Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Analysis of multilevel MCMC-FEM for Bayesian inversion of log-normal diffusions

V.H. Hoang and J.H. Quek and Ch. Schwab

Research Report No. 2019-05

January 2019
Latest revision: June 2019

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

ANALYSIS OF MULTILEVEL MCMC-FEM FOR BAYESIAN INVERSION OF LOG-NORMAL DIFFUSIONS

VIET HA HOANG[1] AND JIA HAO QUEK[1] AND CHRISTOPH SCHWAB[2]

ABSTRACT. We develop the Multilevel Markov Chain Monte Carlo Finite Element Method (MLMCMC-FEM for short) to sample from the posterior density of the Bayesian inverse problems. The unknown is the diffusion coefficient of a linear, second order divergence form elliptic equation in a bounded, polytopal subdomain of $\mathbb{R}^d$. We provide a convergence analysis with absolute mean convergence rate estimates for the proposed modified MLMCMC Finite Element Method (MLMCMC-FEM) showing in particular error vs. work bounds which are explicit in the discretization parameters. This work generalizes the MLMCMC-FEM algorithm and the error vs. work analysis for uniform prior measure from [21] which we also review here, to linear, elliptic, divergence-form PDEs with log-gaussian uncertain coefficient and Gaussian prior measure. In comparison to [21], we show by mathematical proofs and numerical examples that the unboundedness of the parameter range under gaussian prior and the nonuniform ellipticity of the forward model require essential modifications in the MCMC sampling algorithm and in the error analysis. The proposed novel multilevel MCMC sampler applies to general Bayesian inverse problems with log-gaussian coefficients. It only requires a numerical forward solver with essentially optimal complexity for producing an approximation of the posterior expectation of a quantity of interest within a prescribed accuracy. Numerical examples using independence and pCN samplers confirm our error vs. work analysis.

1. INTRODUCTION

In recent years, the field of *computational uncertainty quantification* (UQ for short) has emerged as a broad area of computational science and engineering. It addresses the efficient computational analysis of responses of partial differential equations (PDEs for short) in science and engineering for input data which are either unknown or for which only partial, statistical information is available. *Uncertainty propagation* is, in this situation, aiming at producing computable statistical information of the PDE responses. It is a part of so-called *forward UQ*, where uncertain measurement data and incomplete information on material properties and physical domains are to be converted into quantitative information on the corresponding PDE solutions.

The present paper addresses so-called *inverse UQ* where, for random or statistical PDE input, quantities of interest (QoI's for short) are to be computed. In a *Bayesian framework*, this amounts to numerical estimation of mathematical expectations of PDE responses over all admissible input data, conditional on noisy observation of measurement data.

Numerous papers have appeared with mathematical and computational investigations of a number of methodologies for the Bayesian inversion of PDEs with uncertain inputs; we mention [10, 9] and the references there for a presentation of

MCMC methods for PDEs which account explicitly for the dependence on PDE discretization parameters. MCMC methods for Bayesian PDE inversion can be prohibitively expensive. Accordingly, many attempts have been made to reduce computational complexity. We mention exemplarily the works of Lieberman et al. [24] and Martin et al. [25] on Bayesian inverse problems with log-gaussian priors.

Multi-level versions of SMC, Particle Filters, EnKF etc. for data assimilation and inference under PDE constraints have been recently proposed and analyzed. We refer to [23, 4, 3, 22] and the references there for recent contributions on this. Recent work on multi-level algorithms in uncertainty quantification for PDEs includes in particular the numerical analysis of multilevel methods for the filtering problem, see, e.g. [11, 23, 30]. In these works, while also admitting noisy data, uncertain input of the PDE is limited to the forcing term, similar to the setting of [8].

In [28], MLMCFEM and (single-level)QMC FEM for Bayesian PDE inversion under log-gaussian diffusion coefficient and under gaussian prior measure have been considered; *ratio estimators* as proposed in [29] were investigated there with MLMC integration to directly estimate the expectation of the QoI under the posterior expectation and the normalization constant of the posterior measure. It was assumed in [28, Appendix] that the log-gaussian diffusion coefficient is bounded away from zero by a positive constant.

To the best of our knowledge, none of these references address the problem of sign-indefiniteness in the exponent of posterior densities in the telescoping sums of multi-level Bayesian estimators under gaussian priors, and with log-gaussian diffusion coefficients. The posterior density may become, with positive probability, arbitrary close to zero. This, however, entails mathematical issues (posterior densities of increments between discretization levels may not be integrable w.r. to the gaussian prior measure), and can, as we show in the present paper with numerical examples, foil practical realizations of the MLMCMC FEM methods.

In Section 4 of the present paper, we resolve this mathematical issue by a novel redesign of our MLMCMC-FEM algorithm for uniform prior measure from [21]. We present a complete error vs. work analysis of the independence sampler, and propose a corresponding version of the pCN-based MLMCMC.

Although we detail in the present paper the design of the MLMCMC and its analysis for the Metropolis Hastings type MCMC, the mentioned integrability issue and our proposed modification of the algorithm apply equally well to other variants of MCMC; we mention only sequential MC (see [2, 5] and the references there), and geometric MCMC (see [1] and the references there). The presently proposed modification may therefore also facilitate convergence proofs of these methods. The principal contributions of the present note are as follows: we give the first complete numerical analysis of a MLMCMC-FEM for Bayesian inverse problems for elliptic PDEs with log-gaussian, uncertain coefficient, and under gaussian prior. While the MLMCMC algorithm developed here is similar to our previous work in [21] for uniform prior measure, there are essential differences both to the MLMCMC algorithms and analysis for uniform prior measure as well as to the single level (plain) MCMC (SLMCMC) algorithm for gaussian prior where the forward equation is solved with the same (fine) mesh for all samples, which was analyzed in [19] and also in [13].

The structure of this paper is as follows. In Section 2, we introduce the general class of MCMC samplers which we consider here, and the model linear diffusion problem in a bounded, polytopal domain $D \subset \mathbb{R}^d$. In Section 3, we review results from [21] on the Bayesian inverse problem for uniform priors. In part to motivate our analysis for the gaussian prior, we re-derive the MLMCMC for the uniform prior in Section 3.3, and briefly recapitulate from [21] the key convergence and error vs. work statements. We then proceed in Section 4 to the derivation of the MLMCMC FEM for the Bayesian PDE inversion under gaussian prior on the log-gaussian diffusion coefficient. We present several key estimates for the log-gaussian diffusion problem, its parametric solution, and its Galerkin FE discretization, from [7, 14, 15]. To reduce technicalities in our presentation, we do not work under the weakest conceivable assumptions on the isotropic gaussian diffusion coefficient $a = \exp(R)$ with a scalar gaussian random field (GRF for short) R, or on the source term f or the domain D. We also do not admit the most general mathematical forward model, but rather confine the presentation to a linear, second order, divergence form PDE with isotropic, log-gaussian diffusion coefficient, in a bounded, polyhedral domain D. To minimize FE technicalities, we assume that realizations of R are Lipschitz in D almost sure w.r. to the prior measure, that $f \in L^2(D)$ and that the domain D is convex. This ensures pathwise a.s. $H^2(D)$ regularity of the parametric PDE solution, and obviates discussion and use of FEM with corner- and edge mesh refinement etc. We repeat that these assumptions were only made to simplify the present exposition; they are not essential in the mathematical arguments regarding the convergence rate of the MLMCMC-FEM algorithm which we present here. Extending the analysis to weakest conditions (e.g. bounded Lipschitz domain, anisotropic diffusion, source term f which is also random and of lower regularity than $L^2(D)$, etc.) is possible *verbatim*, albeit at the expense of further "FE-discretization related parameters and technicalities" (such as weighted spaces, fractional convergence orders of the FEM, graded meshes, etc.). For clarity of exposition, and as the line of argument of the present MLMCMC-FEM convergence rate analysis is not affected by these, we do not detail them here. Section 5 presents numerical examples that validate the theoretical results on the convergence of the MLMCMC-FEM. We first present an example where the MLMCMC-FEM developed for the uniform prior in [21] does not converge but the new MLMCMC-FEM developed for the Gaussian prior in this paper converges as theoretically proved. We then present some numerical examples for elliptic equations with coefficient of the so-called "log-gaussian" form in a two dimensional domain. The MLMCMC-FEM is performed using both the independence and pCN samplers. The numerical results confirm the theoretical convergence rate. To show the essential optimality of the complexity of the MLMCMC-FEM, we record results on the CPU time required to perform the method. The results indicate that the CPU time required is indeed of optimal asymptotic order. The paper concludes with some appendices containing technical proofs and the sampling procedure using circulant embedding to sample FE nodes of a GRF. Throughout the paper, by c we denote a generic constant that does not depend on the approximating parameters, whose value can change from one appearance to the next.

2. Bayesian inverse problems for elliptic PDEs

We present in this section the setting of the Bayesian inverse problem for inferring the unknown coefficient K of an elliptic partial differential equation, given noisy observations of the solution in the form of a finite number of linear functionals of this function, perturbed by additive, centered gaussian observation noise.

2.1. **Model Problem.** Let (U, Θ, γ) be a probability space of parameters u and let D be a bounded polytopal domain in $\mathbb{R}^d$. The dimension d of the physical domain D is assumed to equal $1, 2, 3$. Assume further that $K : U \to L^\infty(D)$ is strongly measurable such that for every $u \in U$ there exist constants $c_1(u)$ and $c_2(u)$ such that

$$0 < c_1(u) \leq K(x, u) \leq c_2(u), \tag{2.1}$$

almost everywhere with respect to the Lebesgue measure in $\mathbb{R}^d$. We consider the parametric diffusion problem

$$-\nabla \cdot (K(\cdot, u)\nabla P(u, \cdot)) = f, \quad P(u, x) = 0, \text{ when } x \in \partial D, \tag{2.2}$$

where $f \in V'$ with $V := H_0^1(D)$. Let $\mathcal{O}_1, \dots, \mathcal{O}_k \in V'$. Then, the forward data to observation map $\mathcal{G}(u) : U \to \mathbb{R}^k$ is defined as

$$\mathcal{G}(u) = (\mathcal{O}_1(P(\cdot, u)), \dots, \mathcal{O}_k(P(\cdot, u))). \tag{2.3}$$

We assume at hand observation data δ of the response $\mathcal{G}$ corrupted by additive, centered gaussian observation noise, i.e.

$$\delta = \mathcal{G}(u) + \vartheta$$

where ϑ is a random variable with value in $\mathbb{R}^k$ which follows the normal distribution $N(0, \Sigma)$ where Σ is a known $k \times k$ symmetric and positive definite covariance matrix. Our aim is to compute approximate expectations under the Bayesian posterior probability measure γ^δ, i.e., the conditional probability $\gamma(u|\delta)$. In particular, we wish to approximate the expectation with respect to the measure γ^δ of "quantities of interest", being continuous, linear functionals of the parametric solution.

We first recall the following standard result on the existence and well-posedness of the posterior γ^δ. To this end, we define the Bayesian data "misfit" (or "Bayesian potential") functional

$$\Phi(u; \delta) = \frac{1}{2}|\delta - \mathcal{G}(u)|_\Sigma^2 = \frac{1}{2}(\delta - \mathcal{G}(u))^\top \Sigma^{-1}(\delta - \mathcal{G}(u)). \tag{2.4}$$

Cotter et al. [8] (Theorem 2.1) prove the following general result on the existence of the posterior γ^δ (see also [33])

Proposition 2.1. *If, in* (2.4)*, the parametric forward map* $\mathcal{G} : U \to \mathbb{R}^k$ *is measurable on* (U, Θ)*, then the posterior* γ^δ *is absolutely continuous with respect to the prior* γ*. The Radon-Nikodym derivative is given by*

$$\frac{d\gamma^\delta}{d\gamma} \propto \exp(-\Phi(u; \delta)). \tag{2.5}$$

For the well-posedness of expectation under the posterior measure, we recall the following results in [20] and [21] which slightly generalize [8] in allowing the function G below to be only square summable. We work under the following assumption.

Assumption 2.2. *The potential function* Φ *in* (2.4) *satisfies:*

(i) *For each* $\lambda > 0$ *there is a constant* $\Lambda(\lambda) > 0$ *such that if* $|\delta| < \lambda$ *where* $|\cdot|$ *denotes the Euclidean norm in* $\mathbb{R}^k$

$$\int_U \Phi(u; \delta) d\gamma(u) < \Lambda.$$

(ii) *There is a function* $G:\mathbb{R}\times U\to\mathbb{R}$ *so that for each* $\lambda>0$, $G(\lambda,\cdot)\in L^2(U,\gamma)$ *and for all* $\delta,\delta'\in\mathbb{R}^k$ *with* $|\delta|,|\delta'|<\lambda$ *we have*

$$|\Phi(u;\delta)-\Phi(u;\delta')|\le G(\lambda,u)|\delta-\delta'|.$$

The data-to-posterior map $\delta\mapsto\gamma^\delta$ is Lipschitz in the Hellinger metric.

Proposition 2.3. [20, 21] *Under Assumption 2.2 the posterior* γ^δ *is locally Lipschitz with respect to the Hellinger distance: for each* $\lambda>0$ *in Assumption 2.2(i), there exists a positive constant* $C=C(\lambda)$ *so that*

$$d_{\mathrm{Hell}}(\gamma^\delta,\gamma^{\delta'})\le C|\delta-\delta'|\quad\forall|\delta|,|\delta'|<\lambda. \tag{2.6}$$

2.2. **MCMC.** Let $g:U\to\mathbb{R}$ be γ^δ-measurable. The expectation $\mathbb{E}^{\gamma^\delta}[g]$ can be approximated numerically by Metropolis-Hastings MCMC sampling. Here, a Markov chain $\{u^{(k)}\}_{k=1}^\infty\subset U$ is constructed as follows: given the current state $u^{(k)}$, we draw a proposal $v^{(k)}$ from a probability distribution $q(u^{(k)},dv^{(k)})$. Let $\{w^{(k)}\}_{k\ge1}$ denote an i.i.d sequence with $w^{(1)}\sim\mathcal{U}[0,1]$ and with $w^{(k)}$ independent of both $u^{(k)}$ and $v^{(k)}$. The next state $u^{(k+1)}$ is determined by

$$u^{(k+1)}=\mathbf{1}\big(\alpha(u^{(k)},v^{(k)})\ge w^{(k)}\big)v^{(k)}+\Big(1-\mathbf{1}\big(\alpha(u^{(k)},v^{(k)})\ge w^{(k)}\big)\Big)u^{(k)} \tag{2.7}$$

where the acceptance probability is

$$\alpha(u,v)=\min(1,\frac{d\nu^\top(u,v)}{d\nu(u,v)})$$

with $\nu(du,dv)=q(u,dv)\gamma^\delta(du)$ and $\nu^\top(du,dv)=q(v,du)\gamma^\delta(dv)$. We suppose that the transition kernel q is chosen so that $\nu^\top<<\nu$, and in particular,

$$\frac{d\nu^\top(u,v)}{d\nu(u,v)}=\exp(\Phi(u;\delta)-\Phi(v;\delta)) \tag{2.8}$$

so that

$$\alpha(u,v)=\min(1,\exp(\Phi(u;\delta)-\Phi(v;\delta)))\ . \tag{2.9}$$

We note that (2.8) is not strictly necessary, although the independence and the pCN samplers satisfy (2.8), for example. Thus we choose to move from $u^{(k)}$ to $v^{(k)}$ with probability $\alpha(u^{(k)},v^{(k)})$, and to remain at $u^{(k)}$ with probability $1-\alpha(u^{(k)},v^{(k)})$.

2.3. **Uniform prior. Affine-parametric coefficient K.** Uniform prior probability measures are considered in detail in Hoang et al. [21]. We consider (2.2) with uncertain diffusion coefficients K of *affine-parametric form*

$$K(x,u)=\bar{K}(x)+\sum_{j=1}^\infty u_j\psi_j(x)\ ,\quad x\in D,\ u=(u_j)_{j\ge1}\in U\ , \tag{2.10}$$

where $\bar{K},\psi_j\in L^\infty(D)$ for $j\in\mathbb{N}$, and where the parameters u_j in $u=(u_j)_{j\ge1}$ are assumed to be independent and uniformly identically distributed in $[-1,1]$. This is phrased mathematically by the product probability space (U,Θ,γ) given by

$$U=[-1,1]^{\mathbb{N}},\quad\Theta=\bigotimes_{i=1}^\infty\mathcal{B}([-1,1])\ \text{ and }\gamma=\bigotimes_{i=1}^\infty\frac{du_i}{2} \tag{2.11}$$

where $\mathcal{B}([-1,1])$ is the Borel σ-algebra in $[-1,1]$, and where du_i denotes the Lebesgue measure in $\mathbb{R}^1$. Unless explicitly stated otherwise, we assume the set U to be endowed with the product topology.

For the coefficient K to be uniformly coercive and bounded for all $u \in U$, and for convergence rate bounds of the FE approximation of the solution P of (2.2), we impose the following assumption on the decay of the sequence $(\psi_j)_{j\geq 1}$.

Assumption 2.4. *The functions $\bar{K}, \psi_j \in L^\infty(D)$. Further, there exists a constant $\kappa > 0$ such that*

$$\sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} \leq \frac{\kappa}{1+\kappa} \bar{K}_{\min}$$

where $\bar{K}_{\min} = \text{essinf} \bar{K} > 0$.

With Assumption 2.4,

$$\forall u \in U: \qquad \frac{1}{1+\kappa}\bar{K}_{\min} \leq K(x,u) \leq \bar{K}_{\max} + \frac{\kappa}{1+\kappa}\bar{K}_{\min}$$

where $\bar{K}_{\max} = \text{esssup}_{x\in D}\bar{K}(x)$. Under Assumption 2.4, for each $u \in U$ the parametric forward problem (2.2) admits a unique solution. The parametric solution map $P : U \to V : u \mapsto K(\cdot, u)$ is continuous as U is endowed with the product topology. Hoang et al. [21] show that under Assumption 2.4, the forward functional $\mathcal{G}$ in (2.3) is measurable and Assumption 2.2 holds. Thus, from Propositions 2.1 and 2.3 we have:

Proposition 2.5. *For the coefficient K in* (2.10)*, under Assumption 2.4, with the probability space (U, Θ, γ) of the prior γ as defined in* (2.11)*, the posterior γ^δ is absolutely continuous with respect to the prior γ. Moreover, it depends locally Lipschitz on the data $\delta \in \mathbb{R}^k$.*

The Radon-Nikodym derivative admits a density as in (2.5) *and the Lipschitz estimate* (2.6) *holds.*

2.4. **Gaussian prior. Log-affine coefficient K.** In this section, we present the Bayesian inverse problem with gaussian prior, for diffusion problems with "log-gaussian coefficients", i.e., K in (2.2) is such that $\log K$ is a gaussian random field. The gaussian measure on realizations of the GRF $R = \log K$ will serve as prior in the corresponding Bayesian inverse problem. The numerical analysis of the forward problem (2.2) for a GRF $R = \log K$ in D was studied in detail in Galvis and Sarkis [14], Hoang and Schwab [19], Gittelson [15], Charrier [7] and in the references there. We review some of the results in these references to the extent that they are required in our ensuing MLMCMC-FEM convergence analysis. Denote by $\mathbb{R}^{\mathbb{N}}$ the set of all infinite sequences $(u_1, u_2, \ldots)$ of real numbers. Let $\{\psi_j\}_{j\geq 1} \subset L^\infty(D)$ be such that $\sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)}$ is finite. Ignoring for now the questions of convergence, we formally introduce the parametric, deterministic coefficient $K : D \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ as

$$K(\cdot, u) = K_*(\cdot) + \exp\left(\bar{K}(\cdot) + \sum_{j=1}^{\infty} u_j \psi_j(\cdot)\right) \tag{2.12}$$

for $u = (u_1, u_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$. To specify a prior probability measure on the coefficient space, we assume that the coordinates u_j are independently, identically distributed according to the standard Gaussian measure, i.e. $u_j \sim N(0,1)$. We denote by γ_1 the standard Gaussian measure in $\mathbb{R}^1$. We equip $\mathbb{R}^{\mathbb{N}}$ with the product σ-algebra $\otimes_{j=1}^{\infty}\mathcal{B}(\mathbb{R})$ where $\mathcal{B}$ denotes the Borel σ-algebra on $\mathbb{R}$. The gaussian probability measure γ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ is the product measure (see, e.g., [35, 6]), i.e.

$$\gamma = \bigotimes_{j=1}^{\infty} \gamma_1 \ . \tag{2.13}$$

For K to be a valid diffusion coefficient, γ-a.s., we impose the following assumption on the functions K_*, $\bar{K}$ and ψ_i.

