Semester project report

Solving PDEs with log-normal random field coefficients

Name:	Matteo Calafà
Date:	June 13, 2022
Chair:	CSQI
Supervisor:	Prof. Fabio Nobile
Assistant:	Sundar Ganesh

Contents

Acronyms

FEM Finite Element Method.

KL Karhunen–Loève expansion.

MLMC Multilevel Monte Carlo method.

PDE Partial Differential Equation.

SPDE Stochastic Partial Differential Equation.

1 Introduction

The modelling and resolution of stochastic partial differential equations (SPDE) using Gaussian fields as random coefficients have always been of great relevance in many applications such as geology, hydrocarbon reservoir modelling, meteorology and even biology ([22], [2], [18]). Gaussian random fields appear naturally when representing spatial noises and uncertainty in input parameters to PDEs such as diffusivity. In this research, we aim to investigate a classical elliptic equation with a log-normal random coefficient where the corresponding Gaussian field is governed by the Matérn covariance function ([25]). Again, the problem is chosen due to the relevance of the elliptic PDE in the applications areas listed above. Moreover, this model has already been studied extensively in literature. More precisely, analytical properties such as existence and uniqueness of the solution have already been proven (see [3]). On the other hand, recent research is now focused on proposing different numerical methods to deal with non-integer fractional Laplacian exponents ([23]) to generate Matérn samples with general smoothness. For instance, [10] introduces and analyses the approximation of the solution using expansions with Hermite polynomials. [4] describes another type of numerical solution that combines the finite element method and the rational expansion of x^{-k}. Also, in [5], another method is discussed that is based instead on a sinc-quadrature for the inverse fractional Laplacian operator. Finally, [11] presents a full study on numerical solutions using spectral Galerkin methods and also discusses some important results for more general Galerkin approximations.
In this work, our goal is to study the feasibility and efficiency of a Multi-Level Monte Carlo method (MLMC, [15], [16]) coupled with a standard Finite Element Method (FEM) to estimate the statistics of random output quantities of interest of an elliptic PDE with uncertain inputs. These output quantities of interest are typically functionals of the solution of the SPDE. Instead of using standard methods for the discrete realizations of Gaussian fields (such as Cholesky factorizations, circulant embeddings or Karhunen-Loéve expansions), we will instead exploit the fractional Laplacian equivalent formulation from Whittle ([30]). This choice is specific only for Matérn fields but it is in general less costly than the above mentioned methods. Furthermore, this method will permit us to employ FEM twice, sharing some elements for both the elliptic and Whittle problems and, so, saving a noteworthy quantity of memory.
The article is organized in the following way. In section 2 we state the SPDE model we previously introduced and that we aim to solve. In section 3 we show how the problem can be numerically discretized presenting both the spatial discretization (using FEM) and the sampling techniques (using single level or multilevel Monte Carlo methods). In section 4 we report the results we obtained in the case of single level sampling, i.e. when adopting a standard Monte Carlo method. These outcomes are also fundamental to analyse in section 5 the more complicated multilevel approach. Important results and observations as well as critical points will be shown. These analyses include a detailed study on the mean square error properties pursuing a priori and a posteriori bounds. Sections 4 and 5 also include descriptions of possible algorithms to generate both single level and multi level discrete realizations of the white noise.

2 Formulation of the problem

We present here the analytical formulation of our SPDE problem. The numerical discretization (section 3) and all the following results on convergence errors and optimal resolution methods (sections 4, 5) are uniquely addressed to the analytical problem described in this section.

2.1 The model problem

The goal of this research is the analysis and resolution of a standard elliptic problem in divergence form with Dirichlet boundary conditions and diffusivity defined as a spatial Matérn field. This system is written explicitly in equation (1).

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

where $D \subset \mathbb{R}^d$ is a bounded domain, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $q, a : D \times \Omega \to \mathbb{R}$ are stochastic fields. In particular, a is a log-normal Matérn field ([25]), which means that $a(\mathbf{x},\omega) = e^{\gamma(\mathbf{x},\omega)}$ where $\gamma : D \times \Omega \to \mathbb{R}$ is a Gaussian random field with zero mean and Matérn covariance function that is defined as:

$$C(\mathbf{x},\mathbf{y}) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa r)^\nu \mathcal{K}_\nu(\kappa r), \qquad r = ||\mathbf{x}-\mathbf{y}||_2, \qquad \kappa = \frac{\sqrt{8\nu}}{\lambda}. \tag{2}$$

Here, $\sigma^2 > 0$ represents the variance, $\nu > 0$ is a smoothness parameter and $\lambda > 0$ states the correlation length. In addition, Γ is the Euler Gamma function and $\mathcal{K}_\nu$ is the modified Bessel function of the second kind. We first note that the correlation is isotropic because it only depends on the Euclidean distance between the two points.
We report now one of the main properties of this stochastic field.

Proposition 2.1 (Regularity of the Matérn stochastic field samples)**.** *The Matérn stochastic field admits a version whose trajectories belong to $C^{0,\nu'}(\bar{D})$ for all $\nu' < \nu$.*

Proof. We follow some steps from the proof of Preposition 2.1 in [8] that we specialize in the case of Matérn fields and we generalize to Hölder continuous covariance functions. Notice that the Matérn function is only a function of r that we will rename $C(r)$. Therefore:

$$\begin{aligned} \mathbb{E}[|\gamma(\mathbf{x}) - \gamma(\mathbf{y})|^2] =& \mathbb{E}[\gamma(\mathbf{x})^2] - 2\mathbb{E}[\gamma(\mathbf{x})\gamma(\mathbf{y})] + \mathbb{E}[\gamma\mathbf{y})^2] \\ =& 2(C(0) - C(r)) \\ =& 2\sigma^2 \left(1 - \frac{(\kappa r)^\nu \mathcal{K}_\nu(\kappa r)}{2^{\nu-1}\Gamma(\nu)}\right) \\ \leq& L\|\mathbf{x}-\mathbf{y}\|^{2\nu}, \end{aligned}$$

where the last inequality comes from the expansions of the Bessel function (see for instance [10]), and L is a constant that only depends on the field parameters. A simple property of zero-mean Gaussian random variables states that:

$$\forall p \in \mathbb{N}, \quad \exists c_p > 0 : \quad \mathbb{E}\left[|X|^{2p}\right] \leq c_p \mathbb{E}\left[|X|^2\right]^p$$

for every mean-free normal random variable X. Therefore, since $(\gamma(\mathbf{x}) - \gamma(\mathbf{y}))$ is a mean-free normal random variable, it holds:

$$\mathbb{E}[|\gamma(\mathbf{x}) - \gamma(\mathbf{y})|^{2p}] \leq c_p (2L)^p \|\mathbf{x} - \mathbf{y}\|^{2\nu p}.$$

Hence, Kolmogorov's continuity theorem states that there exists a version of γ which is in $C^{0,\alpha}(\bar{D})$ for every $\alpha < \frac{2\nu p - d}{2p}$. However, since the result holds for every $p \in \mathbb{N}$, we can let $p \to \infty$ and we get the Hölder continuity for every $\alpha < \nu$. □

To conclude this section, we stress again the choice of this model because of its importance in many science fields such as geology, hydrocarbon reservoir modelling, meteorology.

2.2 The Whittle-Matérn fractional Laplacian equation

Equation (1) can potentially be solved with well-known numerical methods for diffusion equations followed by a wise sampling strategy. However, sampling techniques usually require the Gaussian random field a to be discretized as a Gaussian random vector. The simplest method is to discretize the stochastic field evaluating it only on some specific points and computing the covariance matrix as $C_{ij} := C(\mathbf{x}_i, \mathbf{x}_j)$. Then, the Gaussian vector can be sampled using the $H\mathbf{Z}$ product where H is the Cholesky factorization of the covariance matrix and $\mathbf{Z}$ a standard normal sample. However, the Cholesky factorization of dense matrices could be very expensive (indeed, the cost scales as the third power of the number of evaluations). Other methods consist in the truncated expansion through certain basis functions, e.g. Karhunen-Loéve expansions (KL), but generally they require some demanding calculations as well. The circulant embedding method ([17]) is fast but restricts to few simple cases, i.e. when the grid is uniform. Therefore, it is not so simple to choose which method to adopt and, presumably, the most performant strategy in this setting comes instead from the equivalent formulation given by Whittle ([30]). It can be proven that a single realization of the Matérn field can be obtained by solving the following fractional Laplace equation starting from a realization of the white noise:

$$(\mathcal{I} - \kappa^{-2}\Delta)^k u(\mathbf{x}, \omega) = \eta \dot{W}(\mathbf{x}, \omega), \qquad \mathbf{x} \in \mathbb{R}^d, \qquad \omega \in \Omega, \tag{3}$$

where $k = \nu/2 + d/4$, $\mathcal{I}$ is the identity operator and $\dot{W}$ is a spatial white noise. η is instead defined in the following way:

$$\eta = \frac{\sigma}{\hat{\sigma}}, \qquad \hat{\sigma}^2 = \frac{\Gamma(\nu)\nu^{d/2}}{\Gamma(\nu + d/2)} \left(\frac{2}{\pi}\right)^{d/2} \lambda^{-d}.$$

Through this equivalent formulation, one can simply generate the white noise to get a realization of the more complex Matérn field. Obviously, this requires an additional cost to numerically solve equation (3). In addition, we note two more issues. First of all, the equation involves the fractional Laplacian (see [23] for the definition) which is in general not straightforward to solve. Although, the equation is much simpler when k is an integer and it is even a linear diffusion equation when $k = 1$. The second issue is related to the definition of the problem on the entire $\mathbb{R}^d$ space. One therefore needs to truncate the domain to solve the problem numerically. We will show in section 3.1 how to practically solve this difficulty.
On the other hand, this choice is suitable for our type of problem because it employs a finite

element discretization that could be partially used also for equation (1). Moreover, this formulation is very general and only requires the much simpler realization of the white noise.
To conclude, we sum up here the full analytical problem:

$$\begin{cases} -\nabla \cdot \left(e^{u(\mathbf{x},\omega)} \nabla q(\mathbf{x},\omega)\right) = f(\mathbf{x}) & \mathbf{x} \in D, \omega \in \Omega \\ (\mathcal{I} - \kappa^{-2}\Delta)^k u(\mathbf{x},\omega) = \eta \dot{W}(\mathbf{x},\omega) & \mathbf{x} \in \mathbb{R}^d, \omega \in \Omega \\ q(\mathbf{x},\omega) = 0 & \mathbf{x} \in \partial D, \omega \in \Omega \end{cases} \tag{4}$$

3 Discretization of the problem

In this section, we show how to fully discretize the problem in (4) to practically achieve an approximation of its solution. Because of both the differential and stochastic nature of the problem, we will start with the numerical schemes and move later to the sampling techniques.

3.1 Truncation of the infinite domain

We start the discretization analysis by noticing that the original definition of the problem as in (4) is not solvable numerically because of the unbounded domain $\mathbb{R}^d$. What is usually performed is a truncation to a bounded domain $G \subsetneq \mathbb{R}^d$. This approximation is justified by the fact that equation (1) requires $u(\mathbf{x},\omega)$ to be evaluated only on D and not on the full $\mathbb{R}^d$ space. Therefore, a truncation to a bounded domain that is large enough and the addition of homogeneous boundary conditions turn out to be a reasonable approximation of the original problem. Usually, homogeneous Dirichlet boundary conditions are preferred (because of their simplicity and the restrictions to H_0^1 spaces). Hence, we report here the truncated version of the original analytical problem (4):

$$\begin{cases} -\nabla \cdot \left(e^{u(\mathbf{x},\omega)} \nabla q(\mathbf{x},\omega)\right) = f(\mathbf{x}) & \mathbf{x} \in D, \omega \in \Omega & \text{(5a)} \\ (\mathcal{I} - \kappa^{-2}\Delta)^k u(\mathbf{x},\omega) = \eta \dot{W}(\mathbf{x},\omega) & \mathbf{x} \in G, \omega \in \Omega & \text{(5b)} \\ q(\mathbf{x},\omega) = 0 & \mathbf{x} \in \partial D, \omega \in \Omega & \text{(5c)} \\ u(\mathbf{x},\omega) = 0 & \mathbf{x} \in \partial G, \omega \in \Omega. & \text{(5d)} \end{cases}$$

Notice now that the hypothesis that D is a subset of G is clearly necessary to have a Matérn field that can also be evaluated on D in (5a). Let us see now how concretely this approximation influences the error with respect to the original system in (4).
We report the error analysis results from [19]. Here, a spectral expansion of the Matérn covariance function makes it possible to analyse and bound errors in the case of homogeneous Dirichlet, Neumann or periodic boundary conditions applied to (3). $G \supset D$ is defined as a tensor product domain $(0, L)^d$ (i.e., an hypercube of length L) such that the minimum distance from D is $\delta/2$, $\delta > 0$. Furthermore, $l > 0$ represents the maximum length of D. More technical details can be found in [19].

