell_j(x_{j-1}) = \tau_{j,1}$, $\ell_j(x_j) = -\tau_{j,0}$ where

$$\tau_{j,k} = \frac{h_j}{h_{j-k+1} + h_{j-k}} [u_h']_{x_{j-k}} \kappa(x_{j-k}).$$

Clearly, the quantity \mathcal{E}_h is computable up to quadrature error due to the approximation of the local estimators η_j . Let us finally introduce more precisely the higher-order quantity Λ appearing in (3.2), i.e.,

$$\Lambda^2 := h^\zeta \sum_{j=1}^N \int_{K_j} (f(x) + \ell_j'(x))^2 dx, \quad (5.5)$$

where $\zeta \in (0, 1)$ is arbitrary and ℓ_j are the linear functions employed in [Definition 5.1](#). We can now state the main result concerning the estimator \mathcal{E}_h , which summarizes [\[27, Theorems 8.1 and 8.2\]](#).

Theorem 5.2. Let \mathcal{E}_h and Λ be defined in [Definition 5.1](#) and (5.5), respectively. Then, it holds up to higher order terms in h

$$\|u - u_h\|_V \leq C (\mathcal{E}_h^2 + \Lambda^2)^{1/2},$$

for a constant C independent of h and of the solution u . If moreover the family of meshes \mathcal{T}_h is λ -quasi-uniform and if $\kappa \in C^2(D)$ and $f \in C^1(D)$ then, up to higher order terms, it holds

$$C_{\text{low}} \mathcal{E}_h \leq \|u - u_h\|_V \leq C_{\text{up}} \mathcal{E}_h,$$

for constants $C_{\text{low}}, C_{\text{up}}$ independent of h and u .

We recall that in the one-dimensional case the probabilistic error estimators for the RM-FEM are given by

$$\begin{aligned} \tilde{\mathcal{E}}_{h,1} &:= \left(\sum_{i=1}^N \tilde{\eta}_{K_i,1}^2 \right)^{1/2}, \quad \text{with} \quad \tilde{\eta}_{K_i,1}^2 = h_i^{-(p-1)} \mathbb{E} \left[\|u'_h - (\tilde{\mathcal{I}}u_h)' \|_{L^2(\tilde{K}_i)}^2 \right], \\ \tilde{\mathcal{E}}_{h,2} &:= \left(\sum_{i=1}^N \tilde{\eta}_{K_i,2}^2 \right)^{1/2}, \quad \text{with} \quad \tilde{\eta}_{K_i,2}^2 = h_i^{-(2p-3)} \mathbb{E} \left[\|u'_h|_K - (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}} \|^2 \right]. \end{aligned}$$

Our strategy to prove [Theorem 3.6](#) relies on showing that the deterministic estimator \mathcal{E}_h of [Definition 5.1](#), as well as its probabilistic counterparts $\tilde{\mathcal{E}}_{h,1}$ and $\tilde{\mathcal{E}}_{h,2}$ above are all equivalent to the quantity

$$\mathcal{J}(u_h) := \sum_{i=1}^{N-1} \tilde{h}_i \|u'_h\|^2,$$

i.e., the sum of all squared jumps of the derivatives on the internal nodes. We first prove the equivalence for $\tilde{\mathcal{E}}_{h,1}$.

Lemma 5.3. *Let [Assumption 2.1](#) hold. Then, if the mesh is λ -quasi-uniform it holds*

$$\left(\frac{\mathbb{E} |\tilde{\alpha}_1| (1 + \lambda^{-(p-1)})}{2} - 2h^{p-1} \mathbb{E} |\tilde{\alpha}_1|^2 \right) \mathcal{J}^2(u_h) \leq \tilde{\mathcal{E}}_{h,1}^2 \leq \frac{\mathbb{E} |\tilde{\alpha}_1| (1 + \lambda^{p-1})}{2} \mathcal{J}^2(u_h),$$

where $\tilde{\mathcal{E}}_{h,1}$ is given in [Definition 3.1](#).

Proof. Let $\tilde{K}_i, i = 1, \dots, N$, be a generic element of the perturbed mesh and let us compute the derivative of the interpolant on \tilde{K}_i , which is given by

$$(\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} = \frac{u_h(\tilde{x}_i) - u_h(\tilde{x}_{i-1})}{\tilde{x}_i - \tilde{x}_{i-1}},$$

where an exact Taylor expansion allows to compute

$$u_h(\tilde{x}_{i-1}) = u_h(x_{i-1}) + h^p \alpha_{i-1} u'_h(\tilde{x}_{i-1}).$$

Hence, it holds

$$(\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} = \frac{x_i - x_{i-1}}{\tilde{x}_i - \tilde{x}_{i-1}} u'_h|_{K_i} + h^p \frac{\alpha_i u'_h(\tilde{x}_i) - \alpha_{i-1} u'_h(\tilde{x}_{i-1})}{\tilde{x}_i - \tilde{x}_{i-1}},$$

which we can rewrite rearranging terms as

$$(\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} - u'_h|_{K_i} = h^p \frac{\alpha_i (u'_h(\tilde{x}_i) - u'_h|_{K_i}) + \alpha_{i-1} (u'_h|_{K_i} - u'_h(\tilde{x}_{i-1}))}{\tilde{x}_i - \tilde{x}_{i-1}}. \quad (5.6)$$