Assumption 2.6. *The functions $\bar{K}$, K_* and ψ_j in* (2.12) *are in $L^\infty(D)$ and there holds $0 \le \mathrm{essinf} K_*(x)$, and $\boldsymbol{b} := (\|\psi_j\|_{L^\infty(D)})_{j\ge1} \in \ell^1(\mathbb{N})$.*

We emphasize that in Assumption 2.6, $K_* = 0$ is admissible. Assumption 2.6 implies that the set

$$\Gamma_{\boldsymbol{b}} := \{u = (u_j)_{j\ge1} \in \mathbb{R}^{\mathbb{N}}, \ \sum_{j=1}^{\infty} b_j |u_j| < \infty\} \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}) \tag{2.14}$$

has full Gaussian measure, i.e. $\gamma(\Gamma_{\boldsymbol{b}}) = 1$ (see, e.g., [35, p. 153] or [31, Lemma 2.28]). Here, $b_j := \|\psi_j\|_{L^\infty(D)}$. For every $u \in \Gamma_{\boldsymbol{b}}$, the coefficient (2.12) is well-defined as an element of $L^\infty(D)$, and we observe that $\Gamma_{\boldsymbol{b}}$ is in general not a cartesian product of intervals.

Let $\mathcal{A}_{\boldsymbol{b}}$ denote the restriction of the product σ algebra $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ to $\Gamma_{\boldsymbol{b}} \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and let $\gamma_{\boldsymbol{b}}$ denote the restriction of the Gaussian Measure γ to $\Gamma_{\boldsymbol{b}}$. For $u \in \Gamma_{\boldsymbol{b}}$, we define

$$\hat{K}(u) = \mathrm{esssup}_{x\in D} K_*(x) + \exp(\|\bar{K}\|_{L^\infty(D)} + \sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} |u_j|), \tag{2.15}$$

and

$$\check{K}(u) = \mathrm{essinf}_{x\in D} K_*(x) + \exp(\mathrm{essinf}_{x\in D} \bar{K}(x) - \sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} |u_j|) \ . \tag{2.16}$$

For $u \in \Gamma_{\boldsymbol{b}}$ and for $x \in D\backslash\mathcal{N}$ where $\mathcal{N} \subset D$ is a (Lebesgue) nullset, $0 < \check{K}(u) \le K(x,u) \le \hat{K}(u) < \infty$. We observe that $\hat{K}(u)$ and $\check{K}(u)$ are $(\Gamma_{\boldsymbol{b}}, \mathcal{A}_{\boldsymbol{b}})$ measurable. For every $u \in \Gamma_{\boldsymbol{b}}$, the diffusion problem (2.2) admits a unique solution $P(\cdot, u) \in V$.

The solution P of (2.2), when interpreted as a map from $(\Gamma_{\boldsymbol{b}}, \mathcal{A}_{\boldsymbol{b}})$ to $(V, \mathcal{B}(V))$, is strongly measurable (see, for example, in [15] and [14, 7]) so that the forward functional $\mathcal{G}$ is measurable in the measurable space $(\Gamma_{\boldsymbol{b}}, \mathcal{A}_{\boldsymbol{b}})$. Further, Hoang and Schwab [19] show that under Assumption 2.6, Assumption 2.2 holds. From Propositions 2.1 and 2.3, we have

Proposition 2.7. *Under Assumption 2.6, for the log-gaussian coefficient K defined in* (2.12) *and with the prior probability space $(\Gamma_{\boldsymbol{b}}, \mathcal{A}_{\boldsymbol{b}}, \gamma_{\boldsymbol{b}})$, the posterior probability measure γ^δ is absolutely continuous with respect to the prior measure γ. The map $\delta \mapsto \gamma^\delta$ is locally Lipschitz. Formula* (2.5) *for the Radon-Nikodym derivative and the local Lipschitz estimate* (2.6) *hold.*

3. Multilevel Markov Chain Monte Carlo FEM for uniform prior

We recall in this section the MLMCMC FEM developed in Hoang et al. in [21]. We first summarize the approximation of the forward problem (2.2) with the coefficient (2.10) obtained by a finite truncation of the infinite series representation of K and by subsequent numerical solution of the finite-parametric PDE by finite element discretization.

3.1. **Finite element discretization of the forward problem.** For each $J \in \mathbb{N}$, we consider the J-term truncated, parametric coefficient

$$K^J(\cdot, u) = \bar{K}(\cdot) + \sum_{j=1}^{J} u_j \psi_j(\cdot). \tag{3.1}$$

The forward problem (2.2) with the coefficient K in (2.10) is approximated by the "dimension-truncated", finite-parametric problem

$$-\nabla \cdot (K^J(\cdot, u) \nabla P^J(\cdot, u)) = f, \quad P^J \in V. \tag{3.2}$$

We approximate the solution of (3.2) numerically by a standard, primal FE discretization. To this end, we assume that D is a bounded polytope with plane sides (if $d = 2$) resp. plane faces (if $d = 3$). We consider in D a nested sequence $\{\mathcal{T}^l\}_{l=0}^{\infty}$ of regular, simplicial triangulations of D; each triangulation $\mathcal{T}^l$ is obtained by uniform refinement, i.e., by dividing each simplex in $\mathcal{T}^{l-1}$ into 4 congruent triangles when $d = 2$ or into 8 tedrahedra when $d = 3$. We define a nested sequence $\{V^l\}_{l \geq 1}$ of spaces of continuous, piecewise linear functions on $\mathcal{T}^l$ as

$$V^l = \{w \in V : \; w|_T \in \mathbb{P}^1(T) \; \forall \, T \in \mathcal{T}^l\},$$

where $\mathbb{P}^1(T)$ is the set of linear polynomials in T. The finite element approximation is then defined by Galerkin projection: for $u \in U$, find $P^{J,l} \in V^l$ such that

$$\int_D K^J(x, u) \nabla P^{J,l}(x, u) \cdot \nabla \phi(x) dx = \int_D f(x) \phi(x) dx, \quad \forall \, \phi \in V^l. \tag{3.3}$$

For the solution $P^J(\cdot, u)$ of the parametric PDE (3.2) to belong to $H^2(D)$, we impose the folowing regularity on the coefficients in the expansion (2.10).

Assumption 3.1. *The functions $\bar{K}$ and ψ_j ($j = 1, 2, \ldots$) in* (2.10) *belong to $W^{1,\infty}(D)$ and $\sum_{j=1}^{\infty} \|\psi_j\|_{W^{1,\infty}(D)}$ is finite. Moreover, there exist constants $C > 0$ and $s > 1$ so that $\|\psi_j\|_{L^\infty} < Cj^{-s}$ for all $j \in \mathbb{N}$.*

Under Assumption 3.1, for $f \in L^2(D)$ and D a convex polygon, Hoang et al. [21] show that $P^J(\cdot, u) \in H^2(D) \cap V$, with $\sup_{u \in U} \sup_J \|P^J(\cdot, u)\|_{H^2(D)}$ being bounded. This allows establishing the following error estimate for the approximate solution $P^{J,l}$ in (3.3).

Proposition 3.2. *Assume that the domain D is convex, and that $f \in L^2(D)$. Under Assumption 3.1, there exists a constant $C > 0$ such that for every $J, l \in \mathbb{N}$*

$$\|P - P^{J,l}\|_V \leq C(J^{-q} + 2^{-l}) \|f\|_{L^2(D)} \tag{3.4}$$

where $q = s - 1$.

Remark 3.3. *Assumption 3.1 could be weakened considerably, with error bounds such as* (3.4) *still valid: we could admit non-convex polytopal $D \subset \mathbb{R}^d$, with appropriately refined triangulations $\mathcal{T}^l$ in D, and suitable assumptions on higher regularity of the ψ_j; this would require introduction of weighted Sobolev and Hölder spaces in order to state regularity and FE error estimates. All subsequent results have straightforward extensions in these more general settings. As the focus of the present paper is on the analysis of the MLMCMC algorithms, we chose to impose the rather restrictive conditions in Assumption 3.1 to keep the PDE error analysis as simple as possible.*

3.2. **Finite element approximation of the Bayesian posterior.** With the approximate solution $P^{J,l}(\cdot,u)$ of problem (3.3) we associate the approximate forward map

$$\mathcal{G}^{J,l}(u) = (\mathcal{O}_1(P^{J,l}(\cdot,u)),\dots,\mathcal{O}_k(P^{J,l}(\cdot,u))). \tag{3.5}$$

The approximate Bayesian potential is defined as

$$\Phi^{J,l}(\delta,u) = \frac{1}{2}|\delta - \mathcal{G}^{J,l}(u)|_\Sigma^2, \tag{3.6}$$

and the approximate posterior on (U,Θ) is given by

$$\frac{d\gamma^{J,l,\delta}}{d\gamma} \propto \exp(-\Phi^{J,l}(\delta,u)). \tag{3.7}$$

Hoang et al. [21] in Proposition 10 prove the following result on the approximation property of the measure $\gamma^{J,l}$.

Proposition 3.4. *Under Assumptions 2.4 and 3.1, if the domain D is a convex polyhedron and $f\in L^2(D)$, then there is a constant C which only depends on the data bound λ in Assumption 2.2 so that for every $J,l\in\mathbb{N}$ holds*

$$d_{\text{Hell}}(\gamma^\delta,\gamma^{J,l,\delta}) \le C(J^{-q}+2^{-l})\|f\|_{L^2(D)}.$$

To balance the two errors stemming from truncating the coefficient K at J terms as in (3.1) and from the FE discretization (3.3) at mesh level $l\ge 0$ we choose for $q=s-1>0$ as in Assumption 3.1

$$\forall l\in\mathbb{N}: \qquad J = J_l = \lceil 2^{l/q}\rceil\ . \tag{3.8}$$

3.3. **Multilevel Markov Chain Monte Carlo for the uniform prior.** We recapitulate from Hoang et al. [21] the derivation of the Multilevel MCMC FEM for the uniform prior. For conciseness, with the choice $J_l=\lceil 2^{l/q}\rceil$, we denote $\gamma^{J_l,l,\delta}$ simply as γ^l and $P^{J_l,l}$ as P^l. For every $L\in\mathbb{N}$, there holds the telescoping sum

$$\begin{aligned}\mathbb{E}^{\gamma^L}[\ell(P(\cdot,u))] &= \sum_{l=1}^{L}\Big(\mathbb{E}^{\gamma^l}[\ell(P(\cdot,u))] - \mathbb{E}^{\gamma^{l-1}}[\ell(P(\cdot,u))] + \mathbb{E}^{\gamma^0}[\ell(P(\cdot,u))]\Big)\\ &= \sum_{l=1}^{L}\Big(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\Big)[\ell(P(\cdot,u))] + \mathbb{E}^{\gamma^0}[\ell(P(\cdot,u))]\ . \end{aligned} \tag{3.9}$$

For each l, with a discretization level $L'(l)\le L$ to be determined subsequently, we approximate $\mathbb{E}^{\gamma^L}[\ell(P(\cdot,u))]$ by (omitting the arguments of P and its approximations for brevity of notation) the telescoping sum

$$\sum_{l=1}^{L}\Big(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\Big)[\ell(P^{L'(l)})] + \mathbb{E}^{\gamma^0}[\ell(P^{L'(0)})]. \tag{3.10}$$

For each l, this results in

$$\Big(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\Big)[\ell(P^{L'(l)})] = \sum_{l'=1}^{L'(l)}\Big(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\Big)[\ell(P^{l'})-\ell(P^{l'-1})] + \Big(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\Big)[\ell(P^0)]\ . \tag{3.11}$$

Similarly,

$$\mathbb{E}^{\gamma^0}[\ell(P^{L'(0)})] = \sum_{l'=1}^{L'(0)} \mathbb{E}^{\gamma^0}[\ell(P^{l'}) - \ell(P^{l'-1})] + \mathbb{E}^{\gamma^0}[\ell(P^0)].$$

Thus, for each level L of approximation, there holds

(3.12)

$$\begin{aligned} & \sum_{l=1}^{L} \left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[\ell(P^{L'(l)})] + \mathbb{E}^{\gamma^0}[\ell(P^{L'(0)})] \\ = \; & \sum_{l=1}^{L}\sum_{l'=1}^{L'(l)} \left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[\ell(P^{l'}) - \ell(P^{l'-1})] + \sum_{l=1}^{L}\left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[\ell(P^0)] \\ & + \sum_{l'=1}^{L'(0)} \mathbb{E}^{\gamma^0}[\ell(P^{l'}) - \ell(P^{l'-1})] + \mathbb{E}^{\gamma^0}[\ell(P^0)]. \end{aligned}$$

To obtain a computable MLMCMC estimator we approximate each term in (3.12) by sample averages of $M_{ll'}$ many realizations, upon choosing $L'(l)$ judiciously. To select $M_{ll'}$ and $L'(l)$, we observe that, for any measurable function $Q : U \to \mathbb{R}$ which is integrable with respect to the approximate posterior measures γ^l, there holds

$$\begin{aligned} & \left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[Q] \\ & = \frac{1}{Z^l}\int_U \exp(-\Phi^l(u;\delta))Q(u)d\gamma(u) - \frac{1}{Z^{l-1}}\int_U \exp(-\Phi^{l-1}(u;\delta))Q(u)d\gamma(u) \\ & = \frac{1}{Z^l}\int_U \exp(-\Phi^l(u;\delta))\left(1 - \exp(\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta))\right)Q(u)d\gamma(u) \\ & \qquad + \left(\frac{Z^{l-1}}{Z^l} - 1\right)\frac{1}{Z^{l-1}}\int \exp(-\Phi^{l-1}(u;\delta))Q(u)d\gamma(u)\;, \end{aligned}$$

where Φ^l denotes the approximate potential defined in (3.6) and Z^l denotes the approximate normalizing constant in (3.7) with $J = J_l$. We remark that under Assumptions 2.2 and 2.4, the normalization constants Z^l are uniformly (with respect to l) bounded from below away from zero. We note further that

$$\frac{Z^{l-1}}{Z^l} - 1 = \frac{1}{Z^l}\int_U \left(\exp(\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta)) - 1\right)\exp(-\Phi^l(u;\delta))d\gamma(u)\;.$$

Thus an approximation for $Z^{l-1}/Z^l - 1$ can be computed by running MCMC with respect to the approximate posterior γ^l to sample the potential difference $\exp(\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta)) - 1$. To estimate the expectation with respect to the approximated posteriors γ^l and γ^{l-1}, we apply the MCMC algorithm introduced in Section 2. The acceptance probability $\alpha(u,v)$ in (2.9) is, however, replaced by that derived from the FE solution of the finitely-parametric forward problem. In particular, we define by $E^{\gamma^l}_{M_{ll'}}$ the MCMC FEM estimator obtained with the acceptance probability in (2.9) replaced by

$$\alpha^l(u,v) = \min(1, \exp(\Phi^l(u;\delta) - \Phi^l(v;\delta))). \tag{3.13}$$

for the MCMC procedure to sample the target probability measure γ^l.

This led in [21] to the *Multilevel Markov Chain Monte Carlo Finite Element* (MLMCMC-FEM for short) estimator $E_L^{MLMCMC}[\ell(P)]$ of $\mathbb{E}^{\gamma^\delta}[\ell(P)]$ defined by

(3.14)

$$
\begin{aligned}
&E_L^{MLMCMC}[\ell(P)] = \\
&\sum_{l=1}^{L}\sum_{l'=1}^{L'(l)} E_{M_{ll'}}^{\gamma^l}\left[\Big(1-\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))\Big)(\ell(P^{l'})-\ell(P^{l'-1}))\right] \\
&+\sum_{l=1}^{L}\sum_{l'=1}^{L'(l)} E_{M_{ll'}}^{\gamma^l}\left[\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1\right]\cdot E_{M_{ll'}}^{\gamma^{l-1}}\left[\ell(P^{l'})-\ell(P^{l'-1})\right] \\
&+\sum_{l=1}^{L} E_{M_{l0}}^{\gamma^l}\left[\Big(1-\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))\Big)(\ell(P^0))\right] \\
&+\sum_{l=1}^{L} E_{M_{l0}}^{\gamma^l}\left[\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1\right]\cdot E_{M_{l0}}^{\gamma^{l-1}}\left[\ell(P^0))\right] \\
&+\sum_{l'=1}^{L'(0)} E_{M_{0l'}}^{\gamma^0}\left[\ell(P^{l'})-\ell(P^{l'-1})\right]+E_{M_{00}}^{\gamma_0}[\ell(P^0)].
\end{aligned}
$$

As in Hoang et al. [21], we choose the parameters

$$
L'(l)=L-l,\quad M_{ll'}=2^{2(L-(l+l'))}. \tag{3.15}
$$

When evaluating the MLMCMC estimator for each approximation level l, we generate a Markov chain $\mathcal{C}_l\subset\mathbb{R}^{J_l}$. In this way, we realize L pairwise uncorrelated chains. We denote the probability space of these L Markov chains by $\mathbf{C}_L=\{\mathcal{C}_0,\mathcal{C}_1,\mathcal{C}_2,\dots,\mathcal{C}_L\}$, and let $\mathcal{P}^{\gamma,J_l,l}$ denote the probability measure in the space of all Markov chains on the discretized PDE at mesh level l with parameter dimension J_l, running from the initial sample $u^{(0)}$ which we assume to be distributed according to the prior γ. By pairwise independence of the chains $\mathcal{C}_l$, the probability measure on $\mathbf{C}_L$ is

$$
\mathbf{P}_L=\mathcal{P}^{\gamma,J_0,0}\otimes\mathcal{P}^{\gamma,J_1,1}\otimes\mathcal{P}^{\gamma,J_2,2}\otimes\ldots\otimes\mathcal{P}^{\gamma,J_L,L}.
$$

Let $\mathbf{E}_L$ be the expectation in $\mathbf{C}_L$ with respect to $\mathbf{P}_L$. The multilevel MCMC method achieves an approximation with a prescribed absolute mean error using an essentially optimal number of degrees of freedom for any fixed basis of the FE space V^l. A Riesz basis of the V^l affords essentially linear complexity per MCMC sample of the discretized PDE. Then, Hoang et al. [21] show for uniform prior that the MLMCMC FEM uses an asymptotically optimal number of floating point operations.

To develop corresponding results under gaussian prior, we impose the following assumption on the availability of a Riesz finite element basis. This assumption is valid for polytopal domains in space dimensions $d=1,2,3$. Construction of the Riesz basis can be found in [27] and [34].

Assumption 3.5. *For $l \in \mathbb{N}$, there is a set of indices $\mathcal{I}^l \subset \mathbb{N}^d$ of cardinality $N_l = O(2^{-dl})$ and a family of basis functions $w_k^l \in V$ with $k \in \mathcal{I}^l$ such that V^l is the linear span of w_k^l for $k \in \mathcal{I}^l$. Furthermore, there are positive constants c_1 and c_2 which are independent of l such that for $w = \sum_{k\in I^l} c_k^l w_k^l \in V^l$ holds*

$$c_1 \sum_{k\in\mathcal{I}^l} (c_k^l)^2 \le \|w\|_V^2 \le c_2 \sum_{k\in\mathcal{I}^l} (c_k^l)^2.$$

For all $l \in \mathbb{N}_0$ and all $k \in \mathcal{I}^l$, for each $l' \in \mathbb{N}_0$, $\mathrm{supp}(w_k^l)\cap\mathrm{supp}(w_{k'}^{l'})$ has positive mesure for at most $O(\max(1, 2^{l'-l}))$ functions $w_{k'}^{l'}$ for $k' \in \mathcal{I}^{l'}$.

Hoang et al. [21] establish the following result on the error and complexity of the MLMCMC procedure for sampling the posterior measure γ^δ corresponding to uniform prior probability measure γ, using the independence sampler.

Theorem 3.6. *For $d = 2, 3$, under Assumption 3.1, and with the parameter choices* (3.15)*, there exists a constant $c(\delta) > 0$ such that for all $L \ge 1$ there holds*

$$\mathbf{E}_L[|\mathbb{E}^{\gamma(\delta)}[P] - E_L^{MLMCMC}[P]|] \le C(\delta)L^2 2^{-L} \; . \tag{3.16}$$

The total number of degrees of freedom in the FE discretization that is used in running the MLMCMC sampler is bounded by $O(L2^{2L})$ for $d = 2$ and $O(2^{3L})$ for $d = 3$.

Under Assumption 3.5 on the availability of a Riesz finite element basis, the total number of floating point operations required for computing the MLMCMC estimator is bounded by $O(L^{d-1}2^{(d+1/q)L})$.

Denoting the total number of degrees of freedom which enter in running the chain on all discretization levels by N, the error of the MLMCMC estimator is bounded by $O((\log N)^{3/2}N^{-1/2})$ for $d = 2$ and by $O((\log N)^2 N^{-1/3})$ for $d = 3$.

The total number of floating point operations used in running the MLMCMC-FEM algorithm to termination is bounded by $O((\log N)^{-1/(2q)}N^{1+1/(2q)})$ for $d = 2$ and by $O((\log N)^2 N^{1+1/(3q)})$ for $d = 3$.