Theorem 3.1 (Truncation stability ([19]))**.** *Consider equation in (3) truncated to a bounded domain $G \subsetneq \mathbb{R}^d$ with the addition of Dirichlet, Neumann or periodic homogeneous boundary conditions. Define $C(\mathbf{x},\mathbf{y})$ as the original covariance function and $C_G(\mathbf{x},\mathbf{y})$ the covariance function*

obtained from the truncated domain as in (5). Hence:

$$\|C_G(\mathbf{x},\mathbf{y}) - C(\mathbf{x},\mathbf{y})\|_{L^\infty(D)} \leq A \cdot \sigma^2 \mathcal{M}_\nu(\kappa\delta)$$

where $\mathcal{M}_\nu$ *is the unitary Matérn function*

$$\mathcal{M}_\nu(x) := \frac{x^\nu \mathcal{K}_\nu(x)}{2^{\nu-1}\Gamma(\nu)},$$

A *is a constant that depends on the size of* D

$$A = (2^d - 1) \cdot \left(1 + \frac{2^d d! \cdot f(l)}{(1 - f(l))^d}\right)$$

and f *is defined as* $f(x) := \mathcal{M}_{\max\{\nu,1/2\}}(\kappa x)$.

Remark 3.1. It is important to evaluate the error on D because it is where equation (5a) is solved. As already mentioned, only the evaluation of the Matérn field on D is influential on the final solution q. With the above hypotheses on the domains, we conclude that G must include D and, in addition, they cannot coincide.

Remark 3.2. For this result, the domain G is needed to have a tensor-product geometry (i.e. it must be an hypercube $(0, L)^d$) in order to easily apply spectral methods to prove the error bound. However, this is usually not strictly required and what is instead the key point of the hypothesis is the $\delta/2$ distance from D.

Remark 3.3. Observing the error bound, we can see that A is in fact defined as a constant because it depends only on the Matérn parameters and the geometry of D. What is important instead is the relation with the geometry of G. This is expressed by the last term that depends on δ and shows how the error decreases exponentially with respect to the distance between D and G. Therefore, we conclude that the error is bounded and exponentially decreases as we increase the size of $G \supsetneq D$. In addition, this result motivates the standard choice of $\delta/2 \geq \lambda$ (e.g. in [22]) to have negligible truncation errors.

3.2 Finite element discretization

In this research we work on a finite element space that is general enough but, on the other hand, is restricted to effective practical considerations. Therefore, as a first observation, our assumptions do not accept an unbounded domain that is not feasible for implementations. As discussed in section 3.1, we therefore set our problem in $D \subset G \subsetneq \mathbb{R}^d$ and we then refer to problem (5). Moreover, we require the domains D and G to be with Lipschitz border and k to be integer. For the grid, we assume quasi-uniform triangulations $\mathcal{T}_h^D, \mathcal{T}_h^G$ that are represented by the average element edge $h > 0$. To conclude, we refer to the standard finite element spaces with continuous functions that are piece-wise polynomials. Explicitly:

$$\begin{aligned} V_h &= \{q_h \in C^0(\bar{D}) : q_h|_e \in \mathcal{P}^p(e) \quad \forall e \in \mathcal{T}_h^D\}, \\ W_h &= \{u_h \in C^0(\bar{G}) : u_h|_e \in \mathcal{P}^q(e) \quad \forall e \in \mathcal{T}_h^G\}, \end{aligned} \tag{6}$$

where $\mathcal{P}^p(e)$ indicates the space of polynomials of degree at most $p \in \mathbb{N}$ over the element e. At this point, we could set the same polynomial degree $p = q$ for the two spaces and hence

construct $\mathcal{T}_h^D$ to be a subset of $\mathcal{T}_h^G$. This choice would computationally save memory and time. On the other hand, there is no theoretical need to require it as an assumption. We will assume instead the same polynomial degree p to simplify calculations and notations.
As a last observation, we recall here the main property of the white noise duality on the basis functions. If $\{\varphi_i\}_{i=1}^{N_h}$ represents the set of basis functions for a certain finite element space of dimension N_h, then:

$$\{< \dot{W}, \varphi_i >\}_{i=1}^{N_h} \sim \mathcal{N}(\mathbf{0}, M),$$

i.e. the duality vector is distributed as a zero-mean normal vector with covariance matrix equivalent to the finite element mass matrix $M \in \mathbb{R}^{N_h \times N_h}$ (i.e. the matrix defined as $M_{i,j} :=< \varphi_i, \varphi_j >_{L^2}$). This means that white noise has a very simple finite element representation. We can now proceed to the Galerkin formulation. We define N_h^D, N_h^G as the dimensions of V_h and W_h. Moreover, we define as $\{\varphi_i\}_{i=1\dots N_h^D}$ and $\{\phi_i\}_{i=1\dots N_h^G}$ their basis functions. For the diffusion equation we have the usual linear system:

$$K_{u_h}\mathbf{q}_h = \mathbf{f},$$

where $\mathbf{f}$ is the finite element vector of the known term f, $\mathbf{q}_h$ is the unknown finite element vector representing the discretized solution q_h and K_{u_h} is the stiffness matrix obtained from the discretized Matérn realization, i.e.:

$$\{K_{u_h}\}_{i,j} := \int_D e^{u_h(\mathbf{x},w)} \nabla\varphi_i \cdot \nabla\varphi_j \, \mathrm{d}\mathbf{x}.$$

For the Whittle equation, we firstly consider $k = 1$ in order to have a linear equation. This choice yields the following simple linear system:

$$\left(I_{N_h^G} - \kappa^{-2}K\right)\mathbf{u}_h = \eta\mathbf{Z}_M,$$

where $I_{N_h^G}$ is the identity matrix of size $N_h^G \times N_h^G$, $\mathbf{u}_h$ is the FEM vector of the discretized solution, $\mathbf{Z}_M$ is a normal vector distributed according to $\mathcal{N}(\mathbf{0}, M)$ and K is the standard stiffness matrix, i.e.:

$$\{K\}_{i,j} := \int_G \nabla\phi_i \cdot \nabla\phi_j \, \mathrm{d}\mathbf{x}.$$

For a general $k \in \mathbb{N}$, one can solve the following recursive system ([22], [12]):

$$\begin{cases} u_1(\mathbf{x},\omega) - \kappa^{-2}\Delta u_1(\mathbf{x},\omega) = \eta\dot{W}(\mathbf{x},\omega) & \mathbf{x} \in G, \omega \in \Omega \\ u_{i+1}(\mathbf{x},\omega) - \kappa^{-2}\Delta u_{i+1}(\mathbf{x},\omega) = u_i(\mathbf{x},\omega) & \mathbf{x} \in G, \omega \in \Omega, \forall i = 0 \dots k-1 \\ u_i(\mathbf{x},\omega) = 0 & \mathbf{x} \in \partial G, \omega \in \Omega, \forall i = 1, \dots k \end{cases}$$

For a non-integer k, instead, the system to solve is not so simple to write because the main obstacle is the fractional derivative defined through the Fourier transform ([22]). On the other hand, different numerical methods have been recently formulated to solve deterministic fractional Laplacian problems (see for instance [23] and [1]). A specific version for our SPDE equation (5b) is instead the least-square approximation from [22] in the case of $2k \in \mathbb{N}$. Moreover, [4] presents a numerical scheme to solve equation (5b) with non-integer k using rational expansions of x^{-k}. However, in order to work with standard FEM approximations, we restrict ourself to integer exponents. We refer the interested reader to [23], [1] and [4] for further work.

3.3 Monte Carlo method

In this section and section 3.4 we aim to recall the definitions and main properties of the standard and multilevel Monte Carlo methods which represent respectively the independent and coupled sampling techniques. Indeed, these methods permit to deal with the stochastic nature of the problem returning an estimation of the properties of the solutions. The main difference is that MLMC is much more efficient but requires correlated solutions from different grid refinements.
We recall now the general definition for the standard Monte Carlo method as described in [21]. Suppose $X \sim \mathcal{D}$ is a random variable following an unknown probability distribution $\mathcal{D}$ and P is a functional of X. Finally, let $\{X_i\}_{i=1}^N$ be i.i.d. realizations of $X \sim \mathcal{D}$. Therefore, the Monte Carlo estimation for $P(X)$ is:

$$P^{MC,N} := \frac{1}{N}\sum_{i=1}^{N} P(X_i).$$

The main advantages of this method are its simplicity and generality. On the other hand, convergence is quite slow since:

$$\text{MSE}(P^{MC,N}, \mathbb{E}[P(X)]) := \mathbb{E}\left[\left(P^{MC,N} - \mathbb{E}[P(X)]\right)^2\right] = \mathbb{V}[P^{MC,N}] = \frac{1}{N}\mathbb{V}[P(X)], \tag{7}$$

which means that the mean square error order is $\mathcal{O}(N^{-1})$.
Alternative methods have been developed to have a faster convergence such as Variance Reduction techniques and Quasi Monte Carlo methods (see for instance [21], [7]).
To use this method when solving differential equations, for instance equation (1), it is just needed to compute the sampling mean of the functional evaluated in the set of numerical solutions $\{q_{h,i}\}_{i=1}^N$, where each solution is obtained solving the equation with a different and independent sampling of $a(\mathbf{x}, \omega)$.

3.4 Multilevel Monte Carlo method

One powerful technique to improve the standard Monte Carlo method when solving differential equations is the multilevel version that exploits the correlations of sampling between different grid levels. In this section, we refer to the original definitions in [15], [16].
MLMC relies on a sequence of approximations $q_{h_0}, \dots q_{h_L}$ on a set of grids with mesh sizes $h_0 > \cdots > h_L > 0$. Following the approach of [15], we can estimate $\mathbb{E}[P(q)]$ from $\mathbb{E}[P(q_{h_L})]$ which we rewrite with a telescopic sum as follows:

$$\mathbb{E}[P(q_{h_L})] = \mathbb{E}[P(q_{h_0})] + \sum_{l=1}^{L} \mathbb{E}[P(q_{h_l}) - P(q_{h_{l-1}})]$$

From the last equality, we can infer an unbiased estimator of $\mathbb{E}[q_{h_L}]$ defined as:

$$P_{h_L}^{MLMC} := \sum_{l=0}^{L} P_{h_l} \tag{8}$$

where

$$P_{h_l} = \begin{cases} \frac{1}{N_0}\sum_{i=1}^{N_0} P(q_{h_0}^{(i)}) & l = 0 \\ \frac{1}{N_l}\sum_{i=1}^{N_l}\left(P(q_{h_l}^{(i)}) - P(q_{h_{l-1}}^{(i)})\right) & l > 0 \end{cases}$$

In this case, $q_{h_l}^{(i)}$ indicates the numerical solution at discretization level h_l obtained from the i-th sampling of the stochastic terms. In other words, the estimator is composed by numerical solutions at two consecutive grid sizes obtained from the same random sampling. This is opposite to the standard Monte Carlo where all solutions are sampled independently. In addition, it can be challenging to define the same stochastic sampling (in our case, the $\mathbf{Z}_M$ vector) on two different grids. We postpone the solutions to these numerical challenges to section 5 and we now focus on the benefits of this method. The main result for the MLMC follows:

Theorem 3.2 (MLMC complexity ([9]))**.** *Suppose there exist positive constants α, β, c_1, c_2 and c_3 such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and:*

1. $\mathbb{E}[P(q_{h_l}) - P(q)] \leq c_1 h_l^{\alpha}$
2. $\mathbb{V}[q_{h_l} - q_{h_{l-1}}] \leq c_2 h_l^{\beta}$
3. *The cost C_l for a single sample $P\left(q_{h_l}^{(i)}\right) - P\left(q_{h_{l-1}}^{(i)}\right)$ is such that $C_l \leq c_3 h_l^{\gamma}$*

Then, there exists a positive constant c_4 such that for every $\epsilon < e^{-1}$ there are values of L and N_l to make the error of the MLMC estimator in (8) bounded by

$$MSE\left(P_{h_L}^{MLMC}, \mathbb{E}[P(q)]\right) := \mathbb{E}\left[\left(P_{h_L}^{MLMC} - \mathbb{E}[P(q)]\right)^2\right] \leq \epsilon^2$$

and its computational complexity bounded by

$$C \leq \begin{cases} c_4 \epsilon^{-2}, & \beta > \gamma, \\ c_4 \epsilon^{-2} (\log \epsilon)^2, & \beta = \gamma, \\ c_4 \epsilon^{-2-(\gamma-\beta)/\alpha}, & 0 < \beta < \gamma. \end{cases}$$

This result is crucial in many problems where, for a fixed computational cost, the estimation can be much more accurate than the estimations from the standard Monte Carlo sampling method. The following example makes it evident.

Example 3.1. *Consider the following standard linear SDE:*

$$dX(t) = a(t, X)dt + b(t, X)dW(t), \quad 0 < t < T.$$

We aim to solve it with an Euler-Maruyama scheme (step h) coupled with a standard Monte Carlo method (N realizations) to estimate $\mathbb{E}[X(T)]$. We can as usual decompose the MSE in the bias and variance terms. The first term can be bounded because the weak error order of the Euler-Maruyama scheme is known to be 1 (see for instance [20]). The variance term can instead be bounded using (7). We finally get that $\exists c_1, c_2 > 0$ such that

$$MSE(X_N^{MC}, \mathbb{E}[X(T)]) \leq c_1 h^2 + c_2 \frac{1}{N}$$

If we impose $MSE = \mathcal{O}(\epsilon^2)$ we require $h = \mathcal{O}(\epsilon)$ and $N = \mathcal{O}(\epsilon^{-2})$. This choice inevitably implies a total cost of $C = \mathcal{O}(N/h) = \mathcal{O}(\epsilon^{-3})$.
On the other hand, the adoption of a MLMC method improves the cost since Theorem 3.2 holds for $\alpha = 1$ (because the weak error order is 1), $\beta = 1$ (as a consequence of the fact that the strong error order is $1/2$) and $\gamma = 1$ (Euler-Maruyama cost). Therefore, the total cost is $C = \mathcal{O}(\epsilon^{-2}(\log \epsilon)^2)$ that reduces by almost 1 order in ϵ the cost induced by the standard method.