It is clear then that the expression above depends on the signs of the variables α_{i-1} and α_i . For simplicity of notation, we therefore introduce the events $A_{i,j}^{(s_i, s_j)} \in \mathcal{A}$, where $s_i, s_j \in \{+, -\}$, defined as

$$A_{i,j}^{(s_i, s_j)} := \{\omega \in \Omega : \alpha_i(\omega) \in \mathbb{R}^{s_i}, \alpha_j(\omega) \in \mathbb{R}^{s_j}\}.$$

We now define $e_h := u_h - \tilde{\mathcal{I}}u_h$ and write for any $i = 1, \dots, N$

$$\mathbb{E} \|e'_h\|_{L^2(\tilde{K}_i)}^2 = \mathbb{E} (I_{i-1,i} + I_i + I_{i+1,i}),$$

with

$$I_{i,j} := \int_{K_i \cap \tilde{K}_j} (e'_h)^2 dx,$$

and where we write $I_i := I_{i,i}$ and adopt the convention $I_{0,1} = I_{N+1,N} = 0$. In what follows we study $\mathbb{E}I_{i,j}$. We first consider I_i , which we express by the law of total expectation as

$$\mathbb{E}I_i = \sum_{s_{j-1}, s_j \in \{-, +\}} \mathbb{E} \left[I_i \mid A_{i-1,i}^{(s_{j-1}, s_j)} \right] P(A_{i-1,i}^{(s_{j-1}, s_j)}). \quad (5.7)$$

In the trivial case $\alpha_{i-1} > 0$ and $\alpha_i < 0$, i.e., if $A_{i-1,i}^{(+,-)}$ occurs, we have $\tilde{K}_i \cap K_i = \tilde{K}_i$ and therefore $\mathbb{E}[I_i \mid A_{i-1,i}^{(+,-)}] = 0$. If $A_{i-1,i}^{(-,-)}$ occurs, the equality (5.6) simplifies to

$$(\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} - u_h'|_{K_i} = -\frac{h^p \alpha_{i-1}}{\tilde{x}_i - \tilde{x}_{i-1}} \llbracket u_h' \rrbracket_{x_{i-1}}.$$

Since in this case $|K_i \cap \tilde{K}_i| = \tilde{x}_i - x_{i-1}$, integrating yields

$$I_i = \frac{h^{2p}(\tilde{x}_i - x_{i-1})}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \alpha_{i-1}^2 \llbracket u_h' \rrbracket_{x_{i-1}}^2. \quad (5.8)$$

Similar calculations allow to show that if $A_{i-1,i}^{(+,+)}$ occurs, it holds

$$I_i = \frac{h^{2p}(x_i - \tilde{x}_{i-1})}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \alpha_i^2 \llbracket u_h' \rrbracket_{x_i}^2, \quad (5.9)$$

Finally, if $A_{i-1,i}^{(-,+)}$ occurs, we get

$$I_i = \frac{h^{2p}(x_i - x_{i-1})}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \xi_i^2. \quad (5.10)$$

where we denote

$$\xi_i := \alpha_{i-1} \llbracket u_h' \rrbracket_{x_{i-1}} + \alpha_i \llbracket u_h' \rrbracket_{x_i}. \quad (5.11)$$

We thus have an expression for $\mathbb{E}I_i$ due to (5.7). We now turn to $I_{i-1,i}$. Since $\tilde{K}_i \cap K_{i-1} = \emptyset$ if $\alpha_{i-1} > 0$, we have by the law of total expectation

$$\mathbb{E}I_{i-1,i} = \mathbb{E} \left[I_{i-1,i} \mid A_{i-1,i}^{(-,-)} \right] P(A_{i-1,i}^{(-,-)}) + \mathbb{E} \left[I_{i-1,i} \mid A_{i-1,i}^{(-,+)} \right] P(A_{i-1,i}^{(-,+)}). \quad (5.12)$$

Let us remark that adding and subtracting $u_h'|_{K_i}$ yields

$$(\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} - u_h'|_{K_{i-1}} = (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} - u_h'|_{K_i} - \llbracket u_h' \rrbracket_{x_{i-1}}.$$