The logarithmic factor L^2 in the error bound, may be reduced by slightly increasing the sample numbers $M_{ll'}$. The following error bounds are a refinement of those in [21]: choosing in (3.14) the sample numbers $M_{ll'} = (l + l')^\alpha 2^{2(L-(l+l'))}$ when $l \ge 1$ and $l' \ge 1$, the error due to the first two terms of (3.14) is bounded by an absolute multiple of $2^{-L}\sum_{l,l'=1}^{L}(l + l')^{-\alpha/2}$. The following table collects the resulting asymptotic bounds, for various values of α.

α	$M_{ll'}, l, l' > 1$	$M_{l0} = M_{0l}$	M_{00}	Total error
0	$2^{2(L-(l+l')}$	$2^{2(L-l)}/L^2$	$2^{2L}/L^4$	$O(L^2 2^{-L})$
2	$(l + l')^2 2^{2(L-(l+l'))}$	$2^{2(L-l)}$	$2^{2L}/L^2$	$O(L \log L 2^{-L})$
3	$(l + l')^3 2^{2(L-(l+l'))}$	$l2^{2(L-l)}$	$2^{2L}/L$	$O(L^{1/2} 2^{-L})$
4	$(l + l')^4 2^{2(L-(l+l'))}$	$l^2 2^{2(L-l)}$	$2^{2L}/(\log L)^2$	$O(\log L 2^{-L})$

TABLE 1.

4. Multilevel Markov Chain Monte Carlo Finite Element Method (MLMCMC-FEM) for Gaussian prior

We develop the MLMCMC for sampling the posterior measure γ^δ when the coefficient K is of the form (2.12) with the probability space $U = \Gamma_{\boldsymbol{b}}$ defined in (2.14) and prior probability $\gamma = \gamma_{\boldsymbol{b}}$. We show by a numerical example in Section 5 that the MLMCMC algorithm in the previous section may diverge for problems with coefficients of the log-gaussian form. Essential modifications in the algorithm are necessary for the MLMCMC method to work in the case of Gaussian prior.

4.1. **FE approximation of diffusion problem with log-gaussian coefficients.**

We again approximate the forward equation (2.2) by truncating the coefficient and by discretizing the resulting, finitely-parametric equation by the FEM. To this end, we review some results established by Hoang and Schwab in [19] and refer to [19, Section 4] for proofs.

For $u = (u_1, u_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$ and for $J \in \mathbb{N}$, denote $u^J = (u_1, u_2, \ldots, u_J, 0, 0, \ldots)$, i.e., the "$J$-term anchored" parameter sequence, and define

$$K^J(\cdot, u) = K(\cdot, u^J) = K_*(\cdot) + \exp\left(\bar{K}(\cdot) + \sum_{j=1}^{J} u_j \psi_j(\cdot)\right). \tag{4.1}$$

For each $u \in \Gamma_{\boldsymbol{b}}$, the parametric equation with J-term truncated coefficient reads

$$P^J \in V: \quad -\nabla \cdot (K^J(\cdot, u) \nabla P^J(\cdot, u)) = f \text{ in } H^{-1}(D). \tag{4.2}$$

We approximate (4.2): find $P^{J,l} \in V$ so that

$$\int_D K^J(x, u) \nabla P^{J,l}(x, u) \cdot \nabla \phi(x) dx = \int_D f(x) \phi(x) dx \quad \forall \phi \in V \tag{4.3}$$

where K^J is the J-term truncated coefficient in (4.1). This problem has a unique solution that satisfies: there exists a constant $c > 0$ such that for every $J \in \mathbb{N}$ and for every $u \in \Gamma_{\boldsymbol{b}}$ holds

$$\|P^{J,l}(\cdot, u)\|_V \leq \frac{\|f\|_{V^*}}{\check{K}(u^J)} \leq c \exp(\sum_{j=1}^{\infty} b_j |u_j|). \tag{4.4}$$

From Cea's lemma, we obtain

$$\|P^J(\cdot, u) - P^{J,l}(\cdot, u)\|_V \leq \frac{\hat{K}(u^J)}{\check{K}(u^J)} \inf_{Q \in V^l} \|P^J(\cdot, u) - Q\|_V. \tag{4.5}$$

To obtain first order convergence rates of continuous, piecewise linear FEM in D, we impose the following regularity assumption on the parametric coefficient.

Assumption 4.1. *The functions* $K_*, \bar{K}$ *and* ψ_j *in the expansion* (2.12) *belong to* $W^{1,\infty}(D)$ *and* $\bar{\boldsymbol{b}} := (\|\psi_j\|_{W^{1,\infty}(D)})_{j \geq 1} \in \ell^1(\mathbb{N})$.

Under Assumption 4.1, Hoang and Schwab [19, Section 4.2] prove that when the domain D is convex and when $f \in L^2(D)$, there exists $C > 0$ such that, for every

$u \in \Gamma_{\boldsymbol{b}}$ and $J \in \mathbb{N}$ holds

$$\|P^J(\cdot,u)\|_{H^2(D)} \le C\frac{1}{\check{K}(u^J)}(\|\nabla K^J(\cdot,u)\|_{L^\infty(D)}\frac{\|f\|_{V^*}}{\check{K}(u^J)} + \|f\|_{L^2(D)})$$
$$\le C\exp\left(3\sum_{j=1}^J b_j|u_j|\right)\left(1+\sum_{j=1}^J \bar{b}_j|u_j|\right).$$

There exists a constant $C > 0$ such that for every finite $J, l \in \mathbb{N}$ and every $u \in \Gamma_{\boldsymbol{b}}$

$$\|P^J(\cdot,u) - P^{J,l}(\cdot,u)\|_V \le C\exp\left(5\sum_{j=1}^J b_j|u_j|\right)\left(1+\sum_{j=1}^J \bar{b}_j|u_j|\right)2^{-l} \tag{4.6}$$

where C is independent of u, J and l. There holds the following bound on the error due to truncating the log-gaussian coefficient at a finite number J of parameters and including FE discretization of the forward model at level l.

Proposition 4.2. *Under Assumption 4.1, if D is convex and $f \in L^2(D)$, there exists a constant $C > 0$ such that, for every $J, l \in \mathbb{N}$ and for every $u \in \Gamma_{\boldsymbol{b}}$ holds*
(4.7)
$$\|P(\cdot,u) - P^{J,l}(\cdot,u)\|_V \le C\exp\left(5\sum_{j=1}^J b_j|u_j|\right)\left(2^{-l}(1+\sum_{j=1}^J \bar{b}_j|u_j|) + \sum_{j=J+1}^\infty b_j|u_j|\right).$$

We refer to [19] (Lemma 4.4) for the proof. To estimate the complexity of solving the linear system in (4.3), we note the following result of [19, Section 4.5].

Lemma 4.3. *For a constant $C > 0$ independent of l and u, under Assumption 3.5, for any $u \in \Gamma_{\boldsymbol{b}}$ and for every $l \in \mathbb{N}$, the number $j^*(u,l)$ of pcg-iterations in the approximate, iterative solution of the parametric linear system of equations in* (4.3) *for the corresponding iteration error to be bounded by 2^{-l} (in euclidean norm and, by the norm equivalence Assumption 3.5, in the norm $H^1(D)$) is bounded from below by*

$$j^*(u,l) \ge C(l + |\log \check{K}(u)|)\sqrt{\frac{\hat{K}(u)}{\check{K}(u)}}. \tag{4.8}$$

We remark that (4.8) depends on the realization $u \in \Gamma_{\boldsymbol{b}}$ of the GRF Z, so that error vs. work estimates based on (4.8) will only hold "in expectation", or "in the mean". We also point out that in [19] availability of a particular preconditioner was assumed which is constructed via Riesz-basis of the FE spaces. An alternative approach, based on multilevel preconditioning in standard FE bases with estimates similar to (4.8) and admitting gaussian prior distribution on the parameters was developed in [18].

4.2. **Finite element approximation of the posterior measure.** For the FE solution $P^{J,l}$ in (3.3) with the parameter-truncated log-gaussian coefficient K^J in (4.1), the approximate forward operator $\mathcal{G}^{J,l}$ is denoted as

$$\mathcal{G}^{J,l}(u) = (\mathcal{O}_1(P^{J,l}(u)), \dots, \mathcal{O}_k(P^{J,l}(u))) : \ \Gamma_{\boldsymbol{b}} \to \mathbb{R}^k. \tag{4.9}$$

With this approximate forward mapping, the approximate Bayesian potential $\Phi^{J,l}$ is

$$\Phi^{J,l} := \frac{1}{2}|\delta - \mathcal{G}^{J,l}|^2_\Sigma \tag{4.10}$$

and the corresponding approximate Bayesian posterior (3.6) is

$$\frac{d\gamma^{J,l,\delta}}{d\gamma_{\boldsymbol{b}}}(u) \propto \exp(-\Phi^{J,l}(u;\delta)). \tag{4.11}$$

To quantify the error in approximating the Bayesian posterior γ^δ by $\gamma^{J,l,\delta}$, we make the following assumption on the decay rate of $\|\psi_j\|_{L^\infty(D)}$.

Assumption 4.4. *There are $c > 0$ and $s > 1$ such that $\|\psi_j\|_{L^\infty(D)} \le cj^{-s}$.*

We then have

Proposition 4.5. *Under Assumption 4.4, for $q = s - 1 > 0$ there is a positive constant c depending on δ such that for every $J, l \in \mathbb{N}$ holds*

$$d_{\text{Hell}}(\gamma^\delta, \gamma^{J,l,\delta}) \le c(J^{-q} + 2^{-l}).$$

For proofs of these results we refer to [19, Section 4.6]. Choosing $J = J_l = \lceil 2^{l/q} \rceil$, we obtain that the Hellinger distance between the Bayesian posteriors of the exact forward solution and its FE approximation converges as the FE discretization error: $d_{\text{Hell}}(\gamma^\delta, \gamma^{J,l,\delta}) \le c2^{-l}$. For the Multilevel MCMC method that we will develop ahead, we estimate $|\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)|$.

Lemma 4.6. *For every $\lambda > 0$, there is a constant $c(\lambda)$ that depends only on the data bound λ in Assumption 2.2 such that, for every $l \in \mathbb{N}$ and for every $|\delta| < \lambda$,*

$$|\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)| \le c(\lambda) \exp\left(6\sum_{j=1}^{J_l} b_j|u_j|\right)\left(2^{-l}(1 + \sum_{j=1}^{J_l} \bar{b}_j|u_j|) + \sum_{j=J_{l-1}+1}^{J_l} b_j|u_j|\right).$$

Proof The proof follows the argument in the analysis of [19] e.g. the proof of Lemma 4.4 of [19]. From (3.6), we obtain the existence of a constant $c > 0$ such that for every $u \in \Gamma_{\boldsymbol{b}}$ and for every $l \in \mathbb{N}$ holds

$$|\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)| \le c(|\delta| + |\mathcal{G}^{J_l,l}(u)| + |\mathcal{G}^{J_{l-1},l-1}(u)|)|\mathcal{G}^{J_l,l}(u) - \mathcal{G}^{J_{l-1},l-1}(u)|.$$

We note that

$$|\mathcal{G}^{J_l,l}(u) - \mathcal{G}^{J_{l-1},l-1}(u)| \le c\max\{\|\mathcal{O}_i\|_{V^*}\}\|P^{J_l,l}(\cdot,u) - P^{J_{l-1},l-1}(\cdot,u)\|_V.$$

We have from (4.2) that

$$-\nabla\cdot(K^{J_l}(\cdot,u)\nabla(P^{J_l}(\cdot,u) - P^{J_{l-1}}(\cdot,u)) = -\nabla\cdot((K^{J_l}(\cdot,u) - K^{J_{l-1}}(\cdot,u))\nabla P^{J_l}(\cdot,u)).$$

Therefore

$$\begin{aligned}\|P^{J_l}(\cdot,u) - P^{J_{l-1}}(\cdot,u)\|_V &\le \frac{1}{\check{K}(u^{J_l})}\|K^{J_l}(\cdot,u) - K^{J_{l-1}}(\cdot,u)\|_{L^\infty(D)}\|P^{J_l}(\cdot,u)\|_V \\ &\le \frac{1}{\check{K}(u^{J_l})\check{K}(u^{J_l})}\|K^{J_l}(\cdot,u) - K^{J_{l-1}}(\cdot,u)\|_{L^\infty(D)}\|f\|_{V^*}.\end{aligned}$$

Using the inequality $|e^x - e^y| \le |x-y|(e^x + e^y)$ for $x, y \in \mathbb{R}$, we have

$$\|K^{J_l}(\cdot,u) - K^{J_{l-1}}(\cdot,u)\|_{L^\infty(D)} \le 2\exp(b_0 + \sum_{j=1}^{J_l} b_j|u_j|)\sum_{j=J_{l-1}+1}^{J_l} b_j|u_j|.$$

Thus there exists a constant C such that for all $J, l \in \mathbb{N}$

$$\|P^{J_l}(\cdot,u) - P^{J_{l-1}}(\cdot,u)\|_V \le C\exp\left(3\sum_{j=1}^{J_l} b_j|u_j|\right)\sum_{j=J_{l-1}+1}^{J_l} b_j|u_j|.$$

From this and (4.6) we deduce that
(4.12)

$$\|P^{J_l,l}(\cdot,u)-P^{J_{l-1},l-1}(\cdot,u)\|_V \le C\exp\left(5\sum_{j=1}^{J_l} b_j|u_j|\right)\left(2^{-l}(1+\sum_{j=1}^{J_l}\bar{b}_j|u_j|)+\sum_{j=J_{l-1}+1}^{J_l} b_j|u_j|\right).$$

Together with (4.4) this gives the conclusion. □

4.3. **Multilevel Markov Chain Monte Carlo FEM.** The proof of Theorem 3.6 for the Multilevel MCMC FEM for uniform prior in [21] uses in an essential way that the potential $\Phi(u,\delta)$ and its approximation $\Phi^{J_l,l}(u,\delta)$ are uniformly bounded for all $u \in U = [-1,1]^{\mathbb{N}}$. For the log-gaussian coefficient K in (2.12) this is no longer true. We derive the MLMCMC FEM for the gaussian Bayesian prior. While the algorithm is structured as for the uniform prior in Section 3, there are essential differences to it and to the plain (i.e., single-level) MCMC-FEM for gaussian prior. This is due to the fact that the differences $\Phi^{J_l,l}(u,\delta) - \Phi^{J_{l-1},l-1}(u,\delta)$ for successive discretization levels l in MLMCMC-FEM grow, generally, exponentially as $|u| \to \infty$. This is due to the (sharp) bounds (2.15), (2.16). This exponential growth and the structure of the Bayesian posterior density (2.5) imply a possibly doubly-exponential growth of this density on increments in the MLMCMC for gaussian prior which results in a parametric density which is not integrable against the gaussian prior. This issue does not arise in the MCMC-FEM for the single-level discretization under gaussian prior which was analyzed in [19] and, under stronger assumptions than Assumption 2.6, in [28].

Let us turn to the derivation of the MLMCMC-FEM strategy. We run the MCMC as in the uniform case, but in the case of Gaussian prior, we consider both the independence and pCN samplers, with acceptance probability α in (2.9), evaluated with the FE discretization (3.3) of the forward problem. We use the acceptance probability

$$\alpha^{J,l}(u,v) = 1 \wedge \exp(\Phi^{J,l}(u,\delta) - \Phi^{J,l}(v,\delta)) \tag{4.13}$$

where $\Phi^{J,l}$ is determined from the FE solution of the forward equation truncated at J terms of the input expansion with the diffusion equation with log-gaussian input in (4.10). We denote the MCMC sample average for approximating the posterior expectation of a function g as $E^{\gamma^{J,l}}_{M_{ll'}}[g]$. In this section, we always choose J_l as in (3.8). Again, to simplify notation, we denote $\gamma^{J_l,l,\delta}$ by γ^l, $P^{J_l,l}$ by P^l and $\Phi^{J_l,l}$ by Φ^l.

Let $\ell \in V^*$. The derivation of the MLMCMC FEM to approximate the expectation of $\ell(P(\cdot))$ with respect to the posterior probability measure γ^δ on U is as follows.

From Proposition 4.5, we obtain the existence of a constant $c > 0$ such that, for every $L \in \mathbb{N}$, there holds

$$\left|\mathbb{E}^{\gamma^\delta}[\ell(P(u))] - \mathbb{E}^{\gamma^L}[\ell(P(u))]\right| \le c2^{-L}.$$

We recall the telescoping sum (3.12). To obtain convergence rate bounds for the multilevel MCMC-FEM under uniform prior in Section 3.3, it essential that $|\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta)|$ be uniformly bounded with respect to l. For the presently considered log-gaussian coefficient and for the gaussian prior measure, this is no longer the case: under Assumption 2.6 (which admits $K_* \equiv 0$ in (2.12)), *in the* *multilevel algorithm for uniform prior in Section 3.3, this quantity may not even be integrable with respect to the gaussian prior measure γ.*

As we show with a numerical example in Section 5, under gaussian prior and under Assumption 2.6 the MLMCMC-FEM estimator (3.14) diverges, in general. Therefore, even the design of the MLMCMC-FEM algorithm will require essential modifications as compared to uniform prior in [21]. To address this, we propose a new method for sampling the terms in (3.12). To this end, we denote by

$$I^l(u) = \begin{cases} 1 & \text{if} \quad \Phi^l(u;\delta) - \Phi^{l-1}(u;\delta) \le 0, \\ 0 & \text{otherwise} . \end{cases} \tag{4.14}$$

Let Q be a measurable function from U to $\mathbb{R}$. There holds, for $l \ge 1$,

$$\begin{aligned}
&\mathbb{E}^{\gamma^l}[Q(u)] - \mathbb{E}^{\gamma^{l-1}}[Q(u)] \\
&= \frac{1}{Z^l}\int_U \exp(-\Phi^l(u;\delta))Q(u)I^l(u)d\gamma(u) - \frac{1}{Z^{l-1}}\int_U \exp(-\Phi^{l-1}(u;\delta))Q(u)I^l(u)d\gamma(u) \\
&\qquad + \frac{1}{Z^l}\int_U \exp(-\Phi^l(u;\delta))Q(u)(1-I^l(u))d\gamma(u) \\
&\qquad - \frac{1}{Z^{l-1}}\int_U \exp(-\Phi^{l-1}(u;\delta))Q(u)(1-I^l(u))d\gamma(u) \\
&= \frac{1}{Z^l}\int_U \left(\exp(-\Phi^l(u;\delta)) - \exp(-\Phi^{l-1}(u;\delta))\right) Q(u)I^l(u)d\gamma(u) \\
&\qquad + \left(\frac{1}{Z^l} - \frac{1}{Z^{l-1}}\right)\int_U \exp(-\Phi^{l-1}(u;\delta))Q(u)I^l(u)d\gamma(u) \\
&- \frac{1}{Z^{l-1}}\int_U \left(\exp(-\Phi^{l-1}(u;\delta)) - \exp(-\Phi^l(u;\delta))\right) Q(u)(1-I^l(u))d\gamma(u) \\
&\qquad + \left(\frac{1}{Z^l} - \frac{1}{Z^{l-1}}\right)\int_U \exp(-\Phi^l(u;\delta))Q(u)(1-I^l(u))d\gamma(u).
\end{aligned}$$

With the notation $A_1^l = (1-\exp(\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta)))Q(u)I^l(u)$, we have

$$\begin{aligned}
&\frac{1}{Z^l}\int_U \left(\exp(-\Phi^l(u;\delta)) - \exp(-\Phi^{l-1}(u;\delta))\right) Q(u)I^l(u)d\gamma(u) \\
&= \mathbb{E}^{\gamma^l}[(1-\exp(\Phi^l(u;\delta) - \Phi^{l-1}(u;\delta))Q(u)I^l(u)] = \mathbb{E}^{\gamma^l}[A_1^l].
\end{aligned}$$

Introducing $A_2^l = (\exp(\Phi^{l-1}(u;\delta) - \Phi^l(u;\delta)) - 1)Q(u)(1-I^l(u))$, we may write

$$\begin{aligned}
&-\frac{1}{Z^{l-1}}\int_U \left(\exp(-\Phi^{l-1}(u;\delta)) - \exp(-\Phi^l(u;\delta))\right) Q(u)(1-I^l(u))d\gamma(u) \\
&= \mathbb{E}^{\gamma^{l-1}}[(\exp(\Phi^{l-1}(u;\delta) - \Phi^l(u;\delta)) - 1)Q(u)(1-I^l(u))] = \mathbb{E}^{\gamma^{l-1}}[A_2^l] .
\end{aligned}$$