We conclude this theoretical recollection anticipating that Theorem 3.2 represents the main goal for the next sections where we aim to use MLMC's potentiality for our problem taking care of respecting the assumptions 1,2,3 in Theorem 3.2.

4 Study on the single level total error

The goal of this section is the achievement of important results regarding the numerical solution of the problem in (5) using a single level sampling technique as the standard Monte Carlo method. This is motivated by three different reasons:

- To present results that could be useful in the case one prefers or is obliged to pass to a single level sampling method (e.g. when just one grid is available or it is demanding to generate correlated solutions).
- To gradually increase the level of difficulty, starting with the error analysis of a problem that is much easier than the multilevel one.
- To pursue results that can be used to guarantee the assumption 1 in Theorem 3.2. Namely, the fulfilment of such assumption for a certain parameter α permits to assure the convergence and, in addition, returns information about the level of accuracy/cost trade-off for the MLMC applied to our problem.

4.1 Some classical preliminaries

We recall here some classical results that will be used to achieve error bounds in the following sections. Let us start with the classical finite element a priori bound.

Theorem 4.1 (FEM a priori bound ([27]))**.** *Consider the resolution of the following elliptic problem over a bounded domain $B \subset \mathbb{R}^d$:*

$$\text{Find } \phi \in V : a(\phi, v) = F(v) \ \ \forall v \in V, \tag{9}$$

where V is a Hilbert space which is subspace of $H^1(B)$, $a(\cdot,\cdot)$ is a continuous, coercive and bilinear form and $F(\cdot)$ is a linear and bounded functional. Hence, the Lax-Milgram theorem guarantees the existence and uniqueness of the solution. Consider now a regular triangulation $\mathcal{T}_h$ of the domain over which we solve the problem with a finite element method of order r (as presented in section 3.2). Let $\phi \in V$ the solution of (9) and $\phi_h \in V_h$ the solution of the FEM method applied to (9). If $\phi \in H^{r+1}(B)$, then:

$$\|\phi - \phi_h\|_{H^1(B)} \leq \frac{M}{\alpha} C h^r |\phi|_{H^{r+1}(B)}, \tag{10}$$

$$\|\phi - \phi_h\|_{L^2(B)} \leq \frac{M}{\alpha} C h^{r+1} |\phi|_{H^{r+1}(B)} \tag{11}$$

for a certain constant $C > 0$ independent of h and ϕ and where M, α are the continuity and coercive constants of $a(\cdot,\cdot)$. $|\cdot|_{H^r(B)}$ is instead the seminorm of $H^r(B)$.

Remark 4.1. This is certainly a basic result we can use to achieve an a priori bound for our problem. The main issue, we will see, is how to validate the condition that $\phi \in H^{r+1}(B)$.

Consider the following weak problem on a bounded domain $B \subset \mathbb{R}^d$ (elliptic equation in divergence form with homogeneous Dirichlet boundary conditions):

$$\text{find } \phi \in H_0^1(B) : \int_B a\nabla\phi \cdot \nabla v \, \mathrm{d}x = \int_B fv \, \mathrm{d}x \quad \forall v \in H_0^1(B). \tag{12}$$

This is the general formulation of our equation in (1). In the following sections, the following result will also turn out to be fundamental.

Theorem 4.2 (Elliptic stability under diffusivity coefficients)**.** *Suppose we aim to solve the problem in (12) with two different diffusivity parameters, i.e.:*

$$\textit{Find } \phi_i \in H_0^1(B) : \int_B a_i\nabla\phi_i \cdot \nabla v \, dx = \int_B fv \, dx \quad \forall v \in H_0^1(B), \quad i = 1, 2.$$

If $f \in L^2(B)$, $a_i \in L^\infty(B)$ *and* $\min_{\mathbf{x}\in B} a_i(\mathbf{x}) > 0$ *for* $i = 1, 2$, *then* ϕ_1, ϕ_2 *exist and are unique. Furthermore:*

$$|\phi_1 - \phi_2|_{H^1(B)} \leq \|a_1 - a_2\|_{L^\infty(B)} \frac{\|f\|_{L^2(B)}}{\min_{\mathbf{x}\in B} a_1(\mathbf{x}) \cdot \min_{\mathbf{x}\in B} a_2(\mathbf{x})}.$$

If, in addition, $a_i \in L^2(B)$ *for* $i = 1, 2$ *and* $\phi_2 \in W^{1,\infty}(B)$ *, then:*

$$|\phi_1 - \phi_2|_{H^1(B)} \leq \|a_1 - a_2\|_{L^2(B)} \frac{\|\phi_2\|_{W^{1,\infty}(B)}}{\min_{\mathbf{x}\in B} a_1(\mathbf{x})}.$$

Remark 4.2. This is an important result on stability when the diffusivity coefficient a varies. Under certain regularity assumptions, it guarantees that the two different solutions get closer when the two coefficients do. This represents a useful result for our problem where the original diffusivity coefficient is $e^{u(x,\omega)}$ is replaced by the numerical solution $e^{u_h(x,\omega)}$. Therefore, to make a comparison between the analytical and the numerical solution of the original diffusion equation (1), we firstly need to compare the two different diffusivity coefficients.

Remark 4.3. The second inequality can obviously be rewritten with ϕ_1 instead of ϕ_2 and a_2 instead of a_1 because of the complete symmetry when switching the index i.

Proof. Let us firstly define $< \cdot, \cdot >$ as the standard $L^2(B)$ scalar product to simplify the notation. Now, the problem can be stated as follows:

$$\begin{cases} < a_1\nabla\phi_1, \nabla v > = < f, v > & \forall v \in H_0^1(B), \\ < a_2\nabla\phi_2, \nabla v > = < f, v > & \forall v \in H_0^1(B). \end{cases}$$

First of all, we notice that the two bilinear forms have continuity constants $\|\phi_i\|_{L^\infty} < \infty$, $i = 1, 2$. In addition, their coercivity constants are equivalent to $\min_{\mathbf{x}\in B} a_i(\mathbf{x}) > 0$ for $i = 1, 2$ by assumption. Moreover, also the functional on the right hand side is bounded by $\|f\|_{L^2(B)} < \infty$. Therefore, the hypotheses of the Lax-Milgram theorem holds and existence and uniqueness of

the solutions are guaranteed.
We set now $v = e := \phi_1 - \phi_2$ and we subtract the equations:

$$\begin{aligned} & < a_1\nabla\phi_1 - a_2\nabla\phi_2, \nabla e > \; = \; 0 \\ \Rightarrow \quad & < a_1\nabla\phi_1 - a_1\nabla\phi_2, \nabla e > + < a_1\nabla\phi_2 - a_2\nabla\phi_2, \nabla e > \; = \; 0 \\ \Rightarrow \quad & < a_1\nabla e, \nabla e > \; = \; - < (a_1 - a_2)\nabla\phi_2, \nabla e > \\ \Rightarrow \quad & \min_{\mathbf{x}\in B} a_1(\mathbf{x}) \cdot |e|^2_{H^1(B)} \overset{\text{Hölder}}{\leq} \|a_1 - a_2\|_{L^\infty(B)} |\phi_2|_{H^1(B)} |e|_{H^1(B)} \\ & \qquad\qquad \leq \; \|a_1 - a_2\|_{L^\infty(B)} \frac{\|f\|_{L^2(B)}}{\min_{\mathbf{x}\in B} a_2(\mathbf{x})} |e|_{H^1(B)}, \end{aligned}$$

where the last inequality holds because of the continuity of the solution norm from the Lax-Milgram theorem. So that, we obtain:

$$|e|_{H^1(B)} \leq \|a_1 - a_2\|_{L^\infty(B)} \frac{\|f\|_{L^2(B)}}{\min_{\mathbf{x}\in B} a_1(\mathbf{x}) \min_{\mathbf{x}\in B} a_2(\mathbf{x})}.$$

The second result can be easily obtained if we use Hölder inequality in a different way. More precisely, we can bound the right hand side term with:

$$< (a_1 - a_2)\nabla\phi_2, \nabla e > \; \leq \; \|a_1 - a_2\|_{L^2(B)} \|\nabla\phi_2\|_{L^\infty(B)} |e|_{H^1(B)}.$$

□

A last result we will partially use for the study of the posteriori bound is the following.

Theorem 4.3 (Error indicators for elliptic problems ([6], [27]))**.** *Suppose we aim to solve the same elliptic problem as in (12) where the hypotheses for existence and uniqueness hold and $a \in L^\infty(B)$ is piecewise smooth on a non-degenerate mesh $\mathcal{T}_h$. If $\phi \in V$ is the exact solution of (12) and $\phi_h \in V_h$ is the finite element solution, then:*

$$|\phi - \phi_h|_{H^1(B)} \leq \frac{\beta}{\min_{\mathbf{x}\in B} a(\mathbf{x})} \left(\sum_{e\in\mathcal{T}_h} [\mathcal{E}_e(\phi_h)]^2 \right)^{1/2},$$

where β is a constant that only depends on the non-degeneracy of the mesh and $\mathcal{E}_e$ is the error indicator defined as:

$$\mathcal{E}_e(\phi_h) := h_e \|f + \nabla\cdot(a\nabla\phi_h)\|_{L^2(e)} + \frac{1}{2} h_e^{1/2} \|[a\mathbf{n}\cdot\nabla\phi_h]_{\mathbf{n}}\|_{L^2(\partial e)},$$

where $\mathbf{n}$ is the normal vector to the surface ∂e, h_e is the size of the element $e \in \mathcal{T}_h$ and $[\cdot]_{\mathbf{n}}$ is the discontinuity between adjacent elements along the $\mathbf{n}$ direction.

Remark 4.4. The formula for $\mathcal{E}_e$ can be interpreted in the following way: the first part is the residual to the equation in the strong form and the second represents the discontinuity between adjacent elements.

Remark 4.5. This result is basic and general since it can be adapted to multiple linear differential equations. However, we anticipate that the assumption of $f \in L^2(\Omega)$ is not suitable for equation (3) where the sampled white noise can only be treated as a functional.

4.2 Definition of the solution error

We define in this section the error we aim to control in order to assure convergence and regularity of the numerical solution. We remind that, for now, we study the approximation error for a solution obtained from the finite element method as defined in section 3.2 and a standard Monte Carlo method as defined in section 3.3. Let q the analytical stochastic solution of (5), q_h the numerical stochastic solution obtained with a mesh size h and $q_h^{MC,N}$ the N-samples Monte Carlo estimation of q_h. A possible straightforward choice is a slight generalization of the Mean-Squared Error (MSE) already defined in section 3.3:

$$MSE_{L^2(D)}(q_h^{MC,N}, \mathbb{E}[q]) := \mathbb{E}\left[\|q_h^{MC,N} - \mathbb{E}[q]\|^2_{L^2(D)}\right].$$

Simple calculations show that a variance-bias decomposition holds also with this definition, indeed:

$$\begin{aligned} MSE_{L^2(D)}(q_h^{MC,N}, \mathbb{E}[q]) &= \mathbb{E}\Big[\int_D \left(q_h^{MC,N} - \mathbb{E}[q_h]\right)^2 dx \\ &\quad + \int_D \left(\mathbb{E}[q_h] - \mathbb{E}[q]\right)^2 dx \\ &\quad + \int_D 2(q_h^{MC,N} - \mathbb{E}[q_h])(\mathbb{E}[q_h] - \mathbb{E}[q])\, dx\Big] \\ &= \int_D \mathbb{V}[q_h^{MC,N}]\, dx + \int_D \left(\mathbb{E}[q_h - q]\right)^2 dx + 0 \\ &= \overbrace{\frac{||\mathbb{V}[q_h]||_{L^1(D)}}{N}}^{Variance} + \overbrace{||\mathbb{E}[q_h - q]||^2_{L^2(D)}}^{Bias}. \end{aligned}$$

Previously, we used the Fubini's theorem.
Equivalently, one can get:

$$MSE_{H^1(D)}(q_h^{MC,N}, \mathbb{E}[q]) := \mathbb{E}\left[|q_h^{MC,N} - \mathbb{E}[q]|^2_{H^1(D)}\right] = \overbrace{\frac{\sum_{i=1}^d ||\mathbb{V}[\partial_i q_h]||_{L^1(D)}}{N}}^{Variance} + \overbrace{|\mathbb{E}[q_h - q]|^2_{H^1(D)}}^{Bias}$$

under some regularity assumptions, where $\partial_i q_h$ indicates the partial derivative and $|\cdot|_{H^1(D)}$ the $H^1(D)$ seminorm. Henceforth, the MSE with the H^1 seminorm will be preferred because the latter represents the space where to set the weak formulation of problem (5).
The main advantage of such a decomposition is the following: this splitting separates the part that can be improved with a greater number of samples from the bias that instead only depends on the accuracy of the numerical solution. In other words, we can formally analyse the bias term without considering the sampling error due to the Monte Carlo estimation.
From these observations, assuming the variance of q_h to be approximately constant for small h, we can therefore proceed to analyse only the second part of the error that is strictly related to the finite element approximation. At this point, we note that this component is an L^2/H^1-generalization of the weak error between the analytical solution and its stochastic numerical approximation. Furthermore, the following result holds thanks to a double application of the Jensen's inequality:

$$|\mathbb{E}[q_h - q]|^2_{H^1(D)} \leq \mathbb{E}\left[|q_h - q|_{H^1(D)}\right]^2 \leq \mathbb{E}\left[|q_h - q|^2_{H^1(D)}\right], \tag{13}$$

that is a bound of the generalized weak error with the associated generalized strong error.
In this section, we motivated the choice of the generalized weak error as the main object to study. Indeed, it represents the main numerical influence in the MSE total error while the remaining part is only related to the precision of the sampling method. This latter quantity decreases as N^{-1} and does not require necessary further studies. In addition, possible bounds for the strong error hold also for the weak error because of (13) and therefore one can potentially work with all the three definitions of error shown in (13). Finally, the weak error also reflects the hypothesis 1 of Theorem 3.2 and, so, the achievement of explicit bounds may allow us to derive results on the MLMC efficiency.