The same computations employed for I_i allow to conclude that

$$I_{i-1,i} = \begin{cases} -h^p \alpha_{i-1} \left(\frac{h^p \alpha_{i-1}}{\tilde{x}_i - \tilde{x}_{i-1}} + 1 \right)^2 \llbracket u_h' \rrbracket_{x_{i-1}}^2, & \text{if } A_{i-1,i}^{(-,-)} \text{ occurs,} \\ -h^p \alpha_{i-1} \left(\frac{h^p}{\tilde{x}_i - \tilde{x}_{i-1}} \xi_i + \llbracket u_h' \rrbracket_{x_{i-1}} \right)^2, & \text{if } A_{i-1,i}^{(-,+)} \text{ occurs.} \end{cases}$$

which, replaced into (5.12) gives the final expression for $\mathbb{E}I_{i-1,i}$. Similarly, for $I_{i+1,i}$ we have

$$\mathbb{E}I_{i+1,i} = \mathbb{E} \left[I_{i+1,i} \mid A_{i-1,i}^{(+,+)} \right] P(A_{i-1,i}^{(+,+)}) + \mathbb{E} \left[I_{i+1,i} \mid A_{i-1,i}^{(-,+)} \right] P(A_{i-1,i}^{(-,+)}),$$

where

$$I_{i+1,i} = \begin{cases} h^p \alpha_i \left(\frac{h^p \alpha_i}{\tilde{x}_i - \tilde{x}_{i-1}} - 1 \right)^2 \llbracket u_h' \rrbracket_{x_i}^2, & \text{if } A_{i-1,i}^{(+,+)} \text{ occurs,} \\ h^p \alpha_i \left(\frac{h^p}{\tilde{x}_i - \tilde{x}_{i-1}} \xi_i - \llbracket u_h' \rrbracket_{x_i} \right)^2, & \text{if } A_{i-1,i}^{(-,+)} \text{ occurs.} \end{cases}$$

We now reassemble the quantity $I_i + I_{i-1,i} + I_{i+1,i}$ by grouping terms with regard to their conditioning on the sign of (α_{i-1}, α_i) . In particular, some algebraic simplifications yield

$$I_i + I_{i-1,i} + I_{i+1,i} = \begin{cases} h^p \alpha_i \llbracket u'_h \rrbracket_{x_i}^2 - \frac{h^{2p} \alpha_i^2}{\tilde{x}_i - \tilde{x}_{i-1}} \llbracket u'_h \rrbracket_{x_i}^2, & \text{if } A_{i-1,i}^{(+,+)} \text{ occurs,} \\ -h^p \alpha_{i-1} \llbracket u'_h \rrbracket_{x_{i-1}}^2 - \frac{h^{2p} \alpha_{i-1}^2}{\tilde{x}_i - \tilde{x}_{i-1}} \llbracket u'_h \rrbracket_{x_{i-1}}^2, & \text{if } A_{i-1,i}^{(-,-)} \text{ occurs,} \\ h^p \alpha_i \llbracket u'_h \rrbracket_{x_i}^2 - h^p \alpha_{i-1} \llbracket u'_h \rrbracket_{x_{i-1}}^2 - \frac{h^{2p}}{\tilde{x}_i - \tilde{x}_{i-1}} \xi_i^2, & \text{if } A_{i-1,i}^{(-,+)} \text{ occurs.} \end{cases}$$

We now can compute the estimator $\tilde{\mathcal{E}}_{h,1}$ by summing its local contributions, as in

$$\begin{aligned} \tilde{\mathcal{E}}_{h,1}^2 &= \sum_{i=1}^N \eta_{K,1}^2 = \sum_{i=1}^N h_i^{-(p-1)} \|e'_h\|_{L^2(\tilde{K}_i)}^2 \\ &= \sum_{i=1}^N h_i^{-(p-1)} \mathbb{E}(I_i + I_{i-1,i} + I_{i+1,i}) =: J_1 + J_2, \end{aligned}$$

where J_1 and J_2 are given by

$$\begin{aligned} J_1 &:= \frac{h^p}{2} \sum_{i=1}^N h_i^{-(p-1)} \left(\llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E}[\alpha_i \mid \alpha_i > 0] - \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E}[\alpha_{i-1} \mid \alpha_i < 0] \right), \\ J_2 &:= -\frac{h^{2p}}{4} \sum_{i=1}^N h_i^{-(p-1)} \left(\llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} \left[\frac{\alpha_{i-1}^2}{\tilde{x}_i - \tilde{x}_{i-1}} \mid A_{i-1,i}^{(-,-)} \right] + \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} \left[\frac{\alpha_i^2}{\tilde{x}_i - \tilde{x}_{i-1}} \mid A_{i-1,i}^{(+,+)} \right] \right. \\ &\quad \left. + \mathbb{E} \left[\frac{\xi_i^2}{\tilde{x}_i - \tilde{x}_{i-1}} \mid A_{i-1,i}^{(-,+)} \right] \right). \end{aligned}$$

Let us consider J_1 and J_2 separately. Rearranging the sum, noticing that under [Assumption 2.1\(i\)](#) it holds $\mathbb{E}[\alpha_i \mid \alpha_i > 0] = -\mathbb{E}[\alpha_i \mid \alpha_i < 0] = \mathbb{E}|\alpha_i|$ and recalling that $\alpha_i = (\tilde{h}_i h^{-1})^p \tilde{\alpha}_i$, we obtain

$$\begin{aligned} J_1 &= \frac{h^p}{2} \sum_{i=1}^{N-1} \left(h_i^{-(p-1)} + h_{i+1}^{-(p-1)} \right) \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E}[\alpha_i \mid \alpha_i > 0] \\ &= \frac{\mathbb{E}|\tilde{\alpha}_1|}{2} \sum_{i=1}^{N-1} \left(h_i^{-(p-1)} + h_{i+1}^{-(p-1)} \right) \tilde{h}_i^p \llbracket u'_h \rrbracket_{x_i}^2. \end{aligned}$$