We note

$$
\begin{aligned}
&\frac{1}{Z^l}-\frac{1}{Z^{l-1}}\\
&\quad=\frac{1}{Z^lZ^{l-1}}\int_U\big(\exp(-\Phi^{l-1}(u;\delta))-\exp(-\Phi^l(u;\delta))\big)\,(I^l(u)+1-I^l(u))d\gamma(u)\\
&\quad=\frac{1}{Z^lZ^{l-1}}\int_U\exp(-\Phi^l(u;\delta))(\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1)I^l(u)d\gamma(u)\\
&\quad+\frac{1}{Z^lZ^{l-1}}\int_U\exp(-\Phi^{l-1}(u;\delta))(1-\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta)))(1-I^l(u))d\gamma(u)\\
&\quad=\frac{1}{Z^{l-1}}\mathbb{E}^{\gamma^l}[(\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1)I^l(u)]+\\
&\quad+\frac{1}{Z^l}\mathbb{E}^{\gamma^{l-1}}[(1-\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta)))(1-I^l(u))].
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\left(\frac{1}{Z^l}-\frac{1}{Z^{l-1}}\right)\int_U\exp(-\Phi^{l-1}(u;\delta))Q(u)I^l(u)d\gamma(u)\\
&\quad=\mathbb{E}^{\gamma^l}[(\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1)I^l(u)]\cdot\\
&\qquad\frac{1}{Z^{l-1}}\int_U\exp(-\Phi^{l-1}(u;\delta))Q(u)I^l(u)d\gamma(u)+\\
&\qquad\mathbb{E}^{\gamma^{l-1}}[(1-\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta)))(1-I^l(u))]\cdot\\
&\qquad\quad\frac{1}{Z^l}\int_U\exp(-\Phi^l(u;\delta))\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))Q(u)I^l(u)d\gamma(u)\\
&=\mathbb{E}^{\gamma^l}[A_3^l]\mathbb{E}^{\gamma^{l-1}}[A_4^l]+\mathbb{E}^{\gamma^{l-1}}[A_5^l]\mathbb{E}^{\gamma^l}[A_6^l]
\end{aligned}
$$

where we define

$$
\begin{aligned}
A_3^l &= (\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1)I^l(u),\\
A_4^l &= Q(u)I^l(u),\\
A_5^l &= (1-\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta)))(1-I^l(u)),\\
A_6^l &= \exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))Q(u)I^l(u).
\end{aligned}
$$

Similarly, defining for $l\geq 1$

$$
A_7^l=Q(u)(1-I^l(u))\ \text{and}\ \ A_8^l=\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta))Q(u)(1-I^l(u))\ ,
$$

there holds

$$
\begin{aligned}
&\left(\frac{1}{Z^l}-\frac{1}{Z^{l-1}}\right)\int_U\exp(-\Phi^l(u;\delta))Q(u)(1-I^l(u))d\gamma(u)\\
&\quad=\mathbb{E}^{\gamma^{l-1}}[(1-\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta)))(1-I^l(u))]\cdot\\
&\qquad\frac{1}{Z^l}\int_U\exp(-\Phi^l(u;\delta))Q(u)(1-I^l(u))d\gamma(u)+\\
&\qquad\mathbb{E}^{\gamma^l}[(\exp(\Phi^l(u;\delta)-\Phi^{l-1}(u;\delta))-1)I^l(u)]\cdot\\
&\qquad\quad\frac{1}{Z^{l-1}}\int_U\exp(-\Phi^{l-1}(u;\delta))\exp(\Phi^{l-1}(u;\delta)-\Phi^l(u;\delta))Q(u)(1-I^l(u))d\gamma(u)\\
&=\mathbb{E}^{\gamma^{l-1}}[A_5^l]\mathbb{E}^{\gamma^l}[A_7^l]+\mathbb{E}^{\gamma^l}[A_3^l]\mathbb{E}^{\gamma^{l-1}}[A_8^l]\ .
\end{aligned}
$$

We conclude that, for every $l \geq 1$, there holds

$$
\begin{aligned}
&\mathbb{E}^{\gamma^l}[Q(u)] - \mathbb{E}^{\gamma^{l-1}}[Q(u)] \\
&= \mathbb{E}^{\gamma^l}[A_1^l] + \mathbb{E}^{\gamma^{l-1}}[A_2^l] + \mathbb{E}^{\gamma^l}[A_3^l] \cdot \mathbb{E}^{\gamma^{l-1}}[A_4^l + A_8^l] + \mathbb{E}^{\gamma^{l-1}}[A_5^l] \cdot \mathbb{E}^{\gamma^l}[A_6^l + A_7^l].
\end{aligned}
$$

In (3.12), when $Q = \ell(P^{l'} - P^{l'-1})$, we denote A_1^l as $A_1^{ll'}$, A_2^l as $A_2^{ll'}$, A_4^l as $A_4^{ll'}$, A_6^l as $A_6^{ll'}$, A_7^l as $A_7^{ll'}$ and A_8^l as $A_8^{ll'}$. For $Q = \ell(P^0)$, we denote A_1^l as A_1^{l0}, A_2^l as A_2^{l0}, A_4^l as A_4^{l0}, A_6^l as A_6^{l0}, A_7^l as A_7^{l0} and A_8^l as A_8^{l0}. We therefore approximate $\mathbb{E}^{\gamma^L}[\ell(P(u))]$ as

$$
\begin{aligned}
&\sum_{l=1}^{L} \sum_{l'=1}^{L'(l)} \mathbb{E}^{\gamma^l}[A_1^{ll'}] + \mathbb{E}^{\gamma^{l-1}}[A_2^{ll'}] + \mathbb{E}^{\gamma^l}[A_3^l] \cdot \mathbb{E}^{\gamma^{l-1}}[A_4^{ll'} + A_8^{ll'}] + \mathbb{E}^{\gamma^{l-1}}[A_5^l] \cdot \mathbb{E}^{\gamma^l}[A_6^{ll'} + A_7^{ll'}] \\
&\quad + \sum_{l=1}^{L} \mathbb{E}^{\gamma^l}[A_1^{l0}] + \mathbb{E}^{\gamma^{l-1}}[A_2^{l0}] + \mathbb{E}^{\gamma^l}[A_3^l] \cdot \mathbb{E}^{\gamma^{l-1}}[A_4^{l0} + A_8^{l0}] + \mathbb{E}^{\gamma^{l-1}}[A_5^l] \cdot \mathbb{E}^{\gamma^l}[A_6^{l0} + A_7^{l0}] \\
&\quad + \sum_{l'=1}^{L'(0)} \mathbb{E}^{\gamma^0}[\ell(P^{l'} - P^{l'-1})] + \mathbb{E}^{\gamma^0}[\ell(P^0)].
\end{aligned}
\tag{4.15}
$$

As usual, the Multilevel Markov Chain Monte Carlo estimator is defined by replacing the mathematical expectations in the preceding expression by finite sample averages. In this way we arrive at the computable MLMCMC-FEM estimator

$$
\begin{aligned}
&E_L^{MLMCMC}(\ell(P)) \\
&:= \sum_{l=1}^{L} \sum_{l'=1}^{L'(l)} E_{M_{ll'}}^{\gamma^l}[A_1^{ll'}] + E_{M_{ll'}}^{\gamma^{l-1}}[A_2^{ll'}] + E_{M_{ll'}}^{\gamma^l}[A_3^l] \cdot E_{M_{ll'}}^{\gamma^{l-1}}[A_4^{ll'} + A_8^{ll'}] + E_{M_{ll'}}^{\gamma^{l-1}}[A_5^l] \cdot E_{M_{ll'}}^{\gamma^l}[A_6^{ll'} + A_7^{ll'}] \\
&+ \sum_{l=1}^{L} E_{M_{l0}}^{\gamma^l}[A_1^{l0}] + E_{M_{l0}}^{\gamma^{l-1}}[A_2^{l0}] + E_{M_{l0}}^{\gamma^l}[A_3^l] \cdot E_{M_{l0}}^{\gamma^{l-1}}[A_4^{l0} + A_8^{l0}] + E_{M_{l0}}^{\gamma^{l-1}}[A_5^l] \cdot E_{M_{l0}}^{\gamma^l}[A_6^{l0} + A_7^{l0}] \\
&+ \sum_{l'=1}^{L'(0)} E_{M_{0l'}}^{\gamma^0}[\ell(P^{l'} - P^{l'-1})] + E_{M_{00}}^{\gamma^0}[\ell(P^0)].
\end{aligned}
$$

To obtain convergence rate bounds, for each discretization level $l \in \mathbb{N}_0$, we introduce the Markov chain $\mathcal{C}_l = \{u^{(k)}\}_{k \in \mathbb{N}_0} \subset \mathbb{R}^{J_l}$ which is seeded with $u^{(0)} \in \mathbb{R}^{J_l}$ and subsequently generated by the MCMC sampler with the acceptance probability $\alpha^{J,l}$ in (4.13) with the parameter choice (3.8).

From (4.4), there are positive constants c_1 and c_2 such that for every $J, l \in \mathbb{N}$ holds

$$
\forall u \in \Gamma_{\boldsymbol{b}} : \quad \Phi^{J,l}(u) \leq c_1 + c_2 \exp\left(2 \sum_{j=1}^{\infty} b_j |u_j|\right).
$$

We define

$$
\kappa = \int_U \exp\left(-c_1 - c_2 \exp(2 \sum_{j=1}^{\infty} b_j |u_j|)\right) d\gamma_{\boldsymbol{b}}(u).
$$

As shown in [19, Lemma 4.9], κ is strictly positive. Following [19], we define the probability measure $\bar{\gamma}$ by

$$\forall u \in \Gamma_{\boldsymbol{b}}: \quad d\bar{\gamma}(u) = \frac{1}{\kappa}\exp(-c_1 - c_2\exp(2\sum_{j=1}^{\infty} b_j|u_j|))d\gamma_{\boldsymbol{b}}(u)\ . \tag{4.16}$$

Then, there exists a constant $c > 0$ (independent of l) such that

$$\sup_{u\in\Gamma_{\boldsymbol{b}}}\sup_{J,l\in\mathbb{N}} \frac{d\bar{\gamma}}{d\gamma^{J,l,\delta}}(u) \le \frac{1}{\kappa} < c < \infty\ . \tag{4.17}$$

We denote by $\mathcal{P}^{\bar{\gamma},l}$ the probability measure of the probability space that describes the randomness of this Markov chain when the initial state $u^{(0)}$ is distributed arcording to $\bar{\gamma}$. Then, for each discretization level $l = 0, 1, 2, ...,$ the chains $\mathcal{C}_l$ are pairwise independent. For every fixed discretization level L, we denote by $\mathbf{C}_L = \{\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_L\}$ the collection of Markov chains obtained from running the MCMC sampling procedure with the discretizations at level $l = 0, 1, 2, ..., L$. We denote further by $\mathbf{P}_L$ the product probability measure on the probability space generated by the collection of these L independent Markov chains. For each fixed discretization level L, the measure $\mathbf{P}_L$ describes the law of $\mathbf{C}_L = \{\mathcal{C}_l\}_{l=0}^L$:

$$\mathbf{P}_L := \mathcal{P}^{\bar{\gamma},0} \otimes \mathcal{P}^{\bar{\gamma},1} \otimes \mathcal{P}^{\bar{\gamma},2} \otimes \dots \otimes \mathcal{P}^{\bar{\gamma},L}\ .$$

Let $\mathbf{E}_L$ denote the expectation over all realizations of the collection $\mathbf{C}_L$ of chains $\{\mathcal{C}_l\}_{l=0}^L$ with respect to the product measure $\mathbf{P}_L$. With the choice of parameters

$$\begin{aligned} &L'(l) = L - l, \quad M_{ll'} = 2^{2(L-(l+l'))} \ \text{ for } l \ge 1,\ l' \ge 1,\\ &M_{0l} = M_{l0} = 2^{2L}/L^2,\ M_{00} = 2^{2L}/L^4, \end{aligned} \tag{4.18}$$

we have the following result.

Theorem 4.7. *Assume that the domain D is convex and $f \in L^2(D)$. Under Assumptions 4.1, and A.2 for $d = 2, 3$, with the choices* (4.18) *there exists a constant $C(\delta) > 0$ such that for every $L \in \mathbb{N}$ holds*

$$\mathbf{E}_L[|\mathbb{E}^{\gamma^\delta}[P] - E_L^{MLMCMC}[P]|] \le C(\delta)L^2 2^{-L}\ . \tag{4.19}$$

The total number of degrees of freedom used in running the MLMCMC sampler, is bounded by $O(L2^{2L})$ for $d = 2$ and $O(2^{3L})$ for $d = 3$. Further, with the availability of a Riesz finite element basis as in Assumption 3.5, the expectation of the total number of floating point operations in the probability space of all the proposals is bounded by $O(L^{d-1}2^{(d+1/q)L})$. Denoting the expectation of the total number of degrees of freedom which enter in running the Markov chain on all discretization level by N, the error in (4.19) *is bounded by $O((\log N)^{3/2}N^{-1/2})$ for $d = 2$ and by $O((\log N)^2 N^{-1/3})$ for $d = 3$. The expectation of the total number of floating point operations is bounded by $O((\log N)^{-1/(2q)}N^{1+1/(2q)})$ for $d = 2$ and by $O((\log N)^2 N^{1+1/(3q)})$ for $d = 3$.*

We prove this theorem in Appendix A.

We repeat that the assumption on availability of a Riesz basis in $H_0^1(D)$ can be weakened, while obtaining the same error vs. work bounds, by resorting to a probabilistic convergence analysis of multilevel iterative solvers. We refer to [18] for details.

As for the uniform prior, we can reduce the multiplying logarithmic factor L^2 by increasing $M_{ll'}$ as in Table 1.

Remark 4.8. *The MLMCMC method is developed to approximate the expectations in* (4.15). *If we use other sampling methods such as HMC or SMC to approximate these expectations, we can develop corresponding multilevel methods for these sampling procedures.*

Remark 4.9. *We prove that asymptotically, when the smallest mesh width at discretization level L, i.e.* $O(2^{-L})$, *approaches zero, the MLMCMC-FEM achieves the essentially optimal convergence rate for a fixed number of FE degrees of freedom and floating point operations. However, depending on the scaling of the observation data* δ *in Assumption 2.2, the constant* $C(\delta)$ *in the error estimate* (4.19) *can be large which leads to the requirement of very high spatial resolution to obtain a numerical approximation of the forward problem within a prescribed accuracy. The constant* $C(\delta)$ *essentially depends on the constant in* C *in* (A.4) *in Assumption A.2 on geometric ergodicity. Inspecting the proof of Lemma B.2, from* (B.4), *for the independence sampler, this constant can be bounded by* $a(2a^2+4a-4)\mathbb{E}^\gamma[\mathcal{V}^{ll'}]$ *where* $1/a$ *is the uniform lower bound for the normalizing constant* $Z^{J,l}$ *that was established in Lemma B.1 and* $\mathcal{V}^{ll'}$ *is defined in* (A.3). *Proposition B.4 shows that* $\mathbb{E}^\gamma[\mathcal{V}^{ll'}]$ *depends on* $\sum_{j=1}^\infty(b_j^2+\bar{b}_j^2+b_j+\bar{b}_j)$ *where* b_j *and* $\bar{b}_j$ *are defined in Assumptions 2.6 and 4.1, respectively. Therefore, in the case that* $\|\boldsymbol{b}\|_{\ell^1(\mathbb{N})}$ *and* $\|\bar{\boldsymbol{b}}\|_{\ell^1(\mathbb{N})}$ *are large, we expect the constant* $C(\delta)$ *in* (4.19) *to be large. Further from the proof of Lemma B.1, the uniform lower bound* $1/a$ *of the normalizing constant* $Z^{J,l}$ *is small when the observation noise covariance* Σ *is small, i.e. when* $\|\Sigma^{-1}\|_{\mathbb{R}^{k\times k}}$ *is large. This is because the constant* c^* *in the proof of Lemma B.1 is large, which implies that the constant* C *in* (B.1) *must ensure* $c^*/C<1$. *We thus conclude that when both,* $\|\boldsymbol{b}\|_{\ell^1(\mathbb{N})}$ *and* $\|\bar{\boldsymbol{b}}\|_{\ell^1(\mathbb{N})}$, *are large and at the same time the observation noise covariance* Σ *is small, the constant* $C(\delta)$ *in* (4.19) *is large and the FEM mesh width* 2^{-L} *of spatial resolution has to be sufficiently small to achieve a prescribed level of accuracy for the MLMCMC-FEM. We note that this is also the case for the plain (single-level) MCMC-FEM where the FE approximation of the forward equation is solved with the same (small) meshsize for all the samples; it is well-known that here the computational complexity is far larger than the complexity of the MLMCMC-FEM developed above.*

5. Numerical experiments

First we consider an example where the MLMCMC method developed for the uniform prior case in [21] presented in Section 3.3 fails to approximate the posterior expectation of problems with the Gaussian prior. We consider the one dimensional Dirichlet problem in the domain $D=(0,1)$ where

$$-\frac{d}{dx}(K(x,u)\frac{dP}{dx})=f,\quad x\in(0,1)$$

where $P(0)=P(1)=0$. The coefficient

$$K(x,u)=\exp(u\sin(4\pi x))$$

where $u\sim N(0,1)$; and $f(x)=200$. The observation is

$$\mathcal{G}(u)=\int_0^1 x\frac{dP}{dx}dx;$$

and the quantity of interest is

$$\ell(P(u)) = \int_0^1 x^{1.5}\frac{dP}{dx}(x,u)dx.$$

The data is generated by solving the forward PDE exactly for one randomly chosen realization of u and by generating the noise which follows the standard normal distribution by MATLAB's random number generator. We chose one draw being $\delta = -16.5384$. For this data, the reference posterior expectation is evaluated by solving the equation exactly and by using many Gauss-Hermite quadrature points. The tables below show the arithmetic average of the errors for 64 runs of the MLMCMC using the independence sampler. In Table 2, we present the error for the MLMCMC sampler developed for uniform prior measure in [21]. While in this example a few of the 64 runs produce reasonable approximations for the posterior expectation, the table shows that in general the method does not converge. In Table 3, we present the average error of 64 runs of the MLMCMC sampler developed in Section 4. The results clearly shows that the MLMCMC method for Gaussian prior converges as proved theoretically. Indeed, the slope of the best fit straightline is 0.95 which is in reasonable agreement with our theory.

Mesh-Level (L)	Average MLMCMC error
8	2.77959E+23
9	1.96933E+46
10	1683671.3
11	2.8192E+19
12	3.29498E+32
13	2.89131E+41

Table 2. MLMCMC error from using the MLMCMC sampler developed in [21] for uniform prior, as recapitulated in Section 3, for the gaussian prior.

Mesh-Level (L)	Average MLMCMC error
8	1.72670013
9	1.05627325
10	0.5178982
11	0.4255921
12	0.11905266
13	0.06412478

Table 3. MLMCMC error from using the presently proposed MLMCMC sampler as developed for gaussian prior in Section 4.

Now we consider linear, elliptic PDEs in the domain $D = (0,1)\times(0,1)$ with periodic boundary condition, and with log-gaussian diffusion coefficient under Gaussian prior probability measure. The theory developed above holds for periodic boundary condition. The advantage of considering the periodic boundary condition is that the forward equation can be solved with high numerical accuracy by the Fourier collocation method. For $u \in \mathbb{R}$, we consider the parametric forward equation

$$-\nabla \cdot (K(x,u)\nabla P(x,u))) = f(x) \quad \text{for} \quad x \in D,$$

with the periodic boundary condition, where

$$K(x,u) = e^{u(\sin(2\pi x_1)+\sin(2\pi x_2))}, \quad \text{and } f(x) = 200\,(\sin(2\pi x_1) + \sin(2\pi x_2))$$

for $x = (x_1, x_2) \in D$; $u \sim N(0,1)$. Imposing the condition $\int_D P(x,u)dx = 0$, this forward problem has a unique solution. The forward observation functional is

$$\mathcal{G}(u) = \int_D x^\top \nabla_x P(x,u)dx$$

and the quantity of interest is given by

$$\ell(P(u)) = \int_D x_1^{1.5} \frac{\partial P}{\partial x_1}(x,u) + x_2^{1.5} \frac{\partial P}{\partial x_2}(x,u)dx.$$

The data is generated by choosing randomly a realization of u by Matlab random generator. A numerical value of the centered gaussian observation noise ϑ is generated randomly by Matlab random number generator. Here the noisy observation $\delta = -5.8315$ (which was randomly drawn under the prior) was used. To compute a reference posterior expectation, we evaluate

$$\mathbb{E}^{\rho^\delta}[\ell(P)] = \int_{-\infty}^{\infty} \ell(P(u))\; d\rho^\delta(u) = \int_{-\infty}^{\infty} \ell(P(u)) \exp\left(-\frac{1}{2}|\delta - G(u)|^2\right)\; d\rho(u)$$

using 1200 Gauss-Hermite quadrature points. At each quadrature point, the forward equation is solved by a Fourier collocation method with 1024 collocation points.