4.3 A priori bound

In this section, we aim to exploit the above mentioned results (section 4.1) to get an explicit a priori bound to the weak error. Since we deal with a finite element discretization applied to two coupled problems (respectively, (1) and (3) that give (5)), it could be useful to present the following nomenclature.

Definition 1. For the error analysis of problem (5), we define:

i) $V := H_0^1(D), W := H_0^1(G)$, i.e. the reference spaces for the solutions of the problem in (5). Moreover, we define $\|\cdot\|_V$ as the $H^1(D)$ seminorm.

ii) V_h, W_h are the associated finite element spaces as defined in (6) for a certain polynomial degree $p \in \mathbb{N}$.

iii) $q(\cdot,\omega) \in V, u(\cdot,\omega) \in W$ represent the analytical solutions to problem (5).

iv) $u_h(\cdot,\omega) \in W_h$ is the finite element solution of the equations (5b) and (5d).

v) $\hat{q}_h(\cdot,\omega) \in V$ is the analytical solution of (5a) and (5c) when one takes u_h instead of the analytical solution u.

vi) $q_h(\cdot,\omega) \in V_h$ is the fully numerical solution of (5).

The $(\cdot,\omega)$ notation is needed to indicate that these solutions are stochastic fields that, $\mathbb{P}$ almost-every $\omega \in \Omega$, are functions belonging to the specified spaces.
The reason of considering the semi-discrete solution $\hat{q}_h$ is based on the following error decomposition based on the triangular inequality:

$$\mathbb{E}[\|q_h - q\|_V^2] \leq \mathbb{E}[\|q_h - \hat{q}_h\|_V^2] + \mathbb{E}[\|\hat{q}_h - q\|_V^2] \tag{14}$$

It is useful to separate the two components since the first error represents the discretization error due to $\mathcal{T}_h^D$ while the second is the discretization due to $\mathcal{T}_h^G$. Henceforth, we will work on the two components separately.
A last result we report before the achievement of the a priori bound is a regularity property of two terms that will represent the boundedness and coercivity constants of problem (5).

Proposition 4.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\gamma : \mathbb{R}^d \times \Omega \to \mathbb{R}$ be a Gaussian field and $B \subset \mathbb{R}^d$. Hence:*

$$e^{\|u(\cdot,\omega)\|_{L^\infty(B)}}, e^{-\|u(\cdot,\omega)\|_{L^\infty(B)}} \in L^r(\Omega) \quad \forall\, 0 < r < \infty$$

Proof. Similar to the proof of Proposition 2.2 in [8]. It is a simple application of Fernique's theorem ([13]).

□

We can now give the first significant result related to our problem.

Proposition 4.2. *Let $q_h \in V_h, \hat{q}_h \in V$ and $u_h \in W_h$ be defined as in Definition 1. Let the finite element discretization $\mathcal{T}_h^D$ on D have a grid size $h > 0$ and the V_h space composed by polynomials of degree at most $p \in N$. Assume $\hat{q}_h \in L^4(\Omega, H^{p+1}(D))$. Then:*

$$\mathbb{E}\left[\|q_h - \hat{q}_h\|_V^2\right] \leq \mathbb{E}\left[e^{8\|u_h(\cdot,\omega)\|_{L^\infty(D)}}\right]^{\frac{1}{2}} \mathbb{E}\left[\|\hat{q}_h^{(p+1)}\|_{L^2(D)}^4\right]^{\frac{1}{2}} h^{2p},$$
$$\mathbb{E}\left[\|q_h - \hat{q}_h\|_{L^2(D)}^2\right] \leq \mathbb{E}\left[e^{8\|u_h(\cdot,\omega)\|_{L^\infty(D)}}\right]^{\frac{1}{2}} \mathbb{E}\left[\|\hat{q}_h^{(p+1)}\|_{L^2(D)}^4\right]^{\frac{1}{2}} h^{2p+2}.$$

Proof. This results is a direct application of Theorem 4.1. Suppose for now to fix $\omega \in \Omega$ as to analyse the solution for a single realization. $q_h(\cdot,\omega)$ and $\hat{q}_h(\cdot,\omega)$ represent respectively the numerical and analytical solution of the same problem:

$$\begin{cases} -\nabla \cdot \left(e^{u_h(x,\omega)} \nabla \hat{q}_h(x,\omega)\right) = f(x) & x \in D, \omega \in \Omega, \\ \hat{q}_h(x,\omega) = 0 & x \in \partial D, \omega \in \Omega. \end{cases}$$

Therefore, we can apply Theorem 4.1 if we guarantee the boundedness and coercivity of the bilinear form. As already discussed in the proof of Theorem 4.2, the boundedness and coercivity constants are respectively

$$\begin{aligned} M &= \max_{\mathbf{x} \in D} \left(e^{u_h(\mathbf{x},\omega)}\right) = e^{\max_{\mathbf{x}\in D}(u_h(\mathbf{x},\omega))} \leq e^{\|u_h(\cdot,\omega)\|_{L^\infty(D)}}, \\ \alpha &= \min_{\mathbf{x} \in D} \left(e^{u_h(\mathbf{x},\omega)}\right) = e^{\min_{\mathbf{x}\in D}(u_h(\mathbf{x},\omega))} \geq e^{-\|u_h(\cdot,\omega)\|_{L^\infty(D)}}. \end{aligned} \tag{15}$$

Hence, Theorem 4.1 holds for the single realization and gives:

$$\begin{aligned} &\|q_h(\cdot,\omega) - \hat{q}_h(\cdot,\omega)\|_V \leq e^{2\|u_h(\cdot,\omega)\|_{L^\infty(D)}} \|\hat{q}_h^{(p+1)}\|_{L^2(D)} h^p, \\ \Rightarrow \quad &\mathbb{E}\left[\|q_h(\cdot,\omega) - \hat{q}_h(\cdot,\omega)\|_V^2\right] \leq \mathbb{E}\left[e^{4\|u_h(\cdot,\omega)\|_{L^\infty(D)}} \|\hat{q}_h^{(p+1)}\|_{L^2(D)}^2\right] h^{2p} \\ &\qquad \leq \mathbb{E}\left[e^{8\|u_h(\cdot,\omega)\|_{L^\infty(D)}}\right]^{\frac{1}{2}} \mathbb{E}\left[\|\hat{q}_h^{(p+1)}\|_{L^2(D)}^4\right]^{\frac{1}{2}} h^{2p}, \end{aligned}$$

where in the last line we apply Cauchy-Schwartz inequality. The second bound can be obtained by repeating the previous steps using result (11) from Theorem 4.1.

□

Remark 4.6. This proposition provides an a priori bound for the first term of (14). It states that the error decreases with an order of h^p. A note is that this rate only depends on the discretization order in the diffusion problem (1) and not on the order of the Whittle equation (3) discretization.

Remark 4.7. The drawback is instead related to the strong assumption on $\hat{q}_h$ that is challenging to prove formally (even if in practical situations the property holds true). Another observation is related to the assumption that the constant $\mathbb{E}\left[e^{8\|u_h(\cdot,\omega)\|_{L^\infty(D)}}\right]$ is finite. This has not been formally proven yet but it is assumed to be true since the same quantity with u_h replaced by u is finite thanks to Proposition 4.1. In addition, the same quantity can be easily estimated through sampling methods as presented in section 3.3 and section 3.4.

We can now proceed with the second error term in (14).

Proposition 4.3. *Let $\hat{q}_h \in V, q \in V, u_h \in W_h$ and $u \in W$ be defined as in Definition 1. Let $f \in L^2(D)$ be as usual the forcing term in equation (5). We have that:*

$$\mathbb{E}\left[\|\hat{q}_h - q\|_V^2\right] \leq \|f\|_{L^2(D)}^2 \left(\mathbb{E}\left[\|u_h(\cdot,\omega) - u(\cdot,\omega)\|_{L^\infty(G)}^4\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[e^{4\gamma_u(\omega)}\right]\right)^{\frac{1}{2}}, \tag{16}$$

where:

$$\gamma_u(\omega) := \|u_h(\cdot,\omega)\|_{L^\infty(G)} + \|u(\cdot,\omega)\|_{L^\infty(G)} + \max\{\|u_h(\cdot,\omega)\|_{L^\infty(G)}, \|u(\cdot,\omega)\|_{L^\infty(G)}\}.$$

Proof. This bound is an application of Theorem 4.2. Indeed, q_h and $\hat{q}_h$ are the analytical solutions to the same problem but with different diffusivity coefficients, respectively u and u_h. As we did in the previous proof, we start by analysing the single realization $\omega \in \Omega$. Theorem 4.2 states that:

$$\|\hat{q}_h(\cdot,\omega) - q(\cdot,\omega)\|_V \leq \|e^{u_h(\cdot,\omega)} - e^{u(\cdot,\omega)}\|_{L^\infty(G)} \cdot \|f\|_{L^2(D)} \cdot e^{\|u_h(\cdot,\omega)\|_{L^\infty(G)}} \cdot e^{\|u(\cdot,\omega)\|_{L^\infty(G)}},$$

where the rightmost terms are again due to the inequalities in (15).
For the integral mean theorem, $\forall x, y \in \mathbb{R} \;\; \exists\, x \leq \xi \leq y$ such that

$$\int_x^y e^t\, dt = (y-x)e^\xi \quad \Rightarrow \quad |e^y - e^x| \leq |y-x| e^{\max\{x,y\}} \quad \forall x, y \in \mathbb{R}.$$

Applied to the first term of the bound:

$$\|e^{u_h(\cdot,\omega)} - e^{u(\cdot,\omega)}\|_{L^\infty(G)} \leq \|u_h(\cdot,\omega) - u(\cdot,\omega)\|_{L^\infty(G)} e^{\max\{\|u_h(\cdot,\omega)\|_{L^\infty(G)}, \|u(\cdot,\omega)\|_{L^\infty(G)}\}},$$

and hence, the second term can be bounded by:

$$\|q_h(\cdot,\omega) - q(\cdot,\omega)\|_V \leq \|u_h(\cdot,\omega) - u(\cdot,\omega)\|_{L^\infty(G)} \cdot \|f\|_{L^2(D)} \cdot e^{\gamma_u(\omega)},$$

where:

$$\gamma_u(\omega) := \|u_h(\cdot,\omega)\|_{L^\infty} + \|u(\cdot,\omega)\|_{L^\infty} + \max\{\|u_h(\cdot,\omega)\|_{L^\infty}, \|u(\cdot,\omega)\|_{L^\infty}\}.$$

Finally, we can apply the expectation $\mathbb{E}$ and we get:

$$\begin{aligned}
\mathbb{E}\left[\|q_h - q\|_V^2\right] \leq & \|f\|_{L^2(D)}^2 \cdot \mathbb{E}\left[\|u_h(\cdot,\omega) - u(\cdot,\omega)\|_{L^\infty(G)}^2 \cdot e^{2\gamma_u(\omega)}\right] \\
\leq & \|f\|_{L^2(D)}^2 \cdot \left(\mathbb{E}\left[\|u_h(\cdot,\omega) - u(\cdot,\omega)\|_{L^\infty(G)}^4\right]\right)^{\frac{1}{2}} \left(\mathbb{E}\left[e^{4\gamma(\omega)}\right]\right)^{\frac{1}{2}},
\end{aligned}$$

where in the last line we apply the Cauchy-Schwartz inequality.

□

Remark 4.8. Proposition 4.2 and 4.3 give bounds for the $L^2(\Omega, V)$ norms of the error but one can also easily achieve bounds for the the $L^1(\Omega, V)$: it is just needed not to square the left and right terms of the inequality before the application of $\mathbb{E}$.

Remark 4.9. We already know that the last term of the bound in (16) is a finite constant. Moreover, in practice, it can be approximated by $\gamma_u(\omega) \approx \hat{\gamma}_u(\omega) := 3\|u(\cdot,\omega)\|_{L^\infty(G)}$.

On the contrary, the term in the middle is known to be finite but requires a further study in order to achieve a complete a priori bound. For this purpose, we present here an important result from [11] that gives a bound in the Hölder norm.