Now, let us remark that if the mesh is λ -quasi-uniform, it holds

$$(1 + \lambda^{-(p-1)}) \tilde{h}_i \leq \left(h_i^{-(p-1)} + h_{i+1}^{-(p-1)} \right) \tilde{h}_i^p \leq (1 + \lambda^{p-1}) \tilde{h}_i,$$

which implies

$$\frac{\mathbb{E}|\tilde{\alpha}_1| (1 + \lambda^{-(p-1)})}{2} \mathcal{J}^2(u_h) \leq J_1 \leq \frac{\mathbb{E}|\tilde{\alpha}_1| (1 + \lambda^{p-1})}{2} \mathcal{J}^2(u_h). \quad (5.13)$$

We now turn to J_2 . Clearly, we have $J_2 \leq 0$, which implies the desired upper bound together with (5.13). For the lower bound, we remark that in both cases $A_{i-1,i}^{(+,+)}$ and $A_{i-1,i}^{(-,-)}$ occur, we have that $\tilde{x}_i - \tilde{x}_{i-1} \geq h_i/2$, and if $A_{i-1,i}^{(-,+)}$ occurs, we have $\tilde{x}_i - \tilde{x}_{i-1} \geq h_i$. Hence, simplifying the conditioning in the first and second terms, we obtain

$$\begin{aligned} J_2 &\geq -\frac{h^{2p}}{4} \sum_{i=1}^N h_i^{-p} \left(2 \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E}[\alpha_i^2 \mid \alpha_i > 0] + 2 \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E}[\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] \right. \\ &\quad \left. + \mathbb{E}[\xi_i^2 \mid A_{i-1,i}^{(-,+)}] \right). \end{aligned}$$

We now consider ξ_i given in (5.11) and use $(a+b)^2 \leq 2(a^2+b^2)$ for $a = \alpha_{i-1} \llbracket u'_h \rrbracket_{x_{i-1}}$ and $b = \alpha_i \llbracket u'_h \rrbracket_{x_i}$ to obtain

$$\mathbb{E} \left[\xi_i^2 \mid A_{i-1,i}^{(-,+)} \right] \leq 2 \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} [\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] + 2 \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} [\alpha_i^2 \mid \alpha_i > 0].$$

Therefore

$$J_2 \geq -h^{2p} \sum_{i=1}^N h_i^{-p} \left(\llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} [\alpha_i^2 \mid \alpha_i > 0] + \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} [\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] \right).$$

Rewriting the sum and replacing the definition of α_i yields

$$J_2 \geq - \sum_{i=1}^{N-1} \bar{h}_i^{2p} \llbracket u'_h \rrbracket_{x_i}^2 \left(h_i^{-p} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i > 0] + h_{i+1}^{-p} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i < 0] \right).$$

Now $\bar{h}_i = \min\{h_i, h_{i+1}\}$ implies $h_i^{-p} \leq \bar{h}_i^{-p}$ and $h_{i+1}^{-p} \leq \bar{h}_i^{-p}$, which gives

$$\begin{aligned} J_2 &\geq -2 \sum_{i=1}^{N-1} \bar{h}_i^p \llbracket u'_h \rrbracket_{x_i}^2 \left(\frac{1}{2} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i > 0] + \frac{1}{2} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i < 0] \right) \\ &\geq -2 \mathbb{E} |\bar{\alpha}_1|^2 \sum_{i=1}^{N-1} \bar{h}_i^p \llbracket u'_h \rrbracket_{x_i}^2, \end{aligned}$$

where we applied the law of total expectation on the second line. Finally, we have $\bar{h}_i \leq h$ and $p \geq 1$, which yield

$$J_2 \geq -2 \mathbb{E} |\bar{\alpha}_1|^2 h^{p-1} \mathcal{J}^2(u_h).$$

Combining this with (5.13) then yields the desired lower bound and thus concludes the proof. \square

Let us remark that the coefficient appearing in the lower bound of Lemma 5.3 is positive if Assumption 3.4 holds. We now prove the equivalence of the estimator $\tilde{\mathcal{E}}_{h,2}$ given in Definition 3.1 with $\mathcal{J}(u_h)$.

Lemma 5.4. *Let Assumption 2.1 hold and let the mesh \mathcal{T}_h be λ -quasi uniform. Then, it holds*

$$\frac{\mathbb{E} |\bar{\alpha}_1|^2}{2(1+\lambda)^2 \lambda^{2p-1}} \mathcal{J}^2(u_h) \leq \tilde{\mathcal{E}}_{h,2}^2 \leq 3 \mathbb{E} |\bar{\alpha}_1|^2 \mathcal{J}^2(u_h),$$

where $\tilde{\mathcal{E}}_{h,2}$ is given in Definition 3.1.