First we present the numerical experiments with the independence sampler. In the figures below, we plot the arithmetic average of the absolute errors of 64 runs of the MLMCMC-FEM approximation versus the finest resolution meshwidth 2^{-L}.

In Figure 1, we plot the MCMC-FEM error versus the meshsize 2^{-L} for the case where α in Table 1 equals 0. The gradient of the best fit straight line is 0.9312. In Figure 2, we plot the MLMCMC-FEM error versus the meshwidth 2^{-L} for $\alpha = 2$. The gradient of the best fit straight line is 0.93257. Similarly, for $\alpha = 3$ and $\alpha = 4$, the MLMCMC-FEM error versus 2^{-L} is plotted in Figures 3 and 4. The gradient of the best fit straight line is 1.0072 and 1.0804, respectively.

We now present the results for the MLMCMC-FEM with the pCN sampler

$$v^{(k)} = \sqrt{1-\beta^2}u^{(k)} + \beta\xi,$$

where $\xi \sim N(0,1)$ for different values of β. First for $\beta = 1/\sqrt{2}$, we plot the MLMCMC-FEM error versus the meshwidth parameter 2^{-L} for $\alpha = 0, 2, 3$ and 4 in Figures 5, 6, 7 and 8, respectively. The slope of the least-squares fit straight lines are 0.73977, 1.0392, 1.0111 and 0.99257 respectively. We see that except when $\alpha = 0$ (in which case the method slightly underperforms possibly due to the large L^2 multiplying factor in the error bound), the MLMCMC-FEM using the pCN sampler performs as expected from our theoretical results. For $\beta = 1/\sqrt{10}$, we plot the results for $\alpha = 0, 2, 3$ and 4 in Figures 9, 10, 11 and 12, respectively. The slopes of the best fit straight lines in the error plots are 0.44129, 0.97061, 0.9058 and 0.92827, respectively. Again when $\alpha = 0$ the convergence rate is inferior to the optimal rate $O(2^{-L})$ due to the large multiplying factor L^2 in the error, but for other values of α, the observed convergence rate essentially corresponds to the rates in our theorems.

To test the CPU time performance of the method, we record below in tables 4 and 5 the average CPU time for 5 different runs of the MLMCMC-FEM for different values of L (corresponding to the meshwidth $O(2^{-L})$ at the finest discretization) for the independence sampler. The CPU time behaves like $O(2^{2L})$ which is essentially optimal. We obtained similar results for the pCN sampler.

Next we consider the problem where $\log K$ is a stationary gaussian random field. We sample the values of the log-gaussian coefficients at all the FE nodes in D by circulant embedding [12]. We consider

$$-\nabla\cdot(K(x)\nabla P(x)) = \cos(2\pi x_1)\sin(2\pi x_2), \quad x=(x_1,x_2)\in D=(0,1)\times(0,1),$$
$$P(0,x_2)=0, P(1,x_2)=1, \frac{\partial P}{\partial x_2}(x_1,0)=0, \frac{\partial P}{\partial x_2}(x_1,1)=0. \tag{5.1}$$

To denote the dependence of the solution on the coefficient K, we denote it also as $P(K)$. We assume that

$$K=\exp(R)$$

where R is a GRF with mean $\mu=0$ and with covariance function given by

$$C(x,y)=\rho(|x-y|)=\exp\left(-|x-y|^2\right), \quad x,y\in D\,.$$

Here $|\cdot|$ denotes the Euclidean norm. The observation functional is given by

$$\mathcal{G}(K)=\int_D (0.5-x_1)^2\frac{\partial P}{\partial x_1}(K)+(0.5-x_2)^2\frac{\partial P}{\partial x_2}(K)dx$$

and the quantity of interest is

$$\ell(P(K))=\int_D P(x)dx.$$

The additive observation noise $\vartheta\sim N(0,1)$. Following Graham et al. [16], we consider the map $\eta: D\to D$ given by $\eta(x)=(1-x_1,1-x_2)$. Let $K_\eta(x)=K(\eta(x))$ and $P_\eta(x)=1-P(\eta(x)))$. Then P_η is the solution of problem (5.1) with coefficient $K=K_\eta$. Further

$$\mathcal{G}(K)=\mathcal{G}(K_\eta).$$

As K and K_η define GRFs with the same mean and homogeneous covariance, they have the same probability law, which is the prior γ on the space U of continuous functions. Its law is invariant under the map $K\to K_\eta$. We have

$$Z=\int_U \exp\left(-\frac{1}{2}|\delta-\mathcal{G}(K)|^2\right)d\gamma(K)=\int_U\exp\left(-\frac{1}{2}|\delta-\mathcal{G}(K_\eta)|^2\right)d\gamma(K_\eta).$$

The posterior expectation is

$$\mathbb{E}^{\gamma^\delta}[\ell(P(K))]=\frac{1}{Z}\int_U \ell(P(K))\exp\left(-\frac{1}{2}|\delta-\mathcal{G}(K)|^2\right)d\gamma(K).$$

As

$$\int_D P(x)dx=\int_D P(\eta(x))dx,$$

and $\mathcal{G}(K)=\mathcal{G}(K_\eta)$, we have that

$$\mathbb{E}^{\gamma^\delta}[\ell(P(K))]=\frac{1}{Z}\int_U(1-\ell(P(K_\eta))\exp\left(-\frac{1}{2}|\delta-\mathcal{G}(K_\eta)|^2\right)d\gamma(K).$$

For any δ, γ^δ is invariant under the mapping that maps K to K_η. Therefore

$$\mathbb{E}^{\gamma^\delta}[\ell(P(K))] = 1 - \mathbb{E}^{\gamma^\delta}[\ell(P(K))]$$

so

$$\mathbb{E}^{\gamma^\delta}[\ell(P(K))] = 0.5.$$

In Figure 13, we plot the error of $\mathbb{E}^{\gamma^\delta}[\ell(P(K))]$ of the posterior expectation approximated by the MLMCMC-FEM where the independence sampler is used. The stationary GRF is numerically sampled by circulant embedding (see, [12], [16], and Appendix C) to generate the samples for the nodes of the finite element mesh in D. Here we choose $\alpha = 3$. The slope of the best fit straight line in the error convergence plot is 0.910. The MLMCMC-FEM error plotted is numerically estimated as arithmetic average of the errors of 64 independent runs of MLMCMC-FEM. In Figure 14, we plot the MLMCMC-FEM error for the pCN sampling method where β is chosen as $1/\sqrt{2}$ and α is also chosen as 3. The slope of the best fit straight line is 0.912. Again, the error plotted is the arithmetic average of the errors of 64 independent runs of the MLMCMC.

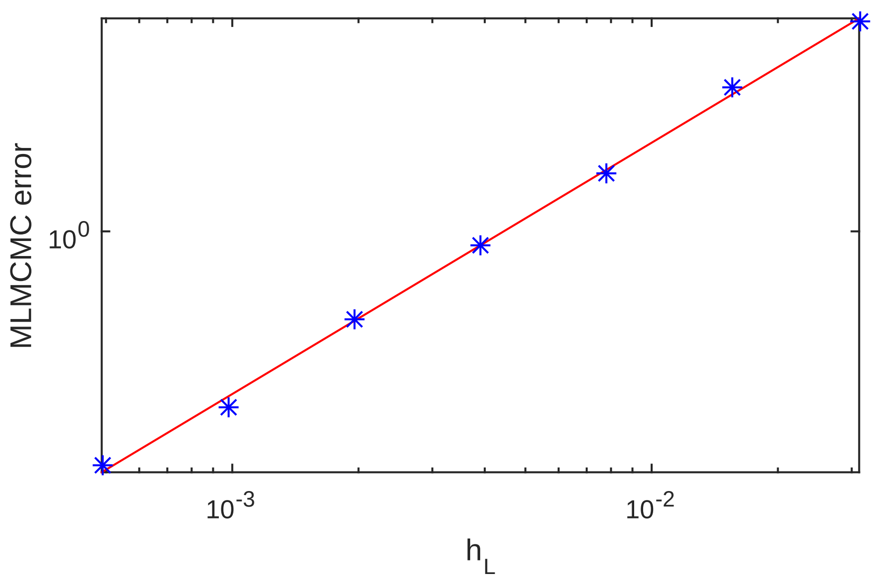

FIGURE 1. Independence sampler, $\alpha = 0$

6. CONCLUSIONS

We proposed and analyzed a multi-level MCMC algorithm for numerical approximation of Bayesian inverse problems for linear, scalar elliptic PDEs with log-gaussian diffusion coefficient. Our convergence rate results and our results on ε-complexity are established under Assumption 2.6. In particular, and distinct

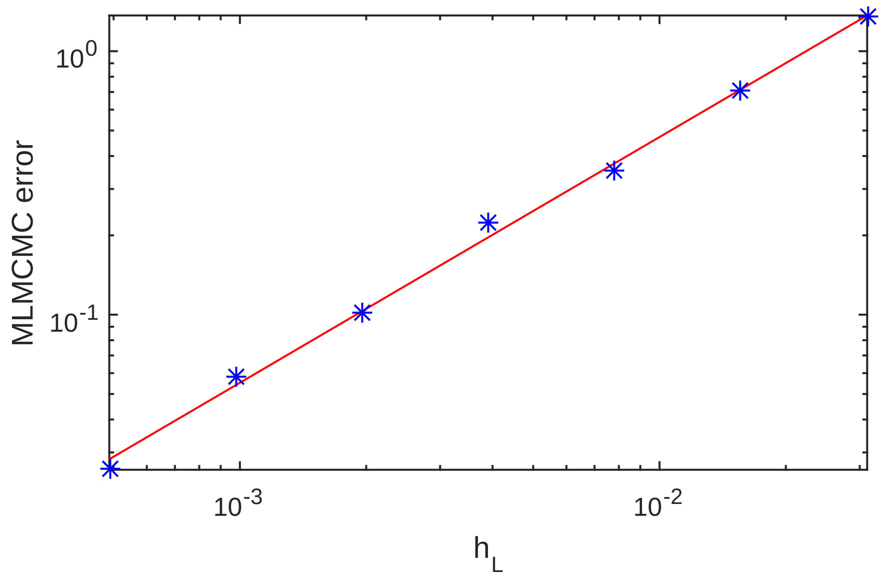

FIGURE 2. Independence sampler, $\alpha = 2$

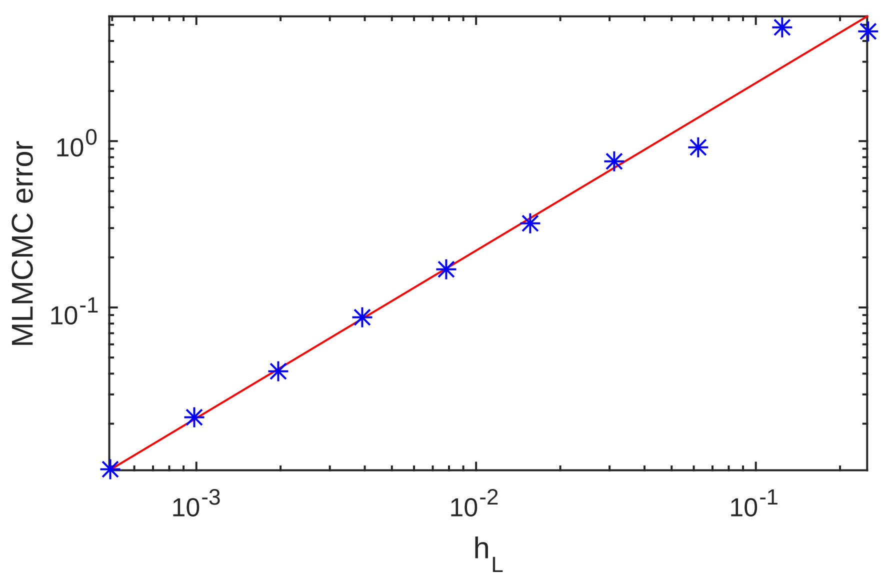

FIGURE 3. Independence sampler, $\alpha = 3$

from other, recent work, such as [28] (Assumption A4, and Appendix), [4], [3], our proofs did not require the parametric diffusion coefficient $K(\cdot, u)$ in (2.12) to be bounded away from zero. This lack of uniform lower bound required, in turn, essential modifications in the MLMCMC sampling algorithm, which led to a novel

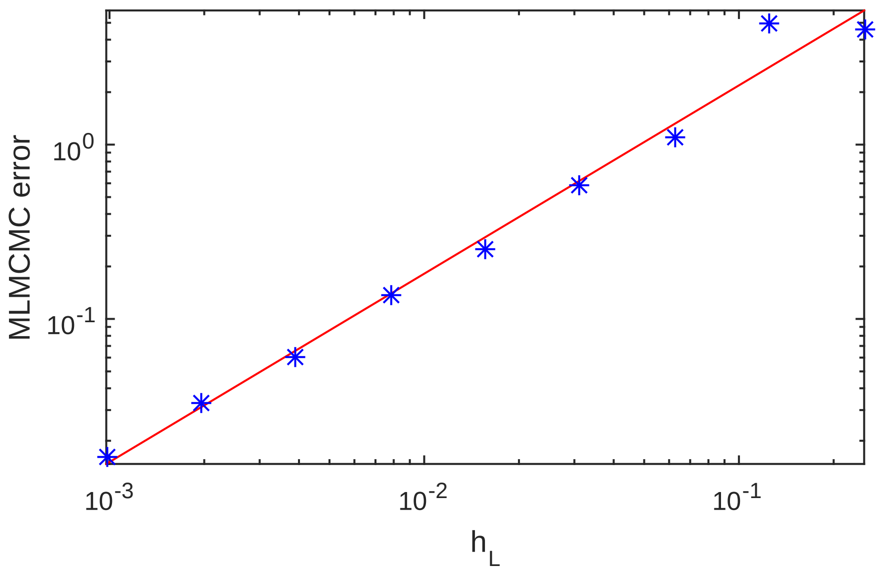

FIGURE 4. Independence sampler, $\alpha = 4$

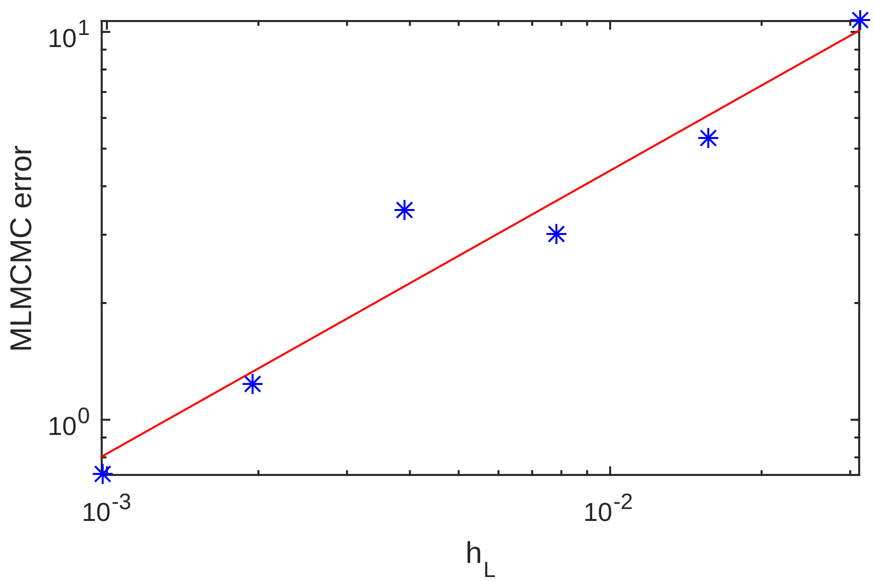

FIGURE 5. pCN sampler, $\beta = \frac{1}{\sqrt{2}}$, $\alpha = 0$

MLMCMC computational strategy, under gaussian prior. Numerical experiments provided strong indication that the presently proposed, novel modifications of the MLMCMC-FEM sampling strategy are indeed necessary to ensure convergence of

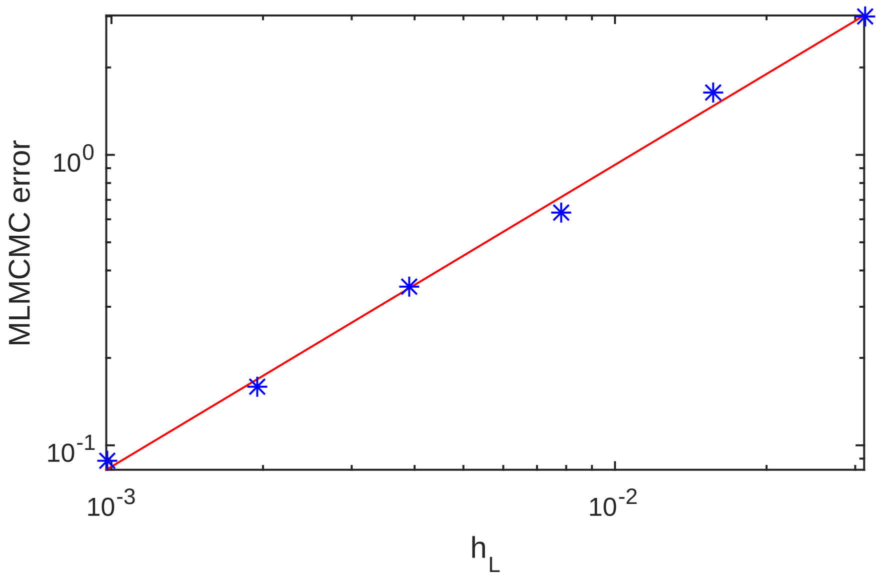

FIGURE 6. pCN sampler, $\beta = \frac{1}{\sqrt{2}}$, $\alpha = 2$

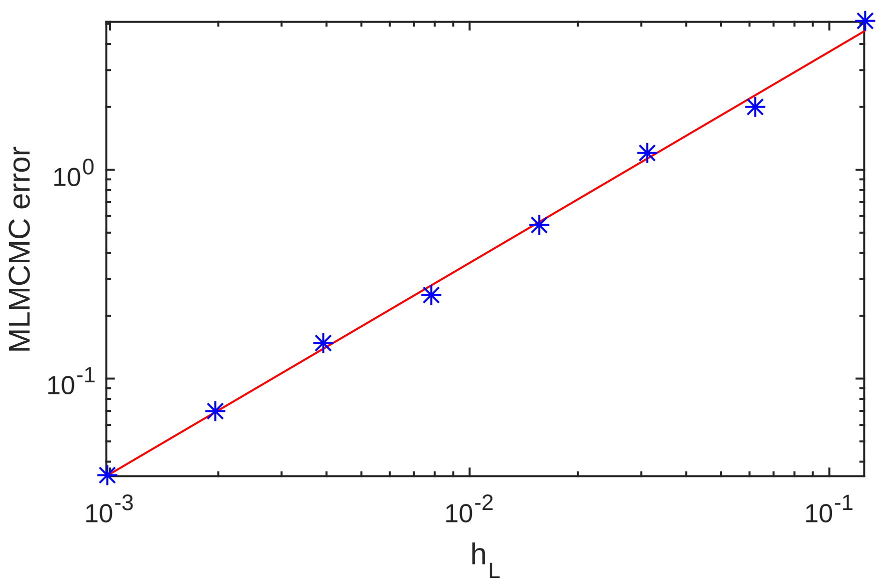

FIGURE 7. pCN sampler, $\beta = \frac{1}{\sqrt{2}}$, $\alpha = 3$

the MLMCMC-FEM process under gaussian prior. The proposed MLMCMC-FEM method achieved essentially optimal error vs. work relation. The novel truncation argument (cf. Eqn. (4.14)) controlling level differences in the MLMCMC sampler

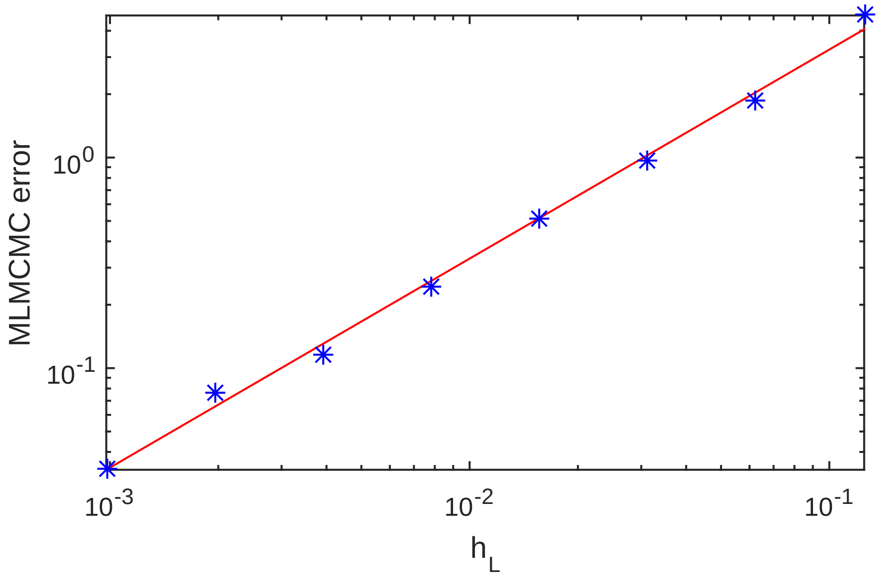

FIGURE 8. pCN sampler, $\beta = \frac{1}{\sqrt{2}}$, $\alpha = 4$

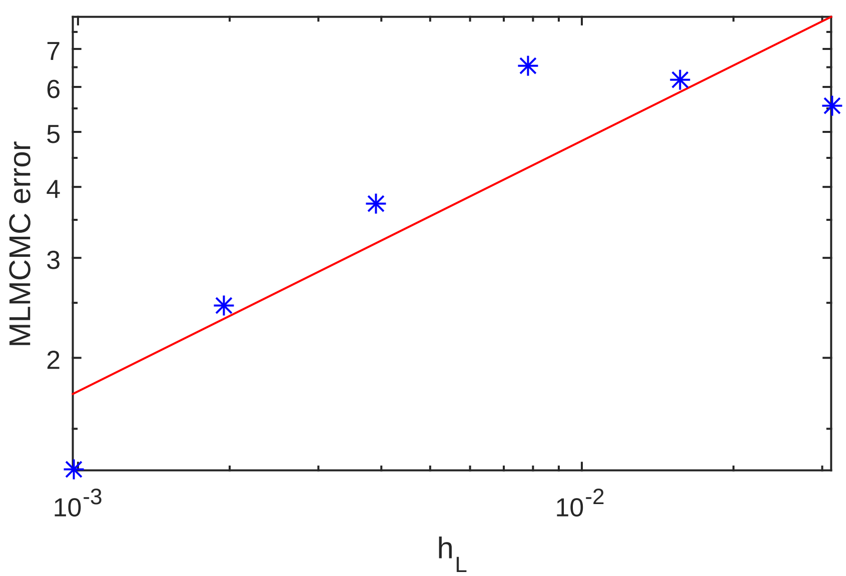

FIGURE 9. pCN sampler, $\beta = \frac{1}{\sqrt{10}}$, $\alpha = 0$

will also allow mathematical convergence analyses of multilevel versions of the HMC and SMC algorithms for Bayesian PDE inversion under gaussian prior, under the weak Assumption 2.6, cp. Remark 4.8.