Theorem 4.4 (Theorem 6.23 from [11]). *Let $u \in W$ and $u_h \in W_h$ be defined as in Definition 1, the finite element polynomial order $p = 1$ and $d = 1$. In addition, let $D \subset \mathbb{R}^d$ be a bounded polytope with Lipschitz border. If $0 < \lambda \leq 1/2$ is such that $2k > \lambda + 1/2$ and $\delta \in (0, \lambda)$, the following bound holds:*

$$\left(\mathbb{E}[\|u_h - u\|^r_{C^{0,\delta}(\bar{D})}]\right)^{1/r} \leq Ch^{\min\{2k-\lambda-1/2-\epsilon, 3/2-\lambda\}}$$

for a certain constant $C > 0$ that does not depend on h or the solution, and for every $r, \epsilon > 0$.

Remark 4.10. The original theorem dealt with "generalized" Matérn fields. Therefore, the statement in [11] requires further assumptions that are actually automatically satisfied by our standard Matérn field (in particular, about the regularity of the differential operator in (5b)). In addition, it is valid for $k \in \mathbb{R}$ despite we initially restricted ourself to $k \in \mathbb{N}$ in order to work with only standard finite element methods. In the case of non-integer k, the theorem holds when equation (5b) is solved using a sinc-quadrature method presented in [5].

Remark 4.11. Despite its generality under the view of finite element spaces and Matérn fields, the main drawback is the assumption of $p = d = 1$. It is presumable that an L^∞ bound works also in higher dimensions and higher polynomial orders but, however, it has not been proved yet.

We can finally present the full a priori bound.

Theorem 4.5 (A priori bound). *Let the definitions and assumptions of Proposition 4.2, Proposition 4.3 and Theorem 4.4 hold for a grid size $h > 0$, polynomial order $p = 1$, dimension $d = 1$ and $k > 1/4$. Then, there exists a constant $C > 0$ independent from h such that:*

$$\mathbb{E}[\|q_h - q\|^2_V] \leq Ch^{\min\{4k-1-\epsilon, 2\}} \tag{17}$$

for every $\epsilon > 0$ small enough.

Proof. We are required to split the error in two parts as in (14). Then, we can apply Proposition 4.2 on the first part and Proposition 4.3 on the second part. From the second part of the error, we can use Theorem 4.4 to bound the following term:

$$\left(\mathbb{E}\left[\|u_h(\cdot,\omega) - u(\cdot,\omega)\|^4_{L^\infty(G)}\right]\right)^{\frac{1}{2}}.$$

More precisely, we apply the limit $\delta \to 0$ and we set $r = 4$. Therefore, the error is bounded by:

$$\mathbb{E}[\|q_h - q\|^2_V] \leq C_1 h^2 + C_2 h^{\min\{4k-2\lambda-1-\epsilon, 3-2\lambda\}}.$$

At this point, we can set $2\delta = \lambda < \epsilon$. The condition $\delta \in (0, \lambda)$ is so automatically satisfied and the new bound can be rewritten as:

$$\mathbb{E}[\|q_h - q\|^2_V] \leq C_1 h^2 + C_2 h^{\min\{4k-2\lambda-1-\epsilon, 3-2\lambda\}} \leq C_1 h^2 + C_2 h^{\min\{4k-1-3\epsilon, 3-2\epsilon\}}$$

under the condition of $k > \epsilon + 1/4$ and $\forall \epsilon \in (0, \frac{1}{2})$. In addition, one can easily show that the condition on k can be relaxed with $k > 1/4$ for every ϵ. To conclude, for ϵ small enough, the

$(3-2\epsilon)$ exponent is no more effective if compared to h^2 on the left. Therefore, we obtain a final bound similar to equation (17).

□

Remark 4.12. Theorem 4.5 gives an explicit rate to the a priori bound. This depends on the parameter k of the Matérn field. Indeed, we have already discussed in section 3.2 that the fractional exponent k has an important role in the numerical discretization of equation (5b) and, therefore, also the finite element method convergence rates depend on it.

At this point, one can investigate how the strong error order explicitly depends on k.

Corollary 4.1 (Order of the H^1 strong error)**.** *Let the same definitions and assumptions of Theorem 4.5 hold. Then:*

$$\left(\mathbb{E}[\|q_h - q\|_V^2]\right)^{1/2} = \begin{cases} \mathcal{O}(h^{2k-1/2-\epsilon}) & k \in (\frac{1}{4}, \frac{3}{4}],\ \epsilon > 0, \\ O(h) & k \in (\frac{3}{4}, \infty). \end{cases}$$

Proof. Simple calculations from the exponent in equation (17) yield the result.

□

Remark 4.13. As mentioned before, the fractional exponent k has a specific role in the convergence rate. In particular, a smaller k decreases the regularity of the solution and worsens the convergence rate.

4.4 Studies on the a posteriori bound

After the achievement of an explicit expression for the a priori bound of the discretization error, we present in this section what has been partially developed to get an a posteriori bound of the same quantity.
As for section 4.3, we start from the first part of error in (14).

Proposition 4.4. *Let $q_h \in V_h$, $\hat{q}_h \in V$, $u_h \in W_h$ be defined as in Definition 1. Assume that $\mathcal{T}_h^D \supset \mathcal{T}_h^G|_D$. Therefore:*

$$\mathbb{E}[||\hat{q}_h - q_h||_V] \leq \beta \left(\mathbb{E}\left[e^{2\|u_h(\cdot,\omega)\|_{L^\infty(G)}}\right]\right)^{\frac{1}{2}} \left(\sum_{e\in\mathcal{T}_h} \bar{\mathcal{E}}_{e,2}(q_h, u_h)\right)^{\frac{1}{2}},$$

where β is a constant that only depends on the non-degeneracy of the mesh and $\bar{\mathcal{E}}_{e,2}(q_h, u_h)$ is the second moment of $\mathcal{E}_e(q_h, u_h)$ that is defined as:

$$\mathcal{E}_e(q_h, u_h) := h_e \|f + \nabla \cdot e^{u_h(\mathbf{x},\omega)} \nabla q_h\|_{L^2(e)} + \frac{1}{2} h_e^{1/2} \left\| \left[\mathbf{n} \cdot e^{u_h(\mathbf{x},\omega)} \nabla q_h\right]_{\mathbf{n}} \right\|_{L^2(\partial e)},$$

where h_e is the size of the element, ∂e is its surface, $\mathbf{n}$ is the normal vector to ∂e and $|\cdot|_{\mathbf{n}}$ is the discontinuity between adjacent elements along the $\mathbf{n}$ direction.

Proof. We already discussed in the previous sections that:

$$\max_{\mathbf{x}\in G}\left(e^{u_h(\mathbf{x},\omega)}\right) \leq e^{\|u_h(\cdot,\omega)\|_{L^\infty(G)}} < \infty \quad \mathbb{P} \text{ a.e. } \omega \in \Omega$$

since it belongs to $L^r(\Omega)\ \forall r > 0$ thanks to Proposition 4.1. It follows that:

$$\min_{\mathbf{x}\in G}\left(e^{u_h(\mathbf{x},\omega)}\right) \geq e^{-\|u_h(\cdot,\omega)\|_{L^\infty(G)}} > 0 \quad \mathbb{P} \text{ a.e. } \omega \in \Omega.$$

In addition, the diffusivity coefficient $e^{u_h(\mathbf{x},\omega)}$ is clearly smooth piece-wise because it belongs to W_h and we have the hypothesis that $\mathcal{T}_h^D \supset \mathcal{T}_h^G|_D$. Hence, we can take a single realization $\omega \in \Omega$ and apply Theorem 4.3 to arrive at the following bound:

$$\begin{aligned}||\hat{q}_h - q_h||_V &\leq \frac{\beta}{\min_{\mathbf{x}\in G} e^{u_h(\mathbf{x},\omega)}.}\left(\sum_{K\in\mathcal{T}_h}[\mathcal{E}_e(q_h,u_h)]^2\right)^{\frac{1}{2}} \\ &\leq \beta\cdot e^{\|u_h(\cdot,\omega)\|_{L^\infty(G)}}\left(\sum_{K\in\mathcal{T}_h}[\mathcal{E}_e(q_h,u_h)]^2\right)^{\frac{1}{2}}.\end{aligned}$$

We can now apply the average operator and use the Cauchy-Schwartz inequality to achieve the expected bound:

$$\begin{aligned}\mathbb{E}[||\hat{q}_h - q_h||_V] &\leq \mathbb{E}\left[\beta\cdot e^{\|u_h(\cdot,\omega)\|_{L^\infty(G)}}\left(\sum_{K\in\mathcal{T}_h}[\mathcal{E}_e(q_h,u_h)]^2\right)^{\frac{1}{2}}\right] \\ &\leq \beta\cdot\mathbb{E}\left[e^{2\|u_h(\cdot,\omega)\|_{L^\infty(G)}}\right]^{\frac{1}{2}}\mathbb{E}\left[\left(\sum_{K\in\mathcal{T}_h}[\mathcal{E}_e(q_h,u_h)]^2\right)\right]^{\frac{1}{2}} \\ &= \beta\cdot\mathbb{E}\left[e^{2\|u_h(\cdot,\omega)\|_{L^\infty(G)}}\right]^{\frac{1}{2}}\left(\sum_{K\in\mathcal{T}_h}\bar{\mathcal{E}}_{e,2}(q_h,u_h)\right)^{\frac{1}{2}}.\end{aligned}$$

□

Remark 4.14. This result can clearly give some useful information to perform mesh adaptivity on D. A practical strategy could be the following: fixing the geometry of $\mathcal{T}_h^G$, one can consider the first two terms of the bound as constants while the averaged indicators $\bar{\mathcal{E}}_{e,2}$ can be estimated from different realizations and computations of (u_h, q_h) with a sampling technique such as the standard Monte Carlo method.

We are interested to get a similar result for the second part of error, i.e. $\|\hat{q}_h - q\|_V$. We have already seen that we can bound this quantity with the error of the generated Matérn field $\|u_h - u\|_{L^\infty(G)}$. Therefore, an a posteriori bound on this term would have two utilities: provide an a posteriori bound to the total error and return information to perform mesh adaptivity on G instead of D.
However, there are two main difficulties that make this objective challenging. First of all, we need to provide a bound in the L^∞ norm. Although this is more challenging than bounds for

standard norms, multiple recent works have proposed different a posteriori bounds suitable for these needs. We cite for instance [26] and [14]. Nevertheless, we face another issue that is the applications of these methods only to classical linear or quasi-linear equations. This is too restrictive for our problem where equation (3) is in general fractional and certainly non linear (except for $k = 1$). In addition, it presents a forcing term (the white noise) in a functional form. In particular, the second point is problematic since most of the a posteriori bounds present the equation residual in a strong form (also in Theorem 4.3).
On the other hand, the best mesh refinement could be deduced from what we know about the properties of the Matérn field (i.e. the analytical solution) and not necessarily from the numerical a posteriori results. First of all, recall that the field is homogeneous. However, we still have to consider two collateral facts: first of all, the solution u_h is defined on G but influences the solution q_h only when evaluated in D. Secondly, there are artificial homogeneous boundary conditions at a certain distance from D. From these observations, it is clear that the central region of G (the one corresponding to D) should be privileged with respect to the areas close to the border. However, this study would presumably require an extensive analysis that we prefer not to tackle in this paper and focus instead on other computational aspects. Although, we recommend future research to achieve this objective with possible concrete results.

4.5 The optimized mixed-mass single level generation method

To conclude this section about the numerical resolution of the problem in (5) with a single grid level h, we introduce a method originally presented in [12] to generate white noise on finite element spaces in optimal times.
We recall that the application of the white noise on the basis functions is represented by a normal vector distributed as $\mathcal{N}(\mathbf{0}, M)$, where M is the mass matrix of the finite element space (see section 3.2). Therefore, to fully estimate $\mathbb{E}[P(q)]$, where q is the analytical solution of (5) and P is a functional of q, we first need to generate independent samplings of this normal vector.
The following algorithm is the standard choice for the generation of a random vector $\sim N(\mathbf{0}, M)$.

Algorithm 1 Standard method to generate white noise on finite element spaces

Compute H such that $HH^T = M$ (Cholesky factorization)
Generate $\mathbf{Z} \sim N(\mathbf{0}, I)$
Compute $\mathbf{Z}_M = H\mathbf{Z}$
Return $\mathbf{Z}_M$

This algorithm returns a vector that is correctly distributed thanks to the properties of the covariance matrix of normal vectors. However, the Cholesky factorization cost could be very demanding ($\approx \mathcal{O}(N_h^3)$) if the number of degrees of freedom N_h is very high.
The algorithm proposed by [12] is based on the properties of the mass matrix itself. Namely, if m_e is the number of degrees of freedom for each element, n is the number of elements, then $M \in \mathbb{R}^{N_h \times N_h}$ can be proved to satisfy:

$$M = L^T \text{diag}(M_e) L,$$

where $L \in \mathbb{R}^{nm_e \times N_h}$ is built as $L^T = [L_1^T \dots L_n^T]$ so that $L_e \in \mathbb{R}^{m_e \times N_h}$ is a boolean matrix that represents the local-to-global map. Furthermore, $\text{diag}(M_e) \in \mathbb{R}^{nm_e \times nm_e}$ is the block diagonal

matrix composed by the local mass matrices. Hence, the Cholesky matrix can be computed as:

$$H = L^T \text{diag}(H_e),$$

where $\text{diag}(H_e)$ is the block diagonal matrix with the Cholesky factorizations of the local mass matrices M_e. We resume the steps of such algorithm.