Proof. As $|K_i| = h_i$, we have

$$\tilde{\mathcal{E}}_{h,2} = \sum_{i=1}^N h_i^{-(2p-3)} \mathbb{E} \left[\left| u'_h|_{K_i} - (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} \right|^2 \right].$$

Proceeding similarly to (5.8), (5.9) and (5.10) and applying the law of total expectation, we obtain

$$\begin{aligned} \mathbb{E} \left[\left| u'_h|_{K_i} - (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} \right|^2 \right] &= \frac{h^{2p}}{4} \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} \left[\frac{\alpha_i^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(+,+)} \right] \\ &\quad + \frac{h^{2p}}{4} \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} \left[\frac{\alpha_{i-1}^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,-)} \right] \\ &\quad + \frac{h^{2p}}{4} \mathbb{E} \left[\frac{\xi_i^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,+)} \right], \end{aligned} \quad (5.14)$$

where we recall the notation ξ_i introduced in (5.11). Let us first consider the lower bound. Since $\xi_i^2 \geq 0$ a.s., and $\tilde{x}_i - \tilde{x}_{i-1} \leq (1+\lambda)h_i$ a.s. under the assumption that the mesh is λ -quasi-uniform, we have

$$\begin{aligned} \mathbb{E} \left[\left| u'_h|_{K_i} - (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} \right|^2 \right] &\geq \frac{h^{2p} h_i^{-2}}{4(1+\lambda)^2} \left(\llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} [\alpha_i^2 \mid \alpha_i > 0] \right. \\ &\quad \left. + \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} [\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] \right). \end{aligned}$$

Assembling the sum, rearranging terms and recalling that $\alpha_i = (h^{-1}\bar{h}_i)^p \bar{\alpha}_i$ with $\bar{h}_i = \min\{h_i, h_{i+1}\}$, we then obtain

$$\begin{aligned}\widehat{\mathcal{E}}_{h,2}^2 &\geq \frac{1}{2(1+\lambda)^2} \sum_{i=1}^{N-1} \bar{h}_i^{2p} h_i^{1-2p} \llbracket u'_h \rrbracket_{x_i}^2 \left(\frac{1}{2} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i > 0] + \frac{1}{2} \mathbb{E} [\bar{\alpha}_i^2 \mid \bar{\alpha}_i < 0] \right) \\ &\geq \frac{\mathbb{E} |\bar{\alpha}_1|^2}{2(1+\lambda)^2 \lambda^{2p-1}} \sum_{i=1}^{N-1} \bar{h}_i \llbracket u'_h \rrbracket_{x_i}^2 = \frac{\mathbb{E} |\bar{\alpha}_1|^2}{2(1+\lambda)^2 \lambda^{2p-1}} \mathcal{J}^2(u_h),\end{aligned}$$

where we employed the law of total expectation and the inequality $h_i^{1-2p} \leq \lambda^{1-2p} \bar{h}_i^{1-2p}$ on the second line. Hence, we proved the lower bound. For the upper bound, using again the inequality $(a+b)^2 \leq 2(a^2+b^2)$ we obtain

$$\xi_i^2 \leq 2\alpha_i^2 \llbracket u'_h \rrbracket_{x_i}^2 + 2\alpha_{i-1}^2 \llbracket u'_h \rrbracket_{x_{i-1}}^2, \quad \text{a.s.},$$

so that

$$\begin{aligned}\mathbb{E} \left[\frac{\xi_i^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,+)} \right] &\leq 2 \llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} \left[\frac{\alpha_i^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,+)} \right] \\ &\quad + 2 \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} \left[\frac{\alpha_{i-1}^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,+)} \right].\end{aligned}$$

Under $A_{i-1,i}^{(-,+)}$, we have $\tilde{x}_i - \tilde{x}_{i-1} \geq h_i$, which implies

$$\mathbb{E} \left[\frac{\xi_i^2}{(\tilde{x}_i - \tilde{x}_{i-1})^2} \mid A_{i-1,i}^{(-,+)} \right] \leq 2h_i^{-2} \left(\llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} [\alpha_i^2 \mid \alpha_i > 0] + \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} [\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] \right).$$

Then, considering that under $A_{i-1,i}^{(+,+)}$ or $A_{i-1,i}^{(-,-)}$ it holds $\tilde{x}_i - \tilde{x}_{i-1} \geq h_i/2$ and plugging into (5.14) we have

$$\mathbb{E} \left[\left| u'_h|_{K_i} - (\tilde{\mathcal{I}}u_h)'|_{\tilde{K}_i} \right|^2 \right] \leq \frac{3}{2} h_i^{-2} h^{2p} \left(\llbracket u'_h \rrbracket_{x_i}^2 \mathbb{E} [\alpha_i^2 \mid \alpha_i > 0] + \llbracket u'_h \rrbracket_{x_{i-1}}^2 \mathbb{E} [\alpha_{i-1}^2 \mid \alpha_{i-1} < 0] \right).$$

We can therefore reassemble and rearrange the sum following the same procedure as for the lower bound, which, together with $h_i^{1-2p} \leq \bar{h}_i^{1-2p}$, yields

$$\widehat{\mathcal{E}}_{h,2}^2 \leq 3 \mathbb{E} |\bar{\alpha}_1|^2 \mathcal{J}^2(u_h),$$

which proves the desired result. \square

We finally prove the equivalence of the deterministic error estimator \mathcal{E}_h given in Definition 5.1 with the quantity $\mathcal{J}(u_h)$.