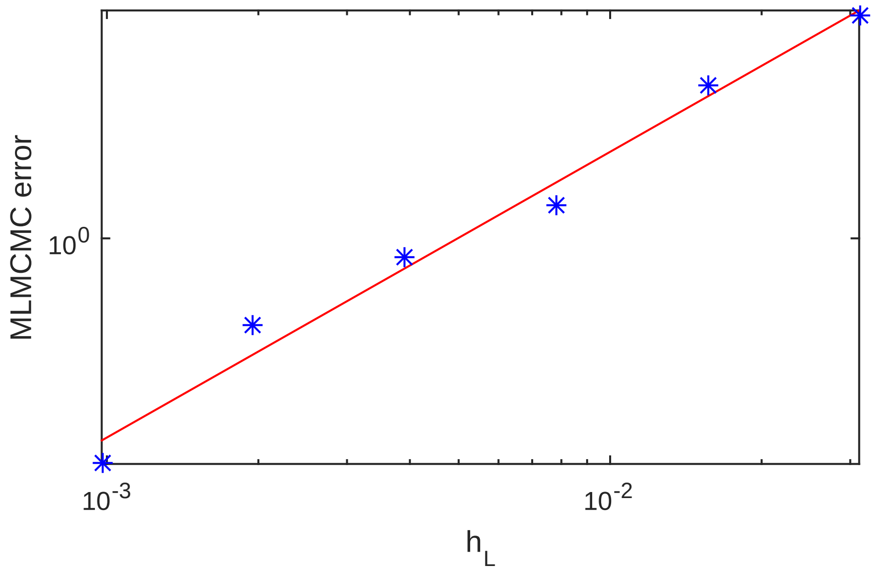

FIGURE 10. pCN sampler, $\beta = \frac{1}{\sqrt{10}}$, $\alpha = 2$

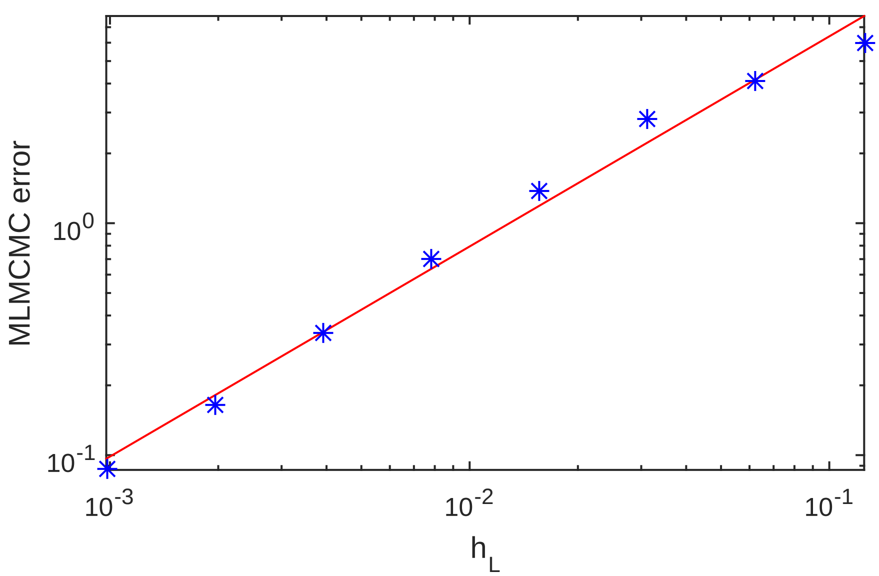

FIGURE 11. pCN, $\beta = \frac{1}{\sqrt{10}}$, $\alpha = 3$

Acknowledgement The research of VHH is supported by the Singapore MOE AcRF Tier 1 grant RG30/16, the MOE Tier 2 grant MOE2017-T2-2-144; QJH is

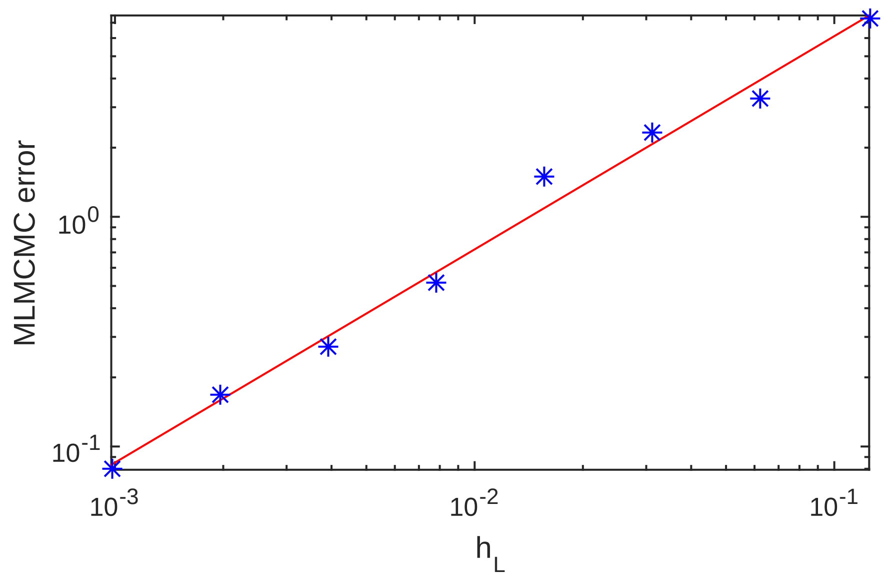

FIGURE 12. pCN, $\beta = \frac{1}{\sqrt{10}}$, $\alpha = 4$

Mesh-Level (L)	Mean CPU time
2	0.052
3	0.067
4	0.105
5	0.256
6	0.767
7	2.387
8	8.863
9	35.467
10	146.463

TABLE 4. CPU time for $\alpha = 0$

Mesh-Level (L)	Mean CPU time
2	0.046
3	0.088
4	0.342
5	1.512
6	6.321
7	27.256
8	115.612
9	490.272
10	2097.376

TABLE 5. CPU time for $\alpha = 2$

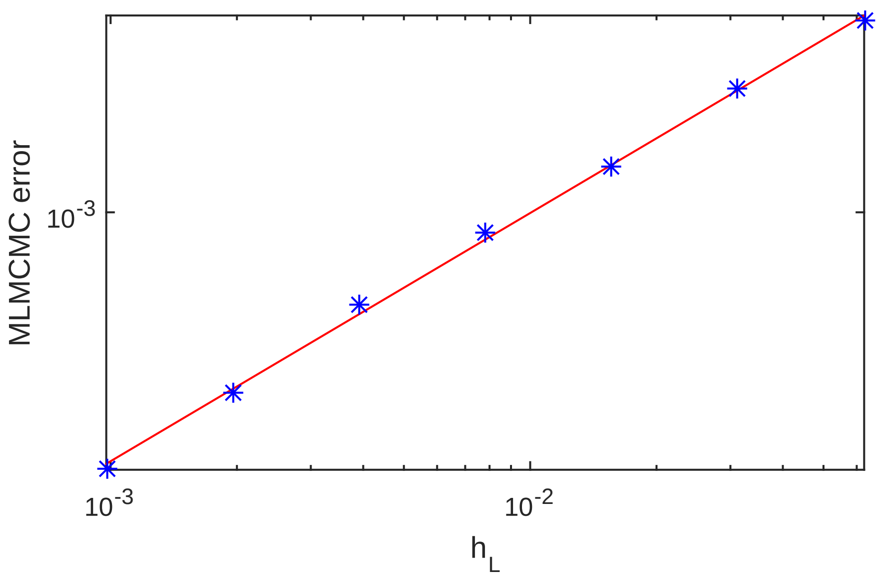

FIGURE 13. MLMCMC error for example (5.1): Independence sampler, $\alpha = 3$

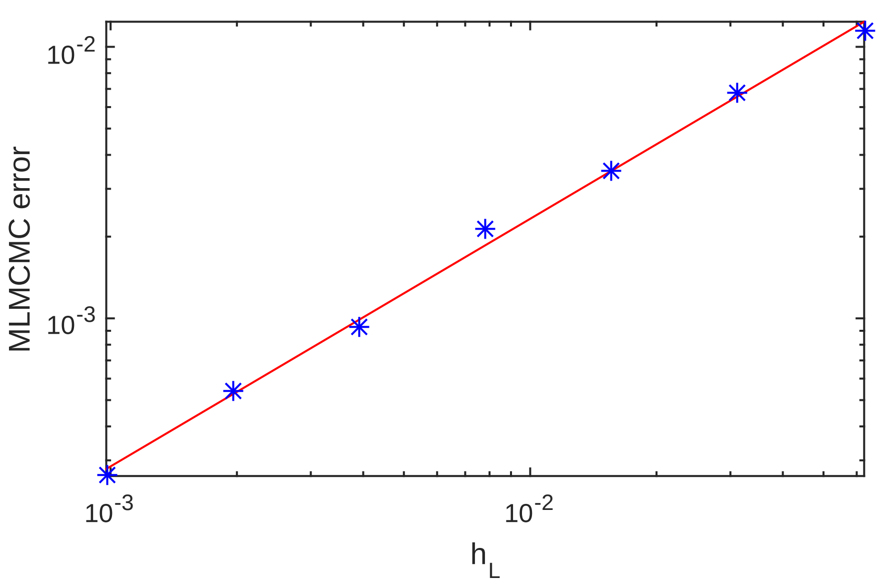

FIGURE 14. MLMCMC error for example (5.1): pCN sampler, $\beta = \frac{1}{\sqrt{2}}$, $\alpha = 3$

supported by a graduate scholarship from Nanyang Technologial University, Singapore; VHH and QJH are supported by a Collaborative Research Award of College of Science, Nanyang Technological University, Singapore; and CS is supported by Swiss National Science Research Foundation. Parts of the research were conducted during the visits of VHH and CS to the Isaac Newton Institute of Mathematical Sciences, Cambridge, UK in Spring 2018 and the E. Schrödinger Institute, Vienna, Austria, in the summer of 2018 and during the visits of VHH to the Institute for Mathematical Research (FIM, ETH Zürich). We thank these institutes for the hospitality and the excellent working condition.

Appendix A.

We prove Theorem 4.7 in this Appendix. To perform the error analysis of the MLMCMC approximation we decompose the error into three terms as follows.

Proposition A.1. *We have*

$$\mathbb{E}^{\gamma^\delta}[\ell(P)] - E_L^{MLMCMC}[\ell(P)] = I_L + II_L + III_L \tag{A.1}$$

where

$$I_L := \mathbb{E}^{\gamma^\delta}[\ell(P)] - \mathbb{E}^{\gamma^L}[\ell(P)],$$
$$II_L = \sum_{l=1}^{L}(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}})[\ell(P) - \ell(P^{L'(l)})] + \mathbb{E}^{\gamma^0}[\ell(P) - \ell(P^{L'(0)})]$$

and

$$\begin{aligned} III_L = &\sum_{l=1}^{L}\sum_{l'=1}^{L'(l)} \mathbb{E}^{\gamma^l}[A_1^{ll'}] + \mathbb{E}^{\gamma^{l-1}}[A_2^{ll'}] + \mathbb{E}^{\gamma^l}[A_3^l]\cdot\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}+A_8^{ll'}] + \mathbb{E}^{\gamma^{l-1}}[A_5^l]\cdot\mathbb{E}^{\gamma^l}[A_6^{ll'}+A_7^{ll'}] \\ &+\sum_{l=1}^{L}\mathbb{E}^{\gamma^l}[A_1^{l0}] + \mathbb{E}^{\gamma^{l-1}}[A_2^{l0}] + \mathbb{E}^{\gamma^l}[A_3^l]\cdot\mathbb{E}^{\gamma^{l-1}}[A_4^{l0}+A_8^{l0}] + \mathbb{E}^{\gamma^{l-1}}[A_5^l]\cdot\mathbb{E}^{\gamma^l}[A_6^{l0}+A_7^{l0}] \\ &+\sum_{l'=1}^{L'(0)}\mathbb{E}^{\gamma^0}[\ell(P^{l'}-P^{l'-1})] + \mathbb{E}^{\gamma^0}[\ell(P^0)] \\ &-E_L^{MLMCMC}[\ell(P)]. \end{aligned}$$

Proof From equation (3.9) we have

$$\mathbb{E}^{\gamma^\delta}[\ell(P)] - \mathbb{E}^{\gamma^L}[\ell(P)] = \mathbb{E}^{\gamma^\delta}[\ell(P)] - \sum_{l=1}^{L}\left(\mathbb{E}^{\gamma^l}[\ell(P)] - \mathbb{E}^{\gamma^{l-1}}[\ell(P)]\right) - \mathbb{E}^{\gamma^0}[\ell(P)]\,. \tag{A.2}$$

It follows that

$$\begin{aligned} &\mathbb{E}^{\gamma^\delta}[\ell(P)] - \mathbb{E}^{\gamma^L}[\ell(P)] = \mathbb{E}^{\gamma^\delta}[\ell(P)] \\ &-\sum_{l=1}^{L}\left(\mathbb{E}^{\gamma^l}[\ell(P^{L'(l)})] - \mathbb{E}^{\gamma^{l-1}}[\ell(P^{L'(l)})]\right) - \mathbb{E}^{\gamma^0}[\ell(P^{L'(0)})] - II_L\,. \end{aligned}$$

Rearranging and using (3.11) gives

$$\begin{aligned}\mathbb{E}^{\gamma^\delta}[\ell(P)] &= I_L + II_L + \sum_{l=1}^{L}\left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[\ell(P^{L'(l)})] + \mathbb{E}^{\gamma^0}[\ell(P^{L'(0)})] \\ &= I_L + II_L + \sum_{l=1}^{L}\sum_{l'=1}^{L'(l)}\left(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}}\right)[\ell(P^{l'}) - \ell(P^{l'-1})] \\ &\quad + \sum_{l=1}^{L}(\mathbb{E}^{\gamma^l} - \mathbb{E}^{\gamma^{l-1}})[\ell(P^0)] + \mathbb{E}^{\gamma^0}[\ell(P^0)] \\ &\quad + \sum_{l'=1}^{L'(0)}\mathbb{E}^{\gamma^0}[\ell(P^{l'}) - \ell(P^{l'-1})]\ .\end{aligned}$$

We then get the conclusion.□

To prove Theorem 4.7, we work under the following assumption of geometric ergodicity. We will discuss sufficient conditions for the validity of this assumption in Appendix B. Let $\mathcal{E}^{\bar{\gamma},l}$ denote the expectation with respect to the probability space generated by the MCMC process with the acceptance probability defined in (2.9), where Φ is replaced by the potential obtained from the FE approximation (4.10) of the forward problem, with the initial sample $u^{(0)}$ distributed according to the probability measure $\bar{\gamma}$ defined in (4.16).

Assumption A.2. *For each l and l' in $\mathbb{N}$, we denote by*

$$\mathcal{V}^{ll'}(u) = \exp\left(11\sum_{j=1}^{\infty}(b_j + \bar{b}_j)|u_j| + \frac{1}{\varepsilon}\sum_{j>J_{l-1}} b_j|u_j| + \frac{1}{\varepsilon'}\sum_{j'>J_{l'-1}} b_{j'}|u_{j'}|\right) \quad \text{(A.3)}$$

where $\varepsilon = \sum_{j>J_{l-1}} b_j$ and $\varepsilon' = \sum_{j>J_{l'-1}} b_j$. Then if $g : \Gamma_{\boldsymbol{b}} \to \mathbb{R}$ is a function such that $|g(u)| \le \mathcal{V}^{ll'}(u)$ for every $u \in \Gamma_{\boldsymbol{b}}$, there exists $C > 0$ independent of l such that for every $M \in \mathbb{N}$ holds

$$\left(\mathcal{E}^{\bar{\gamma},l}\left[\left|\mathbb{E}^{\gamma^l}[g] - E_M^{\gamma^l}[g]\right|^2\right]\right)^{1/2} \le CM^{-1/2}\ . \quad \text{(A.4)}$$

Remark A.3 ("Finite-dimensional noise case")**.** *Assume that the expansion in* (2.12) *has a-priori only a finite number J of random parameters u_j, i.e.*

$$K(\cdot, u) = K_*(\cdot) + \exp\left(\bar{K}(\cdot) + \sum_{j=1}^{J} u_j\psi_j(\cdot)\right)$$

Then we can choose $\mathcal{V}^{ll'}$ as

$$\mathcal{V}^{ll'}(u) = \exp\left(11\sum_{j=1}^{J}(b_j + \bar{b}_j)|u_j|\right). \quad \text{(A.5)}$$

Proof of Theorem 4.7 We derive an error bound by estimating the three terms I_L, II_L and III_L in the error (A.1) separately. Throughout we choose $J_l = 2^{\lceil l/q\rceil}$. For the first term I_L, from [33], we have

$$|(\mathbb{E}^{\gamma^\delta} - \mathbb{E}^{\gamma^L})[\ell(P)]| \le 2\left(\mathbb{E}^{\gamma^\delta}(\ell(P)^2) + \mathbb{E}^{\gamma^L}(\ell(P)^2)\right)^{1/2} d_{\text{Hell}}(\gamma^\delta, \gamma^L).$$

As the normalizing constant in (4.11) is uniformly (with respect to $J, l \in \mathbb{N}$) bounded from below away from zero, the expections $\mathbb{E}^{\gamma^L}(\ell(P)^2)$ are uniformly bounded for all $L \in \mathbb{N}$. Then, there exists a constant $c > 0$ such that

$$\forall L \in \mathbb{N} : \quad |(\mathbb{E}^{\gamma^\delta} - \mathbb{E}^{\gamma^L})[\ell(P)]| \le c 2^{-L} \; .$$

We now bound the term II_L. To this end, we note that

$$\begin{aligned} |II_L| \;\; &\le \;\; \sum_{l=1}^{L} 2 \left(\mathbb{E}^{\gamma^l}(\ell(P - P^{L'(l)})^2) + \mathbb{E}^{\gamma^{l-1}}(\ell(P - P^{L'(l)})^2) \right)^{1/2} d_{\mathrm{Hell}}(\gamma^l, \gamma^{l-1}) \\ &\quad + c\mathbb{E}^{\gamma_{\boldsymbol{b}}}[|\ell(P - P^{L'(0)})|]. \end{aligned}$$