Algorithm 2 Mixed-mass single level generation method ([12])

Compute H_e such that $H_e H_e^T = M_e \ \forall e \in \mathcal{T}_h$ (Cholesky factorization)
Compute L as local-to-global map
Compute $H = L^T \text{diag}(H_e)$
Generate $\mathbf{Z} \sim N(\mathbf{0}, I)$ (of size N_h)
Compute $\mathbf{Z}_M = H\mathbf{Z}$
Return $\mathbf{Z}_M$

One can easily check that this strategy yields a factorization cost of $\mathcal{O}(m_e^3 N_h)$ against the $\mathcal{O}(N_h^3)$ of Algorithm 1. It is now evident that, if the number of elements is sufficiently high, the computational benefit of this second method is noteworthy.

5 Study on the multi level total error

After the discussion on the error induced by the numerical resolution on a single level $h > 0$, we motivate in this section the utility of such analysis for a multilevel approach as presented in section 3.4. Later, we will present two algorithms that permit to generate correlated white noises on two different mesh refinements. Therefore, from the numerical discretization presented in section 3.2, the MLMC method introduced in section 3.4 and one of the following two algorithms, one can effectively perform a multilevel numerical estimation of problem (5).

5.1 The utility of single level results for multilevel methods

As already mentioned, the results in section 4 are not only of practical utility to perform single level estimations of the stochastic solution, but also permit us to improve the estimation efficiency and accuracy when performed with multilevel methods. Let us start with the theoretical motivation and recall hypothesis 1 from Theorem 3.2, which reads:

$$\exists \alpha \quad \text{such that} \quad \mathbb{E}[P(q_{h_l}) - P(q)] \leq c h_l^{\alpha}.$$

What we have obtained at the end of section 4 is instead a bound of the quantity $\mathbb{E}[\|q_h - q\|_V^2]$ (Corollary 4.1). Hence, depending on the regularity assumptions of P (especially, Lipschitz or Hölder continuity), one can directly guarantee hypothesis 1 and determine the parameter α from the a priori bound obtained in section 4.
On the other hand, hypotheses 2 and 3 of Theorem 3.2 can be guaranteed only by analysing the properties of the multilevel generation strategy. Therefore, also for this reason, we believe it is useful to present two of these strategies in sections 5.2 and 5.3.

As already mentioned, the a posteriori bound on the single level (partially but not fully achieved in section 4.4) is useful to perform grid adaptivity. This technique allows us to compute more accurate solutions through a smart refinement of the mesh. Error indicators (as in Theorem 4.3), are each one associated to one element. Therefore, high error indicators represent elements that are preferable to refine. Mesh adaptivity is a popular method to improve solutions without the burden of a full refined mesh. On the other hand, it can be useful also for the improvement of multilevel schemes. In many different contests, it is common that the error convergence does not follow a clear trend when considering few initial levels $h_0, h_1 \dots h_n$. This is often called *pre-asymptotic regime* and it strongly depends on the regularity of the solution and quality of the mesh. However, an effective mesh adaptivity study could considerably reduce this effect. In our study, the pre-asymptotic regime makes a MLMC estimation less effective because it does not fully exploit the improvements from the initial refinements. So, an incisive grid adaptivity study on the single levels would permit the multilevel estimation to fully exploit its potential.

5.2 The optimized mixed-mass multi level generation method

The first strategy we aim to present is the multilevel extension of the algorithm already described in section 4.5 and originally introduced in [12].
Let $\mathcal{T}_l$ be the triangulation for a certain index l, i.e. the level. We therefore aim to generate the correlated $\mathbf{Z}_M^l \sim \mathcal{N}(\mathbf{0}, M^l)$ and $\mathbf{Z}_M^{l-1} \sim \mathcal{N}(\mathbf{0}, M^{l-1})$ where M^l is the mass matrix on $\mathcal{T}_l$. First of all, one needs to construct the supermesh of the two consecutive levels $\mathcal{T}_{l,l-1}$, i.e. a mesh that contains both $\mathcal{T}_l$ and $\mathcal{T}_{l-1}$ (read [12] for more details and see the `libsupermesh`[1] library for the implementation). Clearly, this construction is not needed if $\mathcal{T}_l \supset \mathcal{T}_{l-1}$. Then, one can define the local mass matrices as

$$\left(M_e^l\right)_{ij} := \int_e \varphi_i^l \varphi_j^l \,\mathrm{d}x, \quad \forall e \in \mathcal{T}_{l,l-1},$$
$$\left(M_e^{l_1}\right)_{ij} := \int_e \varphi_i^{l-1} \varphi_j^{l-1} \,\mathrm{d}x, \quad \forall e \in \mathcal{T}_{l,l-1},$$

as well as the local mixed-mass matrix:

$$\left(M_e^{l,l-1}\right)_{ij} := \int_e \varphi_i^l \varphi_j^{l-1} \,\mathrm{d}x, \quad \forall e \in \mathcal{T}_{l,l-1},$$

where φ_i^l indicates as usual the i-th basis function of $\mathcal{T}_l$. One can also define the global mixed-mass matrix that is equivalent to:

$$M^{l,l-1} = \left(L^l\right)^T \mathrm{diag}_e(M_e^{l,l-1}) L^{l-1},$$

where the above terms are defined as in section 4.5 but only specific for a certain level l.
We can finally present the full scheme in Algorithm 3. Here, $N_{l,l-1}$ is the total number of degrees in $\mathcal{T}_{l,l-1}$ and $m_{e,l}$ is the local number of degrees of freedom in $\mathcal{T}_l$.
To conclude, we refer again to [12] to see some results from this generation strategy.

[1] `https://bitbucket.org/libsupermesh/`

Algorithm 3 Mixed-mass multilevel generation method ([12])

Compute $M_e^{l,l-1}\ \forall e \in \mathcal{T}_{l,l-1}$
Compute H_e such that $H_e H_e^T = M_e^l\ \forall e \in \mathcal{T}_{l,l-1}$ (Cholesky factorization)
Compute L^l, L^{l-1} as local-to-global maps
Generate $\mathbf{Z}_e \sim \mathcal{N}(\mathbf{0}, I)$ (of size $m_{e,l}$) $\forall e \in \mathcal{T}_{l,l-1}$
Compute $\mathbf{Z}_{M,e}^l = H_e \mathbf{Z}_e$
Compute $\mathbf{Z}_{M,e}^{l-1} = \left(M^{l,l-1}\right)^T H_e^{-T} \mathbf{Z}_e$
Compute $\mathbf{Z}_M^l = \sum_{e=1}^{N_{l,l-1}} \left(L_e^l\right)^T \mathbf{Z}_{M,e}^l$ and $\mathbf{Z}_M^{l-1} = \sum_{e=1}^{N_{l,l-1}} \left(L_e^{l-1}\right)^T \mathbf{Z}_{M,e}^{l-1}$
Return $\mathbf{Z}_M^l$ and $\mathbf{Z}_M^{l-1}$

5.3 The Gaussian conditional correction generation method

In this section, we instead present a conditional method we recently developed to generate, as in section 5.2, correlated realizations of $\mathbf{Z}_M^l$ and $\mathbf{Z}_M^{l-1}$.
Let us start by noticing that, if $\mathcal{T}_{l-1} \subset \mathcal{T}_l$ and the finite element spaces satisfy some basic properties (our definition in section 3.2 is certainly suitable), then:

$$\exists S \in \mathbb{R}^{N_{l-1} \times N_l} \text{ such that } \varphi_i^{l-1} = \sum_{j=1}^{N_l} S_{ij} \varphi_j^l, \quad \forall i = 1, \dots, N_{l-1}$$

Hence, we might compute the normal realization on the coarse mesh from the exact same white noise realized on the fine mesh:

$$\left(\tilde{\mathbf{Z}}_M^{l-1}\right)_i := < \dot{W}, \varphi_i^{l-1} > = \sum_{j=1}^{N_l} S_{ij} < \dot{W}, \varphi_j^j > =: \sum_{j=1}^{N_l} S_{ij} \left(\mathbf{Z}_M^l\right)_j . \tag{18}$$

One can now check that $\mathbf{Z}_M^{l-1}$ follows the correct distribution:

$$\mathbf{Z}_M^l \sim \mathcal{N}(\mathbf{0}, M_l) \quad \Rightarrow \quad \tilde{\mathbf{Z}}_M^{l-1} \sim \mathcal{N}(\mathbf{0}, S M_l S^T) \overset{d}{=} \mathcal{N}(\mathbf{0}, M_{l-1}).$$

Algorithm 4 Generation from conditional Gaussian distributions (finer to coarser)

Compute $S \in \mathbb{R}^{N_{l-1} \times N_l}$
Generate $\mathbf{Z}_M^l \sim \mathcal{N}(\mathbf{0}, M_l)$
Compute $\mathbf{Z}_M^{l-1} = S \mathbf{Z}_M^l$
Return $\mathbf{Z}_M^{l-1}$

We aim now to achieve a similar algorithm in the inverse way, i.e. returning $\mathbf{Z}_M^l$ from $\mathbf{Z}_M^{l-1}$. The motivation for this choice will be explained at the end of this section.
The starting point is a classical algorithm for the generation of Gaussian vectors from conditional distributions (see for instance [28]). Namely, consider $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ as a n-dimensional Gaussian

vector and the following splitting:

$$\mathbf{X} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_k \\ Z_1 \\ \vdots \\ Z_{n-k} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \Sigma_Y & \Sigma_{YZ} \\ \Sigma_{ZY} & \Sigma_Z \end{bmatrix}\right).$$

Therefore, $\Sigma_Y \in \mathbb{R}^{k\times k}$,$\Sigma_Y \in \mathbb{R}^{(n-k)\times(n-k)}$ and $\Sigma_{YZ} = \Sigma_{ZY}^T \in \mathbb{R}^{n\times(n-k)}$. The purpose of the following algorithm is the conditional generation of $\mathbf{Y}$ when $\mathbf{Z}$ is already known.

Algorithm 5 Generation from conditional Gaussian distributions

Generate $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma_Z)$
Generate $\mathbf{X} = [\mathcal{Y}^T, \mathcal{Z}^T]^T \sim \mathcal{N}(\mathbf{0}, \Sigma)$
Compute $\mathbf{Y} = \mathcal{Y} + \Sigma_{YZ}\Sigma_Z^{-1}(\mathbf{Z} - \mathcal{Z})$
Return $\mathbf{Y}$

One can indeed prove that $\mathbf{Y}$ has the correct distribution.
We aim now to follow this general idea to develop an algorithm that returns $\mathbf{Z}_M^l \sim \mathcal{N}(\mathbf{0}, M_l)$ which is also correlated to the input $\mathbf{Z}_M^{l-1}$. Exploiting the theory of Kalman filters (see [31] for a general description), one can derive the following equation:

$$\mathbf{Z}_M^l = \tilde{\mathbf{Z}}_M^l + M_l S^T M_{l-1}^{-1}\left(\mathbf{Z}_M^{l-1} - S\tilde{\mathbf{Z}}_M^l\right), \tag{19}$$

where $\tilde{\mathbf{Z}}_M^l \sim \mathcal{N}(\mathbf{0}, M_l)$ is independent from $\mathbf{Z}_M^{l-1}$. The following proposition guarantees the correctness of equation (19).

Proposition 5.1. *Let $\mathbf{Z}_M^{l-1} \sim \mathcal{N}(\mathbf{0}, M_{l-1})$ be as usual the application of the white noise on a finite element space of level $l-1$ and let $\tilde{\mathbf{Z}}_M^l \sim \mathcal{N}(\mathbf{0}, M_l)$ be independent from $\mathbf{Z}_M^{l-1}$. Hence, $\mathbf{Z}_M^l$ from (19) is distributed as $\mathcal{N}(\mathbf{0}, M_l)$. Moreover, $\mathbf{Z}_M^{l-1}$ and $\mathbf{Z}_M^l$ are correlated as in Algorithm 4, i.e. as they were generated from the same white noise.*