Lemma 5.5. *Let the mesh \mathcal{T}_h be λ -quasi-uniform. Then, it holds*

$$\frac{\lambda m^2}{6(1+\lambda)^3 M^2} \mathcal{J}^2(u_h) \leq \mathcal{E}_h^2 \leq \frac{2\lambda^2 M^2}{3(1+\lambda)m^2} \mathcal{J}^2(u_h),$$

where \mathcal{E}_h is given in Definition 5.1 and where $m = \underline{\kappa}$ and $M = \|\kappa\|_{L^\infty(D)}$.

Proof. Simple algebraic computations yield

$$\|\ell_j\|_{L^2(K_j)}^2 = \frac{h_j}{3} (\tau_{j,0}^2 - \tau_{j,0}\tau_{j,1} + \tau_{j,1}^2),$$

where ℓ_j are the linear functions employed in Definition 5.1. Applying the inequalities $(a^2+b^2)/2 \leq a^2-ab+b^2 \leq 2(a^2+b^2)$ we obtain

$$\frac{h_j}{6M^2} (\tau_{j,0}^2 + \tau_{j,1}^2) \leq \eta_j^2 \leq \frac{2h_j}{3m^2} (\tau_{j,0}^2 + \tau_{j,1}^2).$$

We now remark that if the mesh \mathcal{T}_h is λ -quasi-uniform and under the assumptions on κ it holds for $k \in \{0, 1\}$

$$\frac{m^2}{(1+\lambda)^2} \llbracket u'_h \rrbracket_{x_{j-k}}^2 \leq \tau_{j,k}^2 \leq \frac{\lambda^2 M^2}{(1+\lambda)^2} \llbracket u'_h \rrbracket_{x_{j-k}}^2,$$

which, in turn, implies

$$\frac{m^2 h_j}{6(1+\lambda)^2 M^2} \left(\llbracket u'_h \rrbracket_{x_{j-1}}^2 + \llbracket u'_h \rrbracket_{x_j}^2 \right) \leq \eta_j^2 \leq \frac{2\lambda^2 M^2 h_j}{3(1+\lambda)^2 m^2} \left(\llbracket u'_h \rrbracket_{x_{j-1}}^2 + \llbracket u'_h \rrbracket_{x_j}^2 \right).$$

We now focus on the upper bound. Reassembling the global error estimator \mathcal{E}_h , we have

$$\begin{aligned} \mathcal{E}_h^2 &\leq \frac{2\lambda^2 M^2}{3(1+\lambda)^2 m^2} \sum_{j=1}^N h_j \left(\llbracket u'_h \rrbracket_{x_{j-1}}^2 + \llbracket u'_h \rrbracket_{x_j}^2 \right) \\ &= \frac{2\lambda^2 M^2}{3(1+\lambda)^2 m^2} \sum_{j=1}^{N-1} (h_j + h_{j+1}) \llbracket u'_h \rrbracket_{x_j}^2 \\ &\leq \frac{2\lambda^2 M^2}{3(1+\lambda)m^2} \mathcal{J}^2(u_h), \end{aligned}$$

where we recall $\bar{h}_j = \min\{h_j, h_{j+1}\}$, so that $h_j + h_{j+1} \leq (1+\lambda)\bar{h}_j$. We conclude the proof proceeding similarly for the lower bound as in [Lemma 5.4](#). \square

We can finally prove [Theorem 3.6](#) and conclude the error analysis.

Proof of Theorem 3.6. Let us first consider $\tilde{\mathcal{E}}_{h,1}$. Under [Assumption 3.4](#), we have for the lower bound of [Lemma 5.3](#)

$$\left(\frac{\mathbb{E} |\bar{\alpha}_1| (1 + \lambda^{-(p-1)})}{2} - 2h^{p-1} \mathbb{E} |\bar{\alpha}_1|^2 \right) \mathcal{J}^2(u_h) \geq C \mathbb{E} |\bar{\alpha}_1| \mathcal{J}^2(u_h)$$

for a constant $C > 0$. Hence, due to [Lemma 5.5](#) we have that there exists a constant \hat{C} such that

$$\tilde{\mathcal{E}}_{h,1} \geq \hat{C} \mathcal{E}_h,$$

and therefore, [Theorem 5.2](#) implies

$$\|u - u_h\|_V \leq C_{\text{up}} \mathcal{E}_h \leq C_{\text{up}} \hat{C} \tilde{\mathcal{E}}_{h,1},$$

which yields the desired upper bound with $\tilde{C}_{\text{up}} = \hat{C} C_{\text{up}}$. The lower bound follows equivalently under the additional regularity required by [Theorem 5.2](#). Similarly, the results for $\tilde{\mathcal{E}}_{h,2}$ follow from [Lemmas 5.3](#) and [5.5](#), together with [Theorem 5.2](#). \square