From (4.7) we have that

$$\mathbb{E}^{\gamma^l}(\ell(P - P^{J_{L'(l)}})^2) \le c\mathbb{E}^{\gamma}(\ell(P - P^{L'(l)})^2) \le c2^{-2L'(l)},$$

$$\mathbb{E}^{\gamma^{l-1}}(\ell(P - P^{L'(l)})^2) \le c\mathbb{E}^{\gamma}(\ell(P - P^{J_{L'(l)}})^2) \le c2^{-2L'(l)},$$

and

$$\mathbb{E}^{\gamma^b}[|\ell(P - P^{L'(l_0)})|] \le c2^{-L'(0)}.$$

From Proposition 4.5, there exists a constant $c > 0$ such that for all $l \in \mathbb{N}$

$$d_{\mathrm{Hell}}(\gamma^l, \gamma^{l-1}) \le c2^{-l}.$$

Therefore

$$|II_L| \le c \sum_{l=0}^{L} 2^{-(l + L'(l))}$$

We now estimate III_L. Using inequalities $1 + x \le \exp(x)$ and $x \le \varepsilon \exp(x/\varepsilon)$ for $x, \varepsilon > 0$, we have from (4.12) that there exists a constant $C > 0$ such that, for every $u \in \Gamma_{\boldsymbol{b}}$ (all sums involving $\bar{b}_j$ in the bounds are finite) and for every l, l', J holds

$$\begin{aligned} & |\ell(P^{J_{l'}, l'}(u) - P^{J_{l'-1}, l'-1}(u))| \\ \le \;\; & C \exp\left(5 \sum_{j=1}^{\infty} b_j |u_j|\right) \left(2^{-l'}(1 + \sum_{j=1}^{J_{l'}} \bar{b}_j |u_j|) + \sum_{j=J_{l'-1}+1}^{\infty} b_j |u_j|\right) \\ \le \;\; & C \exp\left(5 \sum_{j=1}^{\infty} b_j |u_j|\right) \left(2^{-l'} \exp(\sum_{j=1}^{J_{l'}} \bar{b}_j |u_j|) + \varepsilon' \exp(\frac{1}{\varepsilon'} \sum_{j > J_{l'-1}} b_j |u_j|)\right) \\ \text{(A.6)} \le \;\; & C 2^{-l'} \exp\left(5 \sum_{j=1}^{\infty} b_j |u_j| + \sum_{j=1}^{J_{l'}} \bar{b}_j |u_j| + \frac{1}{\varepsilon'} \sum_{j > J_{l'-1}} b_j |u_j|\right) \end{aligned}$$

where $\varepsilon' = \sum_{j > J_{l'-1}} b_j$ is as in the definition of $\mathcal{V}^{ll'}$ in (A.3). Furthermore, for every $u \in \Gamma_{\boldsymbol{b}}$,

$$\begin{aligned} & |1 - \exp(\Phi^{J_l, l}(u; \delta) - \Phi^{J_{l-1}, l-1}(u; \delta))| \\ & \quad \le |\Phi^{J_l, l}(u; \delta) - \Phi^{J_{l-1}, l-1}(u; \delta)|(1 + \exp(\Phi^{J_l, l}(u; \delta) - \Phi^{J_{l-1}, l-1}(u; \delta))) \; . \end{aligned}$$

Therefore

$$|1 - \exp(\Phi^{J_l, l}(u; \delta) - \Phi^{J_{l-1}, l-1}(u; \delta))| I^l(u) \le 2|\Phi^{J_l, l}(u; \delta) - \Phi^{J_{l-1}, l-1}(u; \delta)|.$$

Thus, there exists a constant $C > 0$ such that for every $u \in \Gamma_{\boldsymbol{b}}$ and for every $l, l', J \in \mathbb{N}$ holds

$$
\begin{aligned}
&|1-\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))|I^l(u)\\
&\le C\exp\left(6\sum_{j=1}^{\infty}b_j|u_j|\right)\left(2^{-l}(1+\sum_{j=1}^{J_l}\bar b_j|u_j|)+\sum_{j=J_{l-1}+1}^{\infty}b_j|u_j|\right)\\
&\le C\exp\left(6\sum_{j=1}^{\infty}b_j|u_j|\right)\left(2^{-l}\exp(\sum_{j=1}^{J_l}\bar b_j|u_j|)+\varepsilon\exp(\frac{1}{\varepsilon}\sum_{j>J_{l-1}}b_j|u_j|)\right)\\
&\le C2^{-l}\exp\left(6\sum_{j=1}^{\infty}b_j|u_j|+\sum_{j=1}^{J_l}\bar b_j|u_j|+\frac{1}{\varepsilon}\sum_{j>J_{l-1}}b_j|u_j|\right). \qquad \text{(A.7)}
\end{aligned}
$$

where we define $\varepsilon := \sum_{j>J_{l-1}} b_j$. We thus obtain a constant $c > 0$ such that, for every $u \in \Gamma_{\boldsymbol{b}}$ and for every $l, l', J \in \mathbb{N}$ holds

$$
\begin{aligned}
&|A_1^{ll'}(u)|\\
&=|1-\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))|(\ell(P^{J_{l'},l'}(u))-\ell(P^{J_{l'-1},l'-1}(u)))I^l(u)\\
&\le c2^{-(l+l')}\exp\left(11\sum_{j=1}^{\infty}b_j|u_j|+2\sum_{j=1}^{J_L}\bar b_j|u_j|+\frac{1}{\varepsilon}\sum_{j>J_{l-1}}b_j|u_j|+\frac{1}{\varepsilon'}\sum_{j>J_{l'-1}}b_j|u_j|\right)\\
&\le c2^{-(l+l')}\mathcal{V}^{ll'}(u). \qquad \text{(A.8)}
\end{aligned}
$$

From Assumption A.2, this implies the existence of a constant $C > 0$ such that, for every $l, l' \in \mathbb{N}$,

$$\mathbf{E}_L\left[\left|\mathbb{E}^{\gamma^l}[A_1^{ll'}]-E^{\gamma^l}_{M_{ll'}}[A_1^{ll'}]\right|\right]\le CM_{ll'}^{-1/2}2^{-(l+l')}.$$

Similarly, for every $u \in \Gamma_{\boldsymbol{b}}$ we have $|A_2^{ll'}(u)| \le 2^{-(l+l')}\mathcal{V}^{ll'}(u)$. Therefore, there is a constant $C > 0$ such that for every $u \in \Gamma_{\boldsymbol{b}}$ and for every $l, l' \in \mathbb{N}$ holds

$$\mathbf{E}_L\left[\left|\mathbb{E}^{\gamma^l}[A_2^{ll'}]-E^{\gamma^l}_{M_{ll'}}[A_2^{ll'}]\right|\right]\le CM_{ll'}^{-1/2}2^{-(l+l')}.$$

To estimate the term $|\mathbb{E}^{\gamma^l}[A_3^l]\cdot\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}]-E^{\gamma^l}_{M_{ll'}}[A_3^l]\cdot E^{\gamma^{l-1}}_{M_{ll'}}[A_4^{ll'}]|$, we observe

$$
\begin{aligned}
&\mathbf{E}_L\left[\left|\mathbb{E}^{\gamma^l}[A_3^l]\cdot\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}]-E^{\gamma^l}_{M_{ll'}}[A_3^l]\cdot E^{l-1}_{M_{ll'}}[A_4^{ll'}]\right|\right]\\
&\le\mathbf{E}_L\left[\left|\left(\mathbb{E}^{\gamma^l}[A_3^l]-E^{\gamma^l}_{M_{ll'}}[A_3^l]\right)\cdot\mathbb{E}^{\gamma^{l-1}}[|A_4^{ll'}|]\right|\right]+\mathbf{E}_L\left[\left|\left(\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}]-E^{\gamma^{l-1}}_{M_{ll'}}[A_4^{ll'}]\right)\cdot E^{\gamma^l}_{M_{ll'}}[A_3^l]\right|\right]\\
&\le\mathbf{E}_L\left[\left(\mathbb{E}^{\gamma^l}[A_3^l]-E^{\gamma^l}_{M_{ll'}}[A_3^l]\right)^2\right]^{1/2}\cdot\mathbb{E}^{\gamma^{l-1}}[|A_4^{ll'}|]\\
&+\mathbf{E}_L\left[\left(\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}]-E^{\gamma^{l-1}}_{M_{ll'}}[A_4^{ll'}]\right)^2\right]^{1/2}\cdot\mathbf{E}_L\left[E^{\gamma^l}_{M_{ll'}}[A_3^l]^2\right]^{1/2}.
\end{aligned}
$$

From (A.6) and (A.7), for every $u \in \Gamma_{\boldsymbol{b}}$ and for every $l \in \mathbb{N}$ holds $|A_3^l(u)| \le c2^{-l}\mathcal{V}^{ll'}(u)$ and $|A_4^{ll'}(u)| \le c2^{-l'}\mathcal{V}^{ll'}(u)$. The geometric ergodicity, Assumption A.2, then implies that there exists a constant $c > 0$ such that for every $l, l' \in \mathbb{N}$ holds

$$\mathbf{E}_L\left[\left|\mathbb{E}^{\gamma^l}[A_3^l]\cdot\mathbb{E}^{\gamma^{l-1}}[A_4^{ll'}]-E^{\gamma^l}_{M_{ll'}}[A_3^l]\cdot E^{l-1}_{M_{ll'}}[A_4^{ll'}]\right|\right]\le cM_{ll}^{-1/2}2^{-(l+l')}.$$

The remaining expectations in $E_L^{MLMCMC}[\ell(P)] - III_L$ are similarly estimated, resulting in the same bound. We thus have proved

(A.9)

$$\mathbf{E}_L[|III_L|] \le c\sum_{l=1}^{L}\sum_{l'=1}^{L'(l)} M_{ll'}^{-1/2}2^{-(l+l')} + c\sum_{l=1}^{L} M_{l0}^{-1/2}2^{-l} + c\sum_{l'=0}^{L'(l_0)} M_{0l'}^{-1/2}2^{-l'} + cM_{00}^{-1/2}\ .$$

We choose

$$\text{(A.10)}\quad L'(l) := L - l, \quad \text{and} \quad M_{ll'} := 2^{2(L-(l+l'))} \quad \text{for} \quad l \ge 1,\ l' \ge 1,$$
$$M_{l0} = M_{0l} = 2^{2(L-l)}/L^2 \quad \text{and} \quad M_{00} = 2^{2L}/L^4.$$

We then have

$$\mathbf{E}_L[|III_L|] \le c\sum_{l=0}^{L}(L-l)2^{-L} + cL^2 2^{-L} + cL^2 2^{-L} \le CL^2 2^{-L}\ .$$

This bound is, up to logarithmic terms, of the same order as the discretization error of one instance of the forward problem on the finest mesh level L.

Next, we estimate the total number of degrees of freedom and floating point operations required to realize the MLMCMC-FEM.

For each proposal $v^{(k)}$, the number of degrees of freedom for computing $\Phi^l(v^{(k)})$ is $O(2^{dl})$. The total number of degrees of freedom required for running the MLMCMC-FEM at discretization level L is bounded by

$$\begin{aligned}
&\lesssim \sum_{l=1}^{L}\sum_{l'=0}^{L'(l)} M_{ll'}(2^{dl} + 2^{dl'}) + \sum_{l'=0}^{L'(0)} M_{0l'}2^{dl'} \\
&= 2^{2L}\sum_{l=0}^{L}\sum_{l'=0}^{L'(l)}\left(2^{(d-2)l}\cdot 2^{-2l'} + 2^{-2l}\cdot 2^{(d-2)l'}\right) + 2^{2L}\sum_{l'=0}^{L'(0)} 2^{(d-2)l'} \\
&\lesssim 2^{2L}\left(\sum_{l=0}^{L}2^{(d-2)l} + \sum_{l=0}^{L}2^{-2l}\sum_{l'=0}^{L-l}2^{(d-2)l'} + \sum_{l'=0}^{L}2^{(d-2)l'}\right).
\end{aligned}$$

For space dimension $d = 2$, the number of degrees of freedom in the MLMCMC-FEM at discretization level L is bounded by

$$\lesssim 2^{2L}\left(L + \sum_{l=0}^{L}2^{-2l}(L-l)\right) \lesssim L2^{2L}.$$

For $d = 3$, it is bounded by

$$\lesssim 2^{2L}\left(2^L + \sum_{l=0}^{L}2^{-2l}2^{L-l}\right) \lesssim 2^{2L}\left(2^L + \sum_{l=0}^{L}2^L 2^{-3l}\right) \lesssim 2^{3L}\ .$$

From Lemma 4.3, the number of iterations to approximately solve the system in (4.3) for proposal v at FE meshwidth h is bounded as

$$j^*(v) \ge C\left(\log|h| + |\log \check{K}(v)|\right)\sqrt{\frac{\hat{K}(v)}{\check{K}(v)}}\ .$$

With $h \simeq 2^{-l}$ and with the truncation level $J_l = \lceil 2^{l/q}\rceil$, the total work for performing one step of the MCMC process for solving the linear system for a proposal

$v \in \Gamma_{\boldsymbol{b}}$ is asymptotically, as $l \to \infty$, and on proposal v bounded by

$$\lesssim l^{d-1} 2^{l/q+ld} + (\hat{K}(v))^{1/2} (\check{K}(v))^{-1/2} (1 + |\log \check{K}(v)|) l^d 2^{ld}.$$

The expectation of the overall number of floating point operations required to compute $\Phi^{J_l,l}(v^{(k)};\delta)$ is bounded by

$$\lesssim l^{d-1} 2^{l/q+ld} + l^d 2^{ld} \lesssim l^{d-1} 2^{l/q+ld}.$$

Therefore, the expectation of the total number of floating point operations to run one step of the MLMCMC-FEM at discretization level l is not larger than

$$\lesssim l^{d-1} 2^{l/q+dl} + l'^{d-1} 2^{l'/q+dl'}.$$

With the choice (A.10) for the number of MCMC steps, the total number of floating point operations required to evaluate the MLMCMC-FEM estimator at discretization level L is bounded by

$$\begin{aligned}
&\lesssim \sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} M_{ll'} (l^{d-1} 2^{l/q+dl} + l'^{d-1} 2^{l'/q+dl'}) + \sum_{l'=0}^{L'(0)} M_{0l'} (l'^{d-1} 2^{l'/q+dl'}) \\
&\lesssim 2^{2L} \sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} (l^{d-1} 2^{(d-2+1/q)l} 2^{-2l'} + l'^{d-1} 2^{(d-2+1/q)l'} 2^{-2l}) + 2^{2L} \sum_{l'=0}^{L'(l_0)} l'^{d-1} 2^{(d-2+1/q)l'} \\
&\lesssim L^{d-1} 2^{2L} \left(\sum_{l=0}^{L} 2^{(d-2+1/q)l} + \sum_{l=0}^{L} 2^{(d-2+1/q)(L-l)} 2^{-2l} \right) \\
&\lesssim L^{d-1} 2^{(d+1/q)L}.
\end{aligned}$$

□

Appendix B.

We justify Assumption A.2 in this appendix. We first consider the case of the independence sampler. We then address the pCN sampling method.

Lemma B.1. *For the independence MCMC sampler with acceptance probability $\alpha^{J,l}$ as defined in* (4.13)*, for the equation with the J-term truncated, parametric coefficient* (3.1) *and at discretization level l, the normalizing constant $Z^{J,l}$ is bounded from below away from zero, uniformly for all J and l.*

Proof From (4.4) and (3.6), we have

$$\Phi^{J,l}(u,\delta) \le c \|\Sigma^{-1}\|_{\mathbb{R}^{k\times k}} (|\delta|^2 + |\mathcal{G}^{J,l}(u)|^2) \le c \|\Sigma^{-1}\|_{\mathbb{R}^{k\times k}} (|\delta|^2 + c \exp(\sum_{j=J}^{J} (2b_j |u_j|)))$$

where the constant c is independent of δ and u. For simplicity, we denote the restriction of γ and $\gamma_{\boldsymbol{b}}$ on $\mathbb{R}^J$ as γ. Therefore

$$\int_{\mathbb{R}^J} \Phi^{J,l}(u,\delta) d\gamma(u) < c \|\Sigma^{-1}\|_{\mathbb{R}^{k\times k}} (1 + \delta^2) := c^*$$

uniformly for all J and l. Thus, for each $C > 0$

$$\gamma(\{u : \ \Phi^{J,l}(u;\delta) > C\}) < c^*/C. \tag{B.1}$$

Choosing C sufficiently large,

$$\gamma(\{u : \ \Phi^{J,l}(u;\delta) < C\}) > 1 - c^*/C > c_0 > 0.$$

We have, for every J and l,

$$Z^{J,l} = \int_{\mathbb{R}^J} \exp(-\Phi^{J,l}(v;\delta))d\gamma(v) > \exp(-C)(1 - c^*/C) \ .$$

□

Lemma B.2. *Let $u^{(j)}$ be the jth draw in the Markov chain generated by the MCMC independence sampler with the acceptance probability (4.13); let further $\mathcal{E}^{\tilde{\gamma},J,l}$ denote the expectation with respect to the probability space generated by the Markov chain with the initial sample $u^{(0)}$ being distributed according to the restriction of $\bar{\gamma}$ to $\mathbb{R}^J$, still denoted as $\bar{\gamma}$. For $g \in L^2(U, \gamma_{\boldsymbol{b}})$, let $\bar{g} = g - \mathbb{E}^{\gamma^{J,l}}[g]$. We have*

$$\mathcal{E}^{\tilde{\gamma},J,l}\left[\left|\frac{1}{M}\sum_{k=1}^{M}\bar{g}(u^{(k)})\right|^2\right] \le c\mathbb{E}^{\gamma_{\boldsymbol{b}}}[|g|^2]$$

where c does not depend on g, J and l.

Proof Adopting the notation of [32], we denote for $J, l \in \mathbb{N}$ and for arbitrary $u \in \mathbb{R}^J$

$$w^{J,l}(u) = \frac{d\gamma^{J,l,\delta}(u)}{d\gamma(u)},$$

and define, for each $w \in \mathbb{R}_+$,

$$\tilde{\gamma}^{J,l,\delta}(w) = \gamma^{J,l,\delta}(\{u : w^{J,l}(u) \le w\}) \ . \tag{B.2}$$

For conciseness we will drop the superscript δ in $\gamma^{J,l,\delta}$ and $\tilde{\gamma}^{J,l,\delta}$ in the remainder of the proof.

Let $p^j(u,\cdot)$ be the jth iterate of the transition kernel of the Markov chain. When the current state is u, the probability that a draw is rejected equals

$$\int_{\{v : w^{J,l}(v) \le w^{J,l}(u)\}} \left\{1 - \frac{w^{J,l}(v)}{w^{J,l}(u)}\right\} d\gamma(v). \tag{B.3}$$

This probability only depends on $w^{J,l}(u)$. Following [32], we denote this probability as $\lambda^{J,l}(w)$ when $w = w^{J,l}(u) \in \mathbb{R}_+$.