Proof. One can easily check that $\mathbf{Z}_M^l$ is a zero-mean Gaussian vector. In addition, thanks to the hypotheses on the distributions and independence of $\mathbf{Z}_M^{l-1}$ and $\tilde{\mathbf{Z}}_M^l$, we have that:

$$\mathrm{Cov}[\mathbf{Z}_M^l, \mathbf{Z}_M^{l-1}] = M_l S^T M_{l-1}^{-1}\mathrm{Cov}[\mathbf{Z}_M^{l-1}, \mathbf{Z}_M^{l-1}] = M_l S^T M_{l-1}^{-1} M_{l-1} = M_l S^T = S M_l,$$

where the last equality exploits the symmetries of the mass and covariance matrices. Now, notice that SM_l is the same correlation matrix as in Algorithm 4.
We only miss the computation of the covariance matrix of $\mathbf{Z}_M^l$. Exploiting again equation (19)

and independence of the distributions, we have that:

$$\begin{aligned}
\mathrm{Cov}[\mathbf{Z}_M^l, \mathbf{Z}_M^l] &= \mathrm{Cov}\left[\left(\tilde{\mathbf{Z}}_M^l + M_l S^T M_{l-1}^{-1}\left(\mathbf{Z}_M^{l-1} - S\tilde{\mathbf{Z}}_M^l\right)\right), \left(\tilde{\mathbf{Z}}_M^l + M_l S^T M_{l-1}^{-1}\left(\mathbf{Z}_M^{l-1} - S\tilde{\mathbf{Z}}_M^l\right)\right)\right] \\
&= \mathrm{Cov}[\tilde{\mathbf{Z}}_M^l, \tilde{\mathbf{Z}}_M^l] - \mathrm{Cov}[\tilde{\mathbf{Z}}_M^l, M_l S^T M_{l-1}^{-1} S\tilde{\mathbf{Z}}_M^l] + \\
&\quad + \mathrm{Cov}[M_l S^T M_{l-1}^{-1}\mathbf{Z}_M^{l-1}, M_l S^T M_{l-1}^{-1}\mathbf{Z}_M^{l-1}] - \mathrm{Cov}[M_l S^T M_{l-1}^{-1} S\tilde{\mathbf{Z}}_M^l, \tilde{\mathbf{Z}}_M^l] + \\
&\quad + \mathrm{Cov}[M_l S^T M_{l-1}^{-1} S\tilde{\mathbf{Z}}_M^l, M_l S^T M_{l-1}^{-1} S\tilde{\mathbf{Z}}_M^l] \\
&= M_l - M_l S^T M_{l-1}^{-1} S M_l + M_l S^T M_{l-1}^{-1} M_{l-1} M_{l-1}^{-1} S M_l - M_l S^T M_{l-1}^{-1} S M_l + \\
&\quad + M_l S^T M_{l-1}^{-1} S M_l S^T M_{l-1}^{-1} S M_l
\end{aligned}$$

Since $M_{l-1} = S M_l S^T$, one can show that $M_l^{-1} = S^T M_{l-1}^{-1} S$. Using this information, we obtain:

$$\mathrm{Cov}[\mathbf{Z}_M^l, \mathbf{Z}_M^l] = M_l - M_l M_l^{-1} M_l + M_l M_l^{-1} M_l - M_l M_l^{-1} M_l + M_l M_l^{-1} M_l M_l^{-1} M_l = M_l$$

as expected.

□

Finally, we can write explicitly the algorithm for the conditional generation from the coarser mesh to the finer mesh.

Algorithm 6 Generation from conditional Gaussian distributions (coarser to finer)

Compute $S \in \mathbb{R}^{N_{l-1} \times N_l}$
Generate $\mathbf{Z}_M^{l-1} \sim \mathcal{N}(\mathbf{0}, M_{l-1})$
Generate $\tilde{\mathbf{Z}}_M^l \sim \mathcal{N}(\mathbf{0}, M_l)$
Compute $\mathbf{Z}_M^l = \tilde{\mathbf{Z}}_M^l + M_l S^T M_{l-1}^{-1}\left(\mathbf{Z}_M^{l-1} - S\tilde{\mathbf{Z}}_M^l\right)$
Return $\mathbf{Z}_M^l$

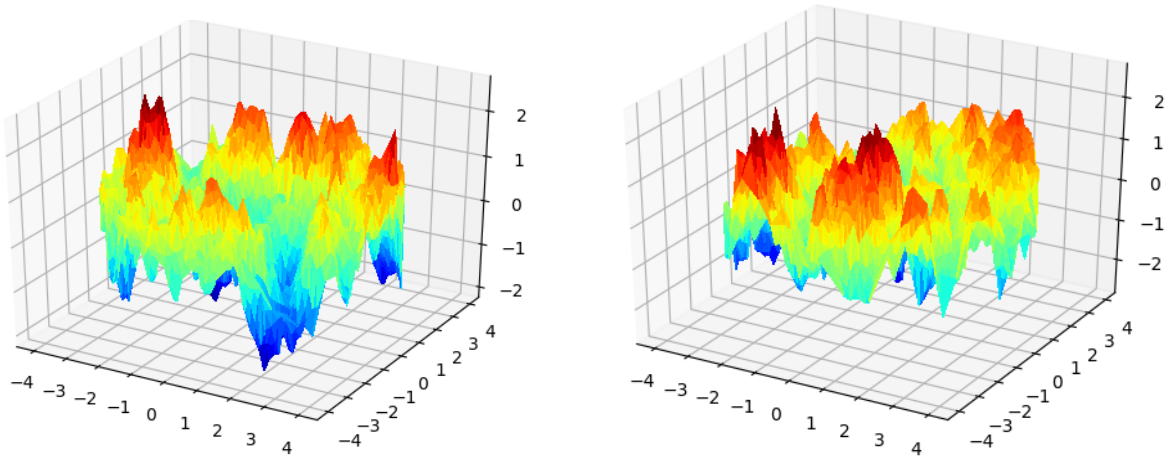
Algorithm 6 is certainly more computationally expensive that Algorithm 4 because of the multiple products and the computation of an inverse matrix. However, this approach could lead to a wise grid adaptivity strategy. Indeed, with the information obtained from the resolution of the problem at level $l-1$ and the pre-computed $\mathbf{Z}_M^l$ with (19), one could develop an a posteriori formulation with error indicators for the level l. This strategy avoids the burden to solve the problem at the level l before having performed grid-adaptivity on its mesh. In addition, it permits to execute grid-adaptivity on multiple levels following this sequential down-top procedure. Then, the sequential grid-adaptivity application on multiple levels $0, \dots l$ provides a set of optimal meshes where to perform the best MLMC estimation.

6 Some numerical results

In this section, we present some numerical results about the single level algorithms introduced in section 4.5. Codes are implemented using the `Fenics`[2] library for Python ([24]). In addition, we fixed the following parameters: dimension $d = 2$, finite element spaces with polynomials of degree $p = 1$, $k = 1$, $\eta = 1$, $\lambda = 1$. The other parameters ν, κ, σ can be directly obtained from the previous ones. An important note is that $k = 1$ and therefore we implemented a standard finite element method for (5b).

6.1 Accuracy of the mixed-mass white noise simulation

We start by observing a single realization of the Matérn field from equation (5b). Defining a domain G as a centred circle of radius 5 and $h = 0.5$, we generated realizations from both the algorithms that are shown in Figure 1.



(a) Algorithm 1 (b) Algorithm 2

Figure 1: Realizations of the Matérn field solving equation (5b) with FEM and two different algorithms for the white noise generation

At a first glance, the two solutions seem to share the same spatial distribution.
We aim now to achieve more concrete results that guarantee the correct generation of the Matérn field. To perform this analysis, we used algorithm 2 and FEM applied to equation (5b) to generate $2 \cdot 10^4$ realizations of the Matérn field. Finally, from these realizations we estimated the covariance function measuring the correlation between randomly picked points at 11 different distances. Therefore, a standard Monte Carlo method applied to 10^5 correlation measurements can return the estimation of the covariance function at every fixed distance $\{r_i\}_{i=1,\dots,11} > 0$. The result is shown in Figure 2.

[2] `https://fenicsproject.org/`

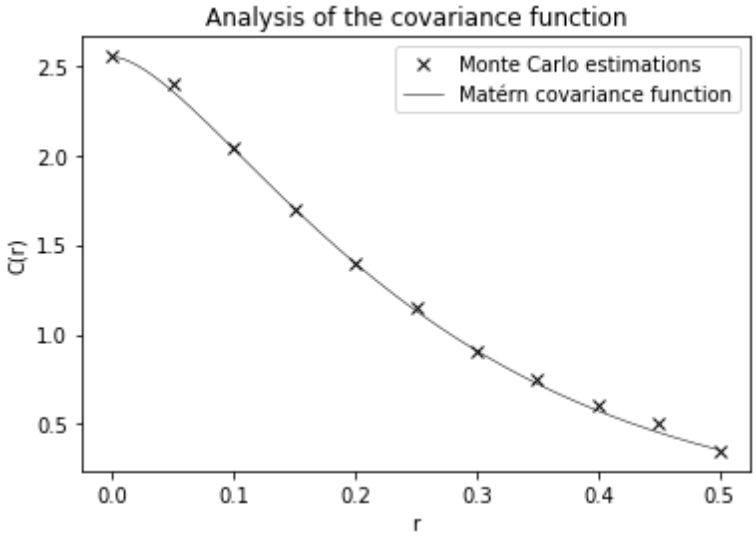


Figure 2: Theoretical Matérn covariance function and experimental estimation from the set of $2 \cdot 10^4$ realizations using algorithm 2.

The plot clearly guarantees the correct generation of the stochastic field. Also algorithm 1 gives similar results.

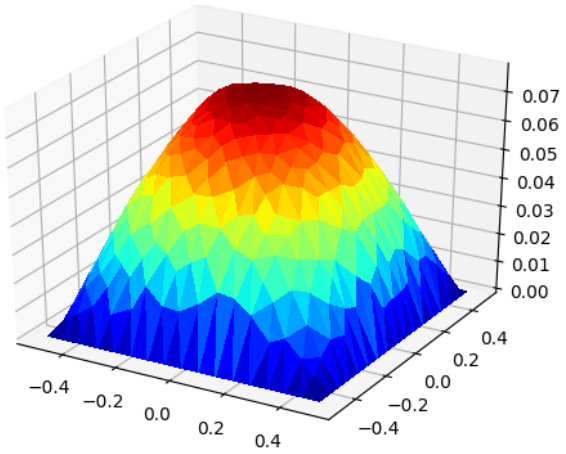
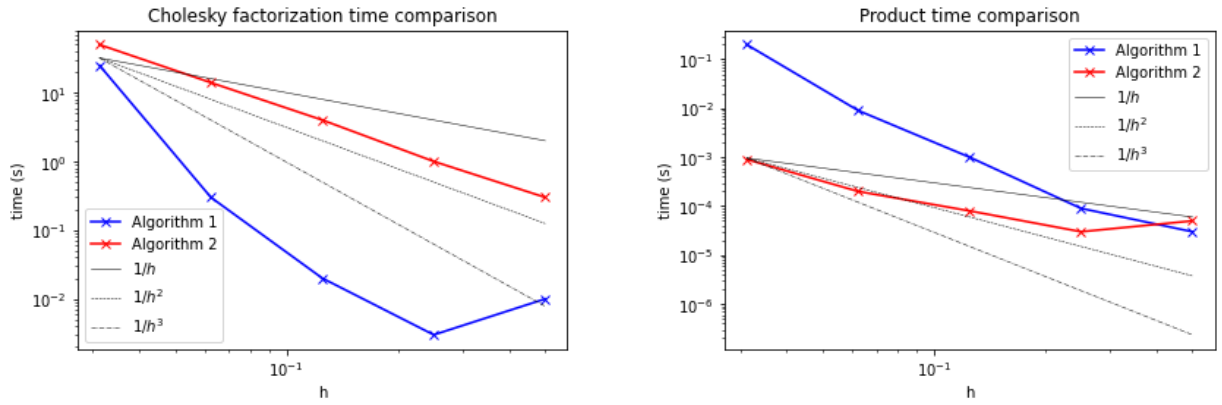


Figure 3: Single realization of the diffusion solution u_h from (5).

Now that we have confirmed the correctness of the generation algorithms and the Whittle model, we plot one realization of the final solution u_h of equation (5a) . We set $D = [-0.5, 0.5]^2$ and $G = [-1.5, 1.5]^2$ (with this setting, the distance between the domains is in fact equal to λ, cfr. Remark 3.3) and $f(\mathbf{x}) = 1$. The contour plot is shown in Figure 3 where we clearly recognize the diffusive regularization effect.

6.2 Efficiency of the mixed-mass white noise simulation

A second experimental analysis is related to the efficiency of algorithm 2 with respect to the standard one (algorithm 1). We recall that algorithm 1 has a cost of $\mathcal{O}(N_h^3)$ while algorithm 2 presents a cost of $\mathcal{O}(m_e^3 N_h)$ where m_e is the number of degrees of freedom for each element. Hence, in our case with $p = 1$, $m_e = 3$. A study has then been performed computing the time (in seconds) to accomplish the Cholesky factorization with different grid levels. Results are shown in Figure 4a.

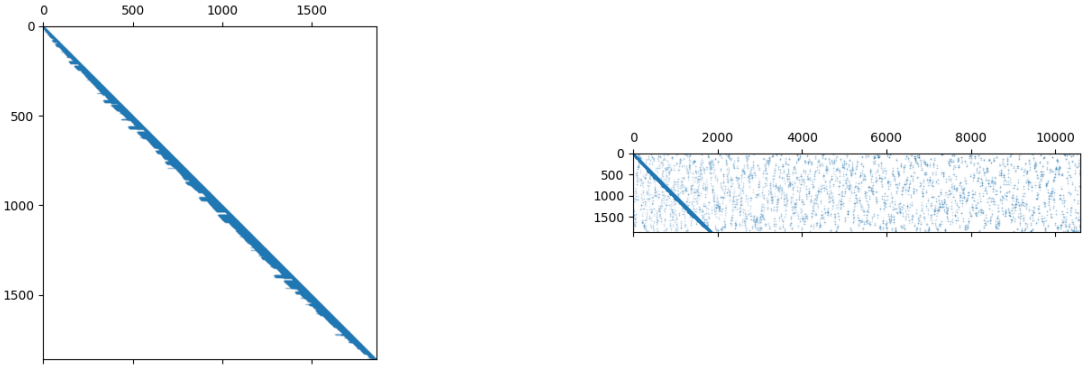


(a) Cholesky factorization time (b) $H\mathbf{Z}$ product time

Figure 4: Experimental computational times for the two algorithms.