6. Conclusion

We have introduced a novel probabilistic methods for PDEs based on the FEM and random meshes, the RM-FEM. We demonstrated how our methodology can be successful when employed in pipelines of computations, such as Bayesian inverse problems. We also show a rigorous use of probabilistic methods for a posteriori error estimators, often speculated in the field. Extending such analysis to PN methods for ODEs would be of interest, thus creating a link between the guiding principles of PN and more classical theories. Generalizing the analysis of the RM-FEM to higher-dimensional PDEs as well as for parabolic or hyperbolic problems represent also interesting future work. We also stress that we have applied so far the RM tool to a simple FEM and in view of the variety of FEMs developed over the years, it would be of interest to build such methodology in other type of methods (e.g. discontinuous methods) and other type of problems (e.g. nonlinear problems) to explore the benefit, the scope and the limitations of RM-FEM solvers and their usefulness for practical engineering applications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are partially supported by the Swiss National Science Foundation, Switzerland, under Grant No. 200020_172710.

References

- [1] C.J. Oates, T.J. Sullivan, A modern retrospective on probabilistic numerics, *Stat. Comput.* 29 (2019) 1335–1351.
- [2] P. Hennig, M.A. Osborne, M. Girolami, Probabilistic numerics and uncertainty in computations, *Proc. A* 471 (2015) 20150142, 17.
- [3] J. Cockayne, C.J. Oates, T.J. Sullivan, M. Girolami, Bayesian probabilistic numerical methods, *SIAM Rev.* 61 (2019) 756–789.
- [4] J. Skilling, Bayesian solution of ordinary differential equations, in: *Maximum Entropy and Bayesian Methods*, Springer, 1992, pp. 23–37.
- [5] H. Kersting, 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.
- [6] H. Kersting, T.J. Sullivan, P. Hennig, Convergence rates of Gaussian ODE filters, *Stat. Comput.* 30 (2020) 1791–1816.
- [7] F. Tronarp, H. Kersting, S. Särkkä, P. Hennig, Probabilistic solutions to ordinary differential equations as nonlinear Bayesian filtering: a new perspective, *Stat. Comput.* 29 (2019) 1297–1315.
- [8] M. Schober, S. Särkkä, P. Hennig, A probabilistic model for the numerical solution of initial value problems, *Stat. Comput.* 29 (2019) 99–122.
- [9] O.A. Chkrebtii, D.A. Campbell, B. Calderhead, M.A. Girolami, Bayesian solution uncertainty quantification for differential equations, *Bayesian Anal.* 11 (2016) 1239–1267.
- [10] M. Schober, D. Duvenaud, P. Hennig, Probabilistic ODE solvers with Runge–Kutta means, in: *Advances in Neural Information Processing Systems 27*, Curran Associates, Inc., 2014, pp. 739–747.
- [11] P.R. Conrad, M. Girolami, S. Särkkä, A.M. Stuart, K. Zygalakis, Statistical analysis of differential equations: introducing probability measures on numerical solutions, *Stat. Comput.* 27 (2017) 1065–1082.
- [12] H.C. Lie, A.M. Stuart, T.J. Sullivan, Strong convergence rates of probabilistic integrators for ordinary differential equations, *Stat. Comput.* 29 (2019) 1265–1283.
- [13] A. Abdulle, G. Garegnani, Random time step probabilistic methods for uncertainty quantification in chaotic and geometric numerical integration, *Stat. Comput.* 30 (2020) 907–932.
- [14] O. Teymur, H.C. Lie, T. Sullivan, B. Calderhead, Implicit probabilistic integrators for ODEs, in: *Advances in Neural Information Processing Systems*, 2018, pp. 7244–7253.
- [15] O. Teymur, K. Zygalakis, B. Calderhead, Probabilistic linear multistep methods, in: *Advances in Neural Information Processing Systems*, 2016, pp. 4321–4328.
- [16] J. Cockayne, C.J. Oates, T.J. Sullivan, M. Girolami, Probabilistic numerical methods for PDE-constrained Bayesian inverse problems, *AIP Conf. Proc.* 1853 (2017) 060001.
- [17] J. Cockayne, C.J. Oates, T.J. Sullivan, M. Girolami, Probabilistic numerical methods for partial differential equations and Bayesian inverse problems, 2017, arXiv preprint arXiv:1605.07811.
- [18] C.J. Oates, J. Cockayne, R.G. Aykroyd, M. Girolami, Bayesian probabilistic numerical methods in time-dependent state estimation for industrial hydrocyclone equipment, *J. Amer. Statist. Assoc.* 114 (2019) 1518–1531.
- [19] H. Owadi, Bayesian numerical homogenization, *Multiscale Model. Simul.* 13 (2015) 812–828.
- [20] H. Owadi, Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games, *SIAM Rev.* 59 (2017) 99–149.
- [21] H. Owadi, L. Zhang, Gamblets for opening the complexity-bottleneck of implicit schemes for hyperbolic and parabolic ODEs/PDEs with rough coefficients, *J. Comput. Phys.* 347 (2017) 99–128.
- [22] M. Raissi, P. Perdikaris, G.E. Karniadakis, Inferring solutions of differential equations using noisy multi-fidelity data, *J. Comput. Phys.* 335 (2017) 736–746.
- [23] M. Raissi, P. Perdikaris, G.E. Karniadakis, Machine learning of linear differential equations using Gaussian processes, *J. Comput. Phys.* 348 (2017) 683–693.
- [24] M. Girolami, E. Febrianto, G. Yin, F. Cirak, The statistical finite element method (statFEM) for coherent synthesis of observation data and model predictions, *Comput. Methods Appl. Mech. Engrg.* 375 (2021) 113533, 32.
- [25] O.A. Chkrebtii, D.A. Campbell, Adaptive step-size selection for state-space probabilistic differential equation solvers, *Stat. Comput.* 29 (2019) 1285–1295.
- [26] N. Bosch, P. Hennig, F. Tronarp, Calibrated adaptive probabilistic ODE solvers, 2020, arXiv preprint arXiv:2012.08202.
- [27] I. Babuška, W.C. Rheinboldt, A posteriori error analysis of finite element solutions for one-dimensional problems, *SIAM J. Numer. Anal.* 18 (1981) 565–589.
- [28] R. Verfürth, A posteriori error estimation techniques for finite element methods, in: *Numerical Mathematics and Scientific Computation*, Oxford University Press, Oxford, 2013.
- [29] M. Ainsworth, J.T. Oden, A Posteriori Error Estimation in Finite Element Analysis, in: *Pure and Applied Mathematics (New York)*, Wiley-Interscience [John Wiley & Sons], New York, 2000.
- [30] R. Verfürth, A posteriori error estimation and adaptive mesh-refinement techniques, *J. Comput. Appl. Math.* 50 (1994) 67–83.
- [31] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. I. The recovery technique, *Internat. J. Numer. Methods Engrg.* 33 (1992) 1331–1364.
- [32] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. II. Error estimates and adaptivity, *Internat. J. Numer. Methods Engrg.* 33 (1992) 1365–1382.
- [33] M. Croci, M.B. Giles, M.E. Rognes, P.E. Farrell, Efficient white noise sampling and coupling for multilevel Monte Carlo with nonnested meshes, *SIAM/ASA J. Uncertain. Quantif.* 6 (2018) 1630–1655.
- [34] M. Croci, P.E. Farrell, Complexity bounds on supermesh construction for quasi-uniform meshes, *J. Comput. Phys.* 414 (2020) 109459, 7.