From Theorem 1 of [32], we have

$$p^j(u, dv) = T_j(w^{J,l}(u) \vee w^{J,l}(v))\gamma^{J,l}(dv) + \lambda^{J,l}(w^{J,l}(u))^j \delta_u(dv),$$

where, for arbitrary $w \in \mathbb{R}_+$ and $j \in \mathbb{N}$, we defined

$$T_j(w) = 1 - \frac{\lambda^{J,l}(w)^j}{\tilde{\gamma}^{J,l}(w)} + \int_{t>w} \frac{\lambda^{J,l}(t)^j}{(\tilde{\gamma}^{J,l}(t))^2} d\tilde{\gamma}^{J,l}(t) \ .$$

We then have

$$p^j(u^{(0)}, dv) - \gamma^{J,l}(dv) = \left(\int_{t > w^{J,l}(u^{(0)}) \vee w^{J,l}(v)} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(t)} d\tilde{\gamma}^{J,l}(t)\right.$$
$$\left. - \frac{\lambda^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))^j}{\tilde{\gamma}^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))}\right) \gamma^{J,l}(dv) + (\lambda^{J,l}(w^{J,l}(u^{(0)})))^j \delta_{u^{(0)}}(dv).$$

As $w^{J,l}(u) = \frac{1}{Z^{J,l}} \exp(-\Phi^{J,l}(u;\delta)) \le \frac{1}{Z^{J,l}} \le a$, where $1/a$ denotes the uniform lower bound of $Z^{J,l}$ proved in the previous lemma, for $w \ge a$ $\tilde{\gamma}^{J,l}(w) = 1$. As shown in [32, Section 3] $\frac{d}{dt}\lambda^{J,l}(t) = \tilde{\gamma}^{J,l}(t)/t^2$. In particular, $\lambda^{J,l}(t)$ is increasing. Moreover, as

$w^{J,l}(u) \le a$, when $t \ge a$, $\lambda^{J,l}(t) = 1 - 1/t$. Thus $\lambda^{J,l}(w^{J,l}(u^{(0)})) \le \lambda^{J,l}(a) = 1 - 1/a$. Moreover,

$$\left| \int_{\mathbb{R}^J} (\lambda^{J,l}(w^{J,l}(u^{(0)})))^j g(v) \delta_{u^{(0)}}(dv) \right| \le \left(1 - \frac{1}{a}\right)^j |g(u^{(0)})|.$$

Therefore

$$\begin{aligned}
&\int_{t > w^{J,l}(u^{(0)}) \vee w^{J,l}(v)} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(t)} d\tilde{\gamma}^{J,l}(t) \\
&= \int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(t)} d\tilde{\gamma}^{J,l}(t) \\
&= \int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} (\lambda^{J,l}(t))^j d(-\frac{1}{\tilde{\gamma}^{J,l}(t)}) \\
&= -\frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)} \Big|_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} + \int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} \frac{1}{\tilde{\gamma}^{J,l}(t)} d(\lambda^{J,l}(t))^j.
\end{aligned}$$

Therefore

$$\begin{aligned}
\int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(t)} d\tilde{\gamma}^{J,l}(t) = -\left(1 - \frac{1}{a}\right)^j + \frac{\lambda^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))^j}{\tilde{\gamma}^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))} \\
+ \int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} \frac{j(\lambda^{J,l}(t))^{j-1}}{t^2} dt.
\end{aligned}$$

From this

$$\begin{aligned}
\int_{t > w^{J,l}(u^{(0)}) \vee w^{J,l}(v)} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(t)} d\tilde{\gamma}^{J,l}(t) - \frac{\lambda^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))^j}{\tilde{\gamma}^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))} \\
= -\left(1 - \frac{1}{a}\right)^j + \int_{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)}^{a} \frac{j(\lambda^{J,l}(t))^{j-1}}{t^2} dt.
\end{aligned}$$

As $d\gamma^{J,l}(v) = w^{J,l}(v) d\gamma(v)$ and $w^{J,l}(v) \le a$ uniformly for all J and l, there exists a constant $c > 0$ such that for every J and l

$$\begin{aligned}
&\int_{\mathbb{R}^J} \left(\int_{t > w^{J,l}(u^{(0)}) \vee w^{J,l}(v)} \frac{(\lambda^{J,l}(t))^j}{\tilde{\gamma}^{J,l}(t)\tilde{\gamma}^{J,l}(v)} d\tilde{\gamma}^{J,l}(t) - \frac{(\lambda^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))^j}{\tilde{\gamma}^{J,l}(w^{J,l}(u^{(0)}) \vee w^{J,l}(v))} \right) g(v) \gamma^{J,l}(dv) \\
&\quad \le \left(1 - \frac{1}{a}\right)^j \int_U |g(v)| \gamma^{J,l}(dv) + j \left(1 - \frac{1}{a}\right)^{j-1} \int_{\mathbb{R}^J} \frac{1}{w^{J,l}(u^{(0)}) \vee w^{J,l}(v)} |g(v)| d\gamma^{J,l}(v) \\
&\quad \le \left(1 - \frac{1}{a}\right)^j \mathbb{E}^{\gamma^{J,l}}[|g|] + j \left(1 - \frac{1}{a}\right)^{j-1} \mathbb{E}^{\gamma}[|g|] .
\end{aligned}$$

This implies

$$\left| (\mathbb{E}^{p^j(u^{(0)},\cdot)} - \mathbb{E}^{\gamma^{J,l}})[g] \right| \le \left(1 - \frac{1}{a}\right)^j |g(u^{(0)})| + \left(1 - \frac{1}{a}\right)^j \mathbb{E}^{\gamma^{J,l}}[|g|] + j \left(1 - \frac{1}{a}\right)^{j-1} \mathbb{E}^{\gamma}[|g|].$$

Let $\mathcal{E}^{\gamma^{J,l}}$ be the expectation with respect to the MCMC process with the initial sample distributed according to $\gamma^{J,l}$. Let further $\mathcal{E}_{u^{(0)}}$ denote the expectation with respect to the MCMC process starting at $u^{(0)}$. Then we calculate, following [26],

$$
\begin{aligned}
\frac{1}{M}\mathcal{E}^{\gamma^{J,l}}\Big[\Big|\sum_{k=1}^{M}\bar g(u^{(k)})\Big|^2\Big] &= \mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=1}^{M}\sum_{j=k+1}^{M}\mathcal{E}^{\gamma^{J,l}}[\bar g(u^{(k)})\bar g(u^{(j)})] \\
&= \mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathcal{E}^{\gamma^{J,l}}[\bar g(u^{(0)})\bar g(u^{(j)})] \\
&= \mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})\mathcal{E}_{u^{(0)}}[\bar g(u^{(j)})]] \\
&\le \mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})^2] \\
&\quad + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathbb{E}^{\gamma^{J,l}}[|\bar g(u^{(0)})||\mathcal{E}_{u^{(0)}}[g(u^{(j)})] - \mathbb{E}^{\gamma^{J,l}}[g]|] \\
&\le \mathbb{E}^{\gamma^{J,l}}[\bar g(u^{(0)})^2] \\
&\quad + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathbb{E}^{\gamma^{J,l}}\Big[|\bar g(u^{(0)})|\cdot \\
&\quad \left(\left(1-\frac{1}{a}\right)^j|g(u^{(0)})| + \left(1-\frac{1}{a}\right)^j\mathbb{E}^{\gamma^{J,l}}[|g|] + j\left(1-\frac{1}{a}\right)^{j-1}\mathbb{E}^{\gamma}[|g|]\right)\Big] \\
&\le \sum_{j=1}^{\infty}\left(1-\frac{1}{a}\right)^j\mathbb{E}^{\gamma^{J,l}}[g^2] + 3\sum_{j=1}^{\infty}\left(1-\frac{1}{a}\right)^j(\mathbb{E}^{\gamma^{J,l}}[|g|])^2 + \\
&\qquad 2\sum_{j=1}^{\infty}j\left(1-\frac{1}{a}\right)^{j-1}\mathbb{E}^{\gamma^{J,l}}[|g|]\mathbb{E}^{\gamma}[|g|]
\end{aligned}
$$

where we have used $|\bar g| \le |g| + \mathbb{E}^{\gamma^{J,l}}[|g|]$. We note that

$$\mathbb{E}^{\gamma^{J,l}}[g^2] \le \frac{1}{Z^{J,l}}\mathbb{E}^{\gamma}[g^2];$$

and $(\mathbb{E}^{\gamma^{J,l}}[|g|])^2 \le \mathbb{E}^{\gamma^{J,l}}[|g|^2]$, $(\mathbb{E}^{\gamma}[|g|])^2 \le \mathbb{E}^{\gamma}[|g|^2]$. Thus

$$\frac{1}{M}\mathcal{E}^{\gamma^{J,l}}\Big[\Big|\sum_{k=1}^{M}\bar g(u^{(k)})\Big|^2\Big] \le a(2a^2+4a-4)\mathbb{E}^{\gamma}[g^2]. \tag{B.4}$$

From (4.17), we get the conclusion. □

Before justifying Assumption A.2, we recall the following result.

Lemma B.3. [19, Appendix] *For any $t > 0$*

$$\int_{-\infty}^{\infty}\exp(-z^2/2 + |z|t)\frac{dz}{\sqrt{2\pi}} \le \exp(t^2/2)\exp(t\sqrt{2/\pi}).$$

Proposition B.4. *For the independence sampler with the acceptance probability* (4.13)*, Assumption A.2 holds.*

Proof It suffices to show that $\mathbb{E}^{\gamma}[\mathcal{V}^{ll'}(\cdot)^2]$ is uniformly bounded with respect to l and l'. We may assume that $J_{l-1} \le J_{l'-1}$ (the argument in the case $J_{l-1} > J_{l'-1}$

is similar). From Lemma B.3, we have the majorization

$$\begin{aligned}
&\mathbb{E}^\gamma[(\mathcal{V}^{ll'})^2]\\
&\le \exp\left(\frac{1}{2}22^2\sum_{j=1}^{J_{l-1}}(b_j+\bar b_j)^2+\sum_{j=J_{l-1}+1}^{J_{l'-1}}\frac{1}{2}\left(22(b_j+\bar b_j)+\frac{2}{\varepsilon}b_j\right)^2+\sum_{j=J_{l'-1}+1}^{\infty}\frac{1}{2}\left(22(b_j+\bar b_j)+\frac{2}{\varepsilon}b_j+\frac{2}{\varepsilon'}b_j\right)^2\right)\\
&\cdot\exp\left(\left(22\sum_{j=1}^{J_{l-1}}(b_j+\bar b_j)+\sum_{j=J_{l-1}+1}^{J_{l'-1}}\left(22(b_j+\bar b_j)+\frac{2}{\varepsilon}b_j\right)+\sum_{j=J_{l'-1}+1}^{\infty}\left(22(b_j+\bar b_j)+\frac{2}{\varepsilon}b_j+\frac{2}{\varepsilon'}b_j\right)\right)\sqrt{\frac{2}{\pi}}\right)\\
&\le \exp\left(c\sum_{j=1}^{\infty}(b_j^2+\bar b_j^2+b_j+\bar b_j)+c\frac{1}{\varepsilon}\sum_{j>J_{l-1}}b_j+c\frac{1}{\varepsilon'}\sum_{j'>J_{l'-1}}b_{j'}\right)
\end{aligned}$$

which is finite. □

The preceding analysis established geometric ergodicity of the independence sampler. For the pCN sampler, the proposal $v^{(k)}\in\mathbb{R}^J$ is chosen as

$$v^{(k)}=\sqrt{1-\beta^2}u^{(k)}+\beta\xi,$$

where $\xi\sim N(0,I)$ in $\mathbb{R}^J$ where I is the $J\times J$ identity matrix. Although the growth conditions which are necessary for the L^2_μ spectral gap results of Hairer et al. [17] to hold have *not* been verified for the forward problem with log-gaussian coefficient (4.1), it is quite straightforward to show that:

Proposition B.5. *Assume that the $L^2_{\gamma^\delta}$ spectral gap result of* [17] *holds. Then, Assumption A.2 on geometric ergodicity holds for the pCN sampler for the forward problem* (2.2) *with log-gaussian coefficient* (4.1).

Appendix C.

For the readers' convenience, we summarize the method using circulant embedding for evaluating samples of the GRF R at evenly spaced FE nodes, as proposed e.g. in [12]. Our presentation follows [16, Sec.5].

In the domain $D=(0,1)\times(0,1)\subset\mathbb{R}^2$, we consider the restriction to D of the stationary (in $\mathbb{R}^2$) GRF R with mean $\mu(x)=0$, and with covariance function

$$C(x,y)=\rho(|x-y|),\quad x,y\in D$$

where $\rho(t)$ attains positive values for $t\ge 0$ ($|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^2$).

Consider a uniform partition of D into axiparallel squares of meshwidth $h=2^{-L}$. On this mesh, we consider affine $\mathbb{Q}_1$ FE-approximations which converge, for solutions with $H^2(D)$-regularity, with the rate $O(h)$ in $H^1(D)$. We define $M=2^L$ so that $h=1/M$ and denote the $(M+1)^2$ equispaced points in $\overline{D}$ which describe the mesh as $x_{i,j}=(ih,jh)$ for $i,j\in\{0,\cdots,M\}$. We identify the FE nodes by the double index (i,j). The FE nodes are enumerated lexicographically, that is, first in the x_1 direction then in the x_2 direction. We consider the $(M+1)^2\times(M+1)^2$ covariance matrix

(C.1)
$$T_{(i,j),(k,l)}=\rho\left(|x_{i,j}-x_{k,l}|\right)=\rho\left(|(ih,jh)-(kh,lh)|\right)=\rho\left(|((i-k)h,(j-l)h)|\right).$$

Our purpose is to generate the random realization of the centered gaussian random vector $R(x_{i,j})$ with covariance T. We note that if $T=SS^\top$ where S is a symmetric matrix then, if Θ is a $(M+1)^2$ random vector where each component follows the standard normal distribution, $S\Theta$ is the random vector with the desired covariance. By embedding T into a larger block circulant matrix, exploiting the homogeneity of the random field R and the even spacing of the FE nodes $x_{i,j}$, we can evaluate the matrix S very efficiently using FFT as we explain next (see also [12]). Let $m \in \mathbb{N}_0$. We consider the covariance vector of points belonging to two different rows that are m rows apart, i.e. we consider the $(M+1)\times(M+1)$ Toeplitz block matrix T_m of T defined as

$$(T_m)_{i,k} = T_{(i,j),(k,j+m)} = \rho\left(|x_{i,j} - x_{k,j+m}|\right) = \rho\left(|((i-k)h, mh)|\right). \tag{C.2}$$

for $i, j, k \in 0, \dots, M$. We note that $(T_m)_{i,k}$ is independent of j. The covariance matrix T is Symmetric Block Toeplitz (SBT for short) with the first block row being $T_0, \cdots, T_M$. We denote this as

$$T = \mathrm{SBT}(T_0, \cdots, T_M).$$

We embed T into a circulant matrix C whose blocks are themselves symmetric circulant. This is done by first possibly extending the domain on which the random field R is defined, beyond the domain D. This process is called padding. For $J \in \mathbb{N}_0$ we define the extended domain $D_J = [0, 1+Jh] \times [0, 1+Jh]$. We consider a uniform partition into axiparallel squares of edgelength h in D_J and the covariance matrix of the GRF sampled at the corresponding mesh points. We embed this matrix into a larger circulant one by *mirroring*. The aim of the padding process is to obtain a positive definite, circulant matrix. In many cases, the padding process is not necessary, i.e. we can just take $J = 0$. For $k, l \in \mathbb{N}_0$, we denote by $\rho_{kl} = \rho(|(kh, lh)|)$ for $(kh, lh) \in \mathbb{R}^2$. We define the circulant block matrix

$$C_j = \mathrm{SBT}(\rho_{0,j}, \cdots, \rho_{M,j}, \rho_{M+1,j}, \cdots, \rho_{M+J,j}, \rho_{M+J-1,j}, \cdots, \rho_{1,j}),$$

for $j = 0, \cdots, M+J$. The block circulant matrix into which we embed the covariance matrix T is

$$C = \mathrm{SBT}(C_0, \cdots, C_M, C_{M+1}, \cdots, C_{M+J}, C_{M+J-1}, \cdots, C_1).$$

The matrix C has dimension $n^2 \times n^2$ where $n = 2(M+J)$. As C is block circulant, it can be efficiently diagonalized by FFT. Indeed, $C = Q\Lambda Q^H$ where

$$Q_{(j_1,j_2),(k_1,k_2)} = \frac{1}{n}\exp\left(\frac{2\pi\sqrt{-1}j_1k_1}{n}\right)\exp\left(\frac{2\pi\sqrt{-1}j_2k_2}{n}\right)$$

for $j_1, j_2, k_1, k_2 = 0, \cdots, n-1$. Here, Λ is the diagonal matrix that contains the eigenvalues of C. Let $G = \mathrm{Re}Q + \mathrm{Im}Q$. Then G is a symmetric and orthogonal matrix such that $C = G\Lambda G^\top$.

For performing the MLMCMC-FEM using circulant embedding, for the FE solution at two consecutive resolution levels of a realization of the forward equation, as the FE meshes are nested, we sample the GRF at the (equispaced nodes at) mesh level l and extract, from this sample, also the values of the GRF at the nodes of mesh level $l-1$, for every discretization level $l \geq 1$.

The stiffness matrix in the FE approximation of the forward equation is numerically evaluated with the trapezoidal rule in each co-ordinate x_1 and x_2 for each mesh cell using the values of the GRF coefficient sample K at the mesh nodes. Newton-Côtes quadrature of fixed order (adapted to the $\mathbb{Q}_1$-FEM in D) can be used by generating evenly spaced quadrature points inside each mesh cell.

References

[1] Alexandros Beskos, Mark Girolami, Shiwei Lan, Patrick E. Farrell, and Andrew M. Stuart. Geometric MCMC for infinite-dimensional inverse problems. *J. Comput. Phys.*, 335:327–351, 2017.

[2] Alexandros Beskos, Ajay Jasra, Nikolas Kantas, and Alexandre Thiery. On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.*, 26(2):1111–1146, 2016.

[3] Alexandros Beskos, Ajay Jasra, Kody Law, Youssef Marzouk, and Yan Zhou. Multilevel sequential Monte Carlo with dimension-independent likelihood-informed proposals. *SIAM/ASA J. Uncertain. Quantif.*, 6(2):762–786, 2018.

[4] Alexandros Beskos, Ajay Jasra, Kody Law, Raul Tempone, and Yan Zhou. Multilevel sequential Monte Carlo samplers. *Stochastic Process. Appl.*, 127(5):1417–1440, 2017.

[5] Alexandros Beskos, Ajay Jasra, Ege A. Muzaffer, and Andrew M. Stuart. Sequential Monte Carlo methods for Bayesian elliptic inverse problems. *Stat. Comput.*, 25(4):727–737, 2015.

[6] Vladimir I. Bogachev. *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[7] Julia Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(1):216–246, 2012.

[8] S. L. Cotter, M. Dashti, J. C. Robinson, and Andrew M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse problems*, 25, 2009.

[9] M. Dashti and A.M.Stuart. The bayesian approach to inverse problems. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*. Springer, 2017.

[10] M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal.*, 49(6):2524–2542, 2011.

[11] Pierre Del Moral, Ajay Jasra, Kody J. H. Law, and Yan Zhou. Multilevel sequential Monte Carlo samplers for normalizing constants. *ACM Trans. Model. Comput. Simul.*, 27(3):Art. 20, 22, 2017.

[12] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, 1997.

[13] T. J. Dodwell, C. Ketelsen, R. Scheichl, and A. L. Teckentrup. A Hierarchical Multilevel Markov Chain Monte Carlo Algorithm with Applications to Uncertainty Quantification in Subsurface Flow. *SIAM/ASA J. Uncertain. Quantif.*, 3(1):1075–1108, 2015.

[14] J. Galvis and M. Sarkis. Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity. *SIAM J. Numer. Anal.*, 47(5):3624–3651, 2009.

[15] Claude J. Gittelson. Stochastic Galerkin discretization of the log-normal isotropic diffusion problem. *Mathematical Models and Methods in Applied Sciences*, 20(2):237–263, 2010.

[16] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications. *J. Comput. Phys.*, 230(10):3668–3694, 2011.

[17] Martin Hairer, Andrew M. Stuart, and Sebastian J. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Probab.*, 24(6):2455–2490, 2014.

[18] Lukas Herrmann. Strong convergence analysis of iterative solvers for random operator equations. Technical Report 2017-35, Seminar for Applied Mathematics, ETH Zürich, 2017. (in review).

[19] V.H. Hoang and Ch. Schwab. Convergence rate analysis of MCMC-FEM for Bayesian inversion of log-normal diffusion problems. Technical Report 2016-19, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2016.

[20] Viet Ha Hoang. Bayesian inverse problems in measure spaces with application to Burgers and Hamilton-Jacobi equations with white noise forcing. *Inverse Problems*, 28(2):025009, 29, 2012.

[21] Viet Ha Hoang, Christoph Schwab, and Andrew Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8), 2013.

[22] Ajay Jasra, Kengo Kamatani, Kody J. H. Law, and Yan Zhou. A multi-index Markov chain Monte Carlo method. *Int. J. Uncertain. Quantif.*, 8(1):61–73, 2018.

[23] Ajay Jasra, Kody J. H. Law, and Yan Zhou. Forward and inverse uncertainty quantification using multilevel Monte Carlo algorithms for an elliptic nonlocal equation. *Int. J. Uncertain. Quantif.*, 6(6):501–514, 2016.

[24] C. Lieberman, K. Willcox, and O. Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):25232542, 2010.
[25] J. Martin, L.C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460A1487, 2012.
[26] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, second edition, 2009.
[27] H. Nguyen. *Finite element wavelets for solving partial differential equations*. PhD thesis, University of Utrecht, 2005.
[28] R. Scheichl, A. M. Stuart, and A. L. Teckentrup. Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):493–518, 2017.
[29] Claudia Schillings and Christoph Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6):065011, 28, 2013.
[30] Claudia Schillings and Andrew M. Stuart. Analysis of the ensemble Kalman filter for inverse problems. *SIAM J. Numer. Anal.*, 55(3):1264–1290, 2017.
[31] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.
[32] Richard Smith and Luke Tierney. Exact transition probabilities for the independence metropolis sampler. Technical report. http://rls.sites.oasis.unc.edu/postscript/rs/exact.ps.
[33] Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 2010.
[34] Karsten Urban. *Wavelet methods for elliptic partial differential equations*. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2009.
[35] Y. Yamasaki. *Measures on infinite-dimensional spaces*, volume 5 of *Series in Pure Mathematics*. World Scientific Publishing Co., Singapore, 1985.

[1] Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, 637371

[2] Seminar for Applied Mathematics, ETH, 8092 Zurich, Switzerland