At a first glance, the optimized algorithm from [12] is slower than the original one. However, a more careful analysis reveals that the order is constant and equal to $\mathcal{O}(h^{-2}) = \mathcal{O}(N_h)$ while the order of algorithm 1 is higher. Therefore, we expect algorithm 2 to be faster for smaller h. In addition, the initial slowness is due to the fact that algorithm 2 needs some explicit loops to assembly the L and $\text{diag}(M_e)$ matrices while algorithm 1 only exploits the optimized Cholesky factorization function[3] from `Scipy` ([29]). Hence, the initial gap is only due to the implementation in the Python[4] environment and should not be present with non interpreted languages.
In our study, we also noticed that algorithm 2 is even more performant when the number of realizations is high. This is explained by Figure 4b that reports the average time to accomplish the $H\mathbf{Z}$ product for different grid levels. Since the product time is much smaller in algorithm 2, the realization of many samples takes much less time.
The presumable reason of this fact is related to the different matrix sparsity. First of all, we show in Figure 5 the sparsity patterns of the two different Cholesky matrices for a certain grid level. We recall that the Cholesky matrix from algorithm 2 is indeed rectangular, more precisely of size (N_h, nm_e). Despite the H matrix obtained from algorithm 1 is smaller, the diagonal is more densely populated by non-zero elements and contains in total more than 4 times the number of non-zeros elements of the matrix obtained from algorithm 2. Therefore, this aspect shows another benefit of the algorithm proposed by [12] that was not originally noticed.

[3] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.linalg.cholesky.html`
[4] Python Software Foundation. Python Language Reference, version 3.9. Available at `http://www.python.org`



(a) Algorithm 1: 96912 non-zeros (b) Algorithm 2: 21180 non-zeros

Figure 5: Sparsity patterns of the Cholesky factorization matrices from the two algorithms ($N_h = 1860$). The matrix from algorithm 2 is bigger but more sparse.

7 Conclusion

In this work we have continued the analysis on an important family of models that deal with stochastic spatial diffusivity coefficients. The analysis of these problems is again motivated by several applications in different scientific contests such as geology, meteorology and biology.
More precisely, we have developed further and important results on diffusion problems with log-normal Matérn diffusivity coefficients. From our single level analysis, we managed to achieve explicit bounds for the discretization error when FEM is applied twice on the two different but coupled problems. In particular, we managed to sum up all the contributions to provide the bound to the generalized MSE of the diffusion solution. On the other hand, results hold in the specific case where $d = p = 1$ because the theory concerning these numerical methods is still very recent and, for now, we do not have access to more general results. We have discussed that this bound can be used to guarantee single level estimations precisions and it also brings important information when one deals with multilevel estimations.
We have also presented a short study on the a posteriori bounds, error indicators and grid adaptivity. We provided some but incomplete results because the definition of white noise is very challenging in the formulation of such bounds. We again recommend future research to proceed in this analysis because of the possible implications of a full a posteriori bound for the formulation of grid adaptivity techniques. These methods would consistently improve the efficiency of a multilevel estimation approach.
To conclude this work, we presented some sampling techniques to generate Gaussian vectors for single and multi-level realizations. In particular, except for the first basic technique, these methods are specific for the discretized realizations of the white noise. Hence, they represent a necessary tool to achieve solutions from the problem discussed in this work. So, the application of one of such methods together with the presented finite element schemes can indeed provide explicit solutions as shown in the section regarding the numerical results. Here, we validated the correctness of the Whittle model and the generation algorithm proposed by [12]. In addition, we confirmed the optimal cost of such a method and we also showed another important advantage related to the sparsity of the factorization matrices.

Bibliography

[1] Mark Ainsworth and Christian Glusa. Towards an efficient finite element method for the integral fractional laplacian on polygonal domains. pages 17–57, 2018. doi: 10.1007/978-3-319-72456-0_2.

[2] M. Aziz. Stochastic partial differential equations applied to hydrocarbon reservoir modelling. Master's thesis, University of Oxford, 2016.

[3] G. Benfatto, G. Gallavotti, and F. Nicoló. Elliptic equations and gaussian processes. *Journal of Functional Analysis*, 36(3):343–400, 1980. ISSN 0022-1236. doi: https://doi.org/10.1016/0022-1236(80)90094-4.

[4] David Bolin and Kristin Kirchner. The rational spde approach for gaussian random fields with general smoothness. *Journal of Computational and Graphical Statistics*, 29(2):274–285, 2020. doi: 10.1080/10618600.2019.1665537.

[5] David Bolin, Kristin Kirchner, and Mihály Kovács. Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. *IMA Journal of Numerical Analysis*, 40(2):1051–1073, 12 2018. ISSN 0272-4979. doi: 10.1093/imanum/dry091.

[6] Susanne Brenner and L. Ridgway Scott. *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics. Springer New York, 2002. ISBN 9780387954516.

[7] John M. Charnes. Using simulation for option pricing. 1:151–157, 2002. doi: 10.1109/WSC.2000.899710.

[8] Julia Charrier. Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients. June 2010.

[9] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Computing and Visualization in Science*, 14, 2011. doi: 10.1007/s00791-011-0160-x.

[10] Albert Cohen and Giovanni Migliorati. Near-optimal approximation methods for elliptic pdes with lognormal coefficients. 2021. doi: 10.48550/ARXIV.2103.13935.

[11] Sonja G. Cox and Kristin Kirchner. Regularity and convergence analysis in sobolev and hölder spaces for generalized whittle–matérn fields. *Numerische Mathematik*, 146(4):819–873, nov 2020. doi: 10.1007/s00211-020-01151-x.

[12] Matteo Croci, Michael B. Giles, Marie E. Rognes, and Patrick E. Farrell. Efficient white noise sampling and coupling for multilevel monte carlo with non-nested meshes. 2018.

[13] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 2014. doi: 10.1017/CBO9781107295513.

[14] Alan Demlow and Natalia Kopteva. Maximum-norm a posteriori error estimates for singularly perturbed elliptic reaction-diffusion problems. *Numerische Mathematik*, 133(4):707–742, August 2016. ISSN 0945-3245. doi: 10.1007/s00211-015-0763-0.

[15] Michael B. Giles. Multilevel monte carlo path simulation. *Operations Research*, 56:607–617, 06 2008. doi: 10.1287/opre.1070.0496.

[16] Michael B. Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015. doi: 10.1017/S096249291500001X.

[17] Ivan G. Graham, Frances Y. Kuo, Dirk Nuyens, Rob Scheichl, and Ian H. Sloan. Analysis of circulant embedding methods for sampling stationary random fields. *SIAM Journal on Numerical Analysis*, 56(3):1871–1895, 2018. doi: 10.1137/17M1149730.

[18] Rüdiger Hewer, Petra Friederichs, Andreas Hense, and Martin Schlather. A matérn-based multivariate gaussian random process for a consistent model of the horizontal wind components and related variables. *Journal of the Atmospheric Sciences*, 74(11):3833 – 3845, 2017. doi: 10.1175/JAS-D-16-0369.1.

[19] Ustim Khristenko, Laura Scarabosio, Piotr Swierczynski, Elisabeth Ullmann, and Barbara Wohlmuth. Analysis of boundary effects on pde-based sampling of whittle-matérn random fields. 2018. doi: 10.48550/ARXIV.1809.07570.

[20] E. Kloeden, Peter and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992. ISBN 978-3-662-12616-5. doi: https://doi.org/10.1007/978-3-662-12616-5.

[21] Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2011. ISBN 978-0-470-17793-8.

[22] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach [with discussion]. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 73(4):423–498, 2011. ISSN 13697412, 14679868.

[23] Anna Lischke, Guofei Pang, Mamikon Gulian, Fangying Song, Christian Glusa, Xiaoning Zheng, Zhiping Mao, Wei Cai, Mark M. Meerschaert, Mark Ainsworth, and George Em Karniadakis. What is the fractional laplacian? 2018. doi: 10.48550/ARXIV.1801.09767.

[24] J. Hake A. Johansson B. Kehlet A. Logg C. Richardson J. Ring M. E. Rognes M. S. Alnaes, J. Blechta and G. N. Wells. The FEniCS project version 1.5. *Archive of Numerical Software*, 3, 2015. doi: 10.11588/ans.2015.100.20553.

[25] Bertil Matérn. *Spatial Variation; Stochastic Models and Their Application to Some Problems in Forest Surveys and Other Sampling Investigations*. Stockholm. Statens Skogsforskningsinstitut. Meddelanden. University of Sweden, 1966.

[26] Ricardo Nochetto, Alfred Schmidt, Kunibert Siebert, and Andreas Veeser. Pointwise a posteriori error estimates for monotone semi-linear equations. *Numerische Mathematik*, 104:515–538, 09 2006. doi: 10.1007/s00211-006-0027-0.

[27] Alfio Quarteroni. *Numerical Models for Differential Problems*. MS&A. Springer Milan, 2014. ISBN 9788847055223.

[28] Ernest M. Scheuer and David S. Stoller. On the generation of normal random vectors. *Technometrics*, 4(2):278–281, 1962. ISSN 00401706.

[29] Pauli Virtanen et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[30] Peter Whittle. On stationary processes in the plane. *Biometrika*, 41(3/4):434–449, 1954. ISSN 00063444.

[31] P. Zarchan and H. Musoff. *Fundamentals of Kalman Filtering: A Practical Approach*. Progress in astronautics and aeronautics. American Institute of Aeronautics and Astronautics, Incorporated, 2000. ISBN 9781563472558.

8 Appendix

Exponential symmetric property of the Matérn field

Proposition 8.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $u : \mathbb{R}^d \times \Omega \to \mathbb{R}$ be a zero-mean Gaussian stochastic field with a radially symmetric covariance function (i.e. $C(x, y) = C(-x, -y)$ for all $x, y \in \mathbb{R}^d$) and $B \subset \mathbb{R}^d$. Then:*

$$\mathbb{E}\left[e^{-\min_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right] = \mathbb{E}\left[e^{\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right] \lneq \mathbb{E}\left[e^{\|u(\cdot,\omega)\|_{L^\infty(B)}}\right].$$

Proof. We use "$\sim$" to refer to equality in distribution. First of all, it is easy to check that a Gaussian field with zero mean and radially symmetric covariance function is such that $u \sim -u$. It trivially implies that:

$$\max_{\mathbf{x}\in B} u(\mathbf{x}) \sim \max_{\mathbf{x}\in B} -u(\mathbf{x}) = -\min_{\mathbf{x}\in B} u(\mathbf{x}). \tag{20}$$

Now define $\mathrm{d}\gamma_{min}$ and $\mathrm{d}\gamma_{max}$ as the probability measures associated to the distribution of the random variables $\min_{\mathbf{x}\in B} u(\mathbf{x})$ and $\max_{\mathbf{x}\in B} u(\mathbf{x})$. Then:

$$\begin{aligned}
\mathbb{E}\left[e^{-\min_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right] &= \int_{\mathbb{R}} e^{-y}\mathrm{d}\gamma_{min}(y) \overset{\text{(Change of variables)}}{=} \\
&= \int_{\mathbb{R}} e^{t}\mathrm{d}\gamma_{min}(-t) \overset{(20)}{=} \\
&= \int_{\mathbb{R}} e^{t}\mathrm{d}\gamma_{max}(t) \\
&= \mathbb{E}\left[e^{\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right].
\end{aligned}$$

To prove the strict inequality, recall that Gaussian fields have continuous trajectories and then:

$$\|u(\cdot,\omega)\|_{L^\infty(B)} = \max_{\mathbf{x}\in B} |u(\mathbf{x},\omega)| = \max\{-\min_{\mathbf{x}\in B} u(\mathbf{x},\omega), \max_{\mathbf{x}\in B} u(\mathbf{x},\omega)\}.$$

Finally, we have that $\|u(\cdot,\omega)\|_{L^\infty(B)} \geq \max_{\mathbf{x}\in B} u(\mathbf{x},\omega)$. Due to the monotonicity of the exponential function and integral operator, we have that:

$$\mathbb{E}\left[e^{\|u(\cdot,\omega)\|_{L^\infty(B)}}\right] \geq \mathbb{E}\left[e^{\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right]$$

On the other hand, since $-\min_{\mathbf{x}\in B} u(\mathbf{x},\omega)$ and $\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)$ have the same distribution,

$$\begin{aligned}
&\exists A \in \mathcal{F} \text{ s.t. } \mathbb{P}(A) > 0 \text{ and } -\min_{\mathbf{x}\in B} u(\mathbf{x},\omega) > \max_{\mathbf{x}\in B} u(\mathbf{x},\omega) \ \ \forall \omega \in A \\
&\Rightarrow \|u(\cdot,\omega)\|_{L^\infty(B)} > \max_{\mathbf{x}\in B} u(\mathbf{x},\omega) \ \ \forall \omega \in A \\
&\Rightarrow e^{\|u(\cdot,\omega)\|_{L^\infty(B)}} > e^{\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)} \ \ \forall \omega \in A \\
&\Rightarrow \mathbb{E}\left[e^{\|u(\cdot,\omega)\|_{L^\infty(B)}}\right] \gneq \mathbb{E}\left[e^{\max_{\mathbf{x}\in B} u(\mathbf{x},\omega)}\right].
\end{aligned}$$

□