- [35] H.C. Lie, T.J. Sullivan, A.L. Teckentrup, Random forward models and log-likelihoods in Bayesian inverse problems, *SIAM/ASA J. Uncertain. Quantif.* 6 (2018) 1600–1629.
- [36] M. Dashti, A.M. Stuart, The Bayesian Approach to Inverse Problems, in: *Handbook of Uncertainty Quantification*, Springer, 2016, pp. 1–118.
- [37] A.M. Stuart, Inverse problems: a Bayesian perspective, *Acta Numer.* 19 (2010) 451–559.
- [38] T.J. Sullivan, Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors, *Inverse Probl. Imaging* 11 (2017) 857–874.
- [39] D. Calvetti, M. Dunlop, E. Somersalo, A.M. Stuart, Iterative updating of model error for Bayesian inversion, *Inverse Problems* 34 (2018) 025008, 38.
- [40] D. Calvetti, O. Ernst, E. Somersalo, Dynamic updating of numerical model discrepancy using sequential sampling, *Inverse Problems* 30 (2014) 114019, 19.
- [41] A. Abdulle, A. Di Blasio, A Bayesian numerical homogenization method for elliptic multiscale inverse problems, *SIAM/ASA J. Uncertain. Quantif.* 8 (2020) 414–450.
- [42] A. Abdulle, G. Garegnani, A. Zaroni, Ensemble Kalman filter for multiscale inverse problems, *Multiscale Model. Simul.* 18 (2020) 1565–1594.
- [43] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, in: *Applied Mathematical Sciences*, vol. 160, Springer-Verlag, New York, 2005.
- [44] D.P. Kroese, T. Taimre, Z.I. Botev, *Handbook of Monte Carlo Methods*, Vol. 706, John Wiley & Sons, 2013.
- [45] M. Vihola, Robust adaptive Metropolis algorithm with coerced acceptance rate, *Stat. Comput.* 22 (2012) 997–1008.
- [46] S.L. Cotter, G.O. Roberts, A.M. Stuart, D. White, MCMC methods for functions: modifying old algorithms to make them faster, *Statist. Sci.* 28 (2013) 424–446.
- [47] M. Hairer, A.M. Stuart, S.J. Vollmer, Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions, *Ann. Appl. Probab.* 24 (2014) 2455–2490.
- [48] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, in: *Classics Appl. Math.*, vol. 40, SIAM, Philadelphia, 2002.
- [49] A. Quarteroni, Numerical Models for Differential Problems, in: *Modeling, Simulation & Applications*, vol. 2, Springer, 2009.
- [50] S.C. Brenner, L.R. Scott, *The Mathematical Theory of Finite Element Methods*, third ed., in: *Texts in Applied Mathematics*, vol. 15, Springer, New York, 2008.