Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Multilevel MCMC Bayesian Inversion of Parabolic PDEs under Gaussian Prior

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

Research Report No. 2020-48
July 2020

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

MULTILEVEL MCMC BAYESIAN INVERSION OF PARABOLIC PDES UNDER GAUSSIAN PRIOR

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

Abstract. We analyze the convergence of a multi-level Markov Chain Monte-Carlo (MLMCMC) algorithm for the Bayesian estimation of solution functionals for linear, parabolic partial differential equations subject to uncertain diffusion coefficient. The multilevel convergence analysis is performed for a time-independent, log-gaussian diffusion coefficient and for observations which are assumed to be corrupted by additive, centered gaussian observation noise. The elliptic spatial part of the parabolic PDE is neither uniformly coercive nor uniformly bounded in terms of the realizations of the unknown gaussian random field. The path-wise, multi-level discretization in space and time considered is based on standard, first order, Lagrangean simplicial Finite Elements in the spatial domain and on first order, implicit timestepping of backward Euler type, ensuring good dissipation and unconditional stability, and resulting in first order convergence in terms of the spatial meshwidth and the timestep. The MCMC algorithms covered by our analysis comprise the standard, indepence sampler as well as various variants, such as pCN. We prove that the proposed MLMCMC algorithm delivers approximate Bayesian estimates of quantities of interest consistent to first order in the discretization parameter on the finest spatial / temporal discretization stepsize in overall work which scales essentially (i.e., up to terms which depend logarithmically on the discretization parameters) as that of one deterministic solve on the finest mesh. Our convergence analysis is based on the discretization-level dependent truncation of the increments, introduced first in [15] for the corresponding elliptic forward problems. This is required to address measurability and integrability issues encountered in the Bayesian posterior density evaluated at consecutive discretization levels with respect to the gaussian prior. Both, independence sampler and pCN are analyzed in detail. Applicability of our analysis to other versions of MCMC is discussed.

1. Introduction

The numerical analysis of multi-level algorithms for the assimilation of noisy observation data into partial differential equations with uncertain function space input has attracted substantial attention in recent years. Numerous mathematical frameworks for precise formulations and numerical treatment of it have been proposed. We mention only [2, 19] and the references there for an account of recent developments.

In the present paper, we perform a mathematical analysis of several multi-level, Markov-chain Monte Carlo (MCMC) algorithms for numerical approximation in the *Bayesian setting* of PDE inversion, which has been promoted in the series of papers [10, 11, 9, 22], and in the references there.

We adapt our multi-level MCMC framework developed in the elliptic setting in [15] to linear, parabolic evolution equations. Importantly, and distinct from other recent works, e.g. [6, 4, 5], we neither require truncation of the Gaussian prior distribution nor any boundedness from above and below of the pathwise solution to the forward partial differential equations, neither of which holds in the case of the Gaussian prior. Further, as we showed in [15] analytically and numerically, under Gaussian prior, multilevel sampling of the Bayesian posterior can be severely imprecise if the unboundedness of the solution with respect to a random draw of the coefficient from the prior distribution is ignored. In this paper, we develop a multilevel MCMC method for the Bayesian inversion of a linear, parabolic evolution equation subject to a Gaussian prior, given observations on its solution at some moments of time. Our mathematical analysis does not require uniform, almost sure lower-boundedness away from zero of the diffusion coefficient. Therefore, our method is applicable to the most general log-gaussian diffusion coefficient subject to a Gaussian prior.

In the past decade, significant progress has been achieved in the mathematical analysis of convergence of MCMC for PDEs with uncertain function space input. We only mention [20, 14, 17]. In these works, in particular the significance of *geometric ergodicity* and of a *spectral gap condition* for uniform w.r. to the discretization parameters convergence rates of the MCMC algorithms has been identified.

The principal contributions of the present paper are: i) we develop a multilevel MCMC space-time discretization for Bayesian coefficient inversion of a forward problem being a linear parabolic PDE, under gaussian prior, ii) we provide sufficient conditions on the regularity and sparsity of the data to show that the MLMCMC method developed here is capable of numerically approximating the Bayesian estimate of a quantity of interest to full consistency on the finest discretization levels in space and time, of the forward parabolic problem, in computational work equally essentially (i.e., up to logarithmic terms) to that of one deterministic solve of (one instance of) the forward problem.

The outline of this paper is as follows. In Section 2, we present the problem formulation, including in particular our assumptions on the uncertain, log-gaussian diffusion coefficient. We emphasize that, unlike other recent works, the present MLMCMC analysis does not require this coefficient to be lower-bounded away from zero almost surely. Section 3 addresses the existence of the posterior measure and the well-posedness of the Bayesian inverse problem. In Section 4, we address the numerical approximations in the multi-level algorithm, in particular the J-term truncation of the gaussian random field input, and the discretization in the temporal and in the spatial domain. We consider standard, first order discretizations in space and time, to keep the pathwise regularity requirements for the solution moderate, and to have a simulation algorithm which resembles methods used in computational practice. Higher order discretizations, or more sophisticated space-time discretizations could equally be considered. Section 6 presents two series of numerical experiments of our algorithm for a model problem in two space dimensions. One set of experiments each for the indenpendence sampler and one for the pCN sampler. We run these examples with the algorithmic parameters selected according to our ML theory. The numerical results confirm our theoretical analysis and strongly indicate that our theoretical results are sharp.

2. Problem formulation

We present the BIP under consideration. In Section 2.1, we address the parametric description of the log-gaussian random field used to model the uncertain diffusion coefficient. For numerical purpose, and as in many earlier works (e.g. [8, 1, 13, 23, 11] and the references there) we introduce a countably-parametric family of gaussian random fields which arises, for example (but not only) through a Karhúnen-Loève expansion. The countably-parametric functional form is exploited in the ensuing design of a multilevel MCMC algorithm to access the GRF approximately, at various levels of fidelity, through finitely truncated expansions with controlled errors. The gaussian prior measure in the BIP is obtained as the product measure on (sequences of) coefficients in the Karhúnen-Loève expansion.

2.1. **Log-gaussian diffusion coefficient.** In a bounded Lipschitz domain $\mathrm{D} \subset \mathbb{R}^d$ and a bounded time interval $[0, T]$, we consider a linear, second order divergence form parabolic equation with an uncertain coefficient. Specifically, we assume given $K : \mathrm{D} \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ in parametric, log-normal form, which is *formally* given as

$$K(x, z) = K_*(x) + \exp\left(\bar{K}(x) + \sum_{j=1}^{\infty} z_j \psi_j(x)\right), \tag{2.1}$$

where $z = (z_1, z_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$. Further assumptions are required to render (2.1) meaningful.

Assumption 2.1. *The functions $K_*, \bar{K} \in L^\infty(\mathrm{D})$ are non-negative. The functions $\psi_j \in L^\infty(\mathrm{D})$ for $j = 1, 2, \ldots$ are such that $\sum_{j=1}^\infty \|\psi_j\|_{L^\infty(\mathrm{D})}$ is finite.*

Note that $\inf K_*(x) = 0$ is permitted by Assumption 2.1 so that depending on the selection of the parameter z the coefficient $K(x, z)$ is still nonnegative, but could be arbitrarily close to 0.

We assume the parameters z_j to be i.i.d, standard normal random variables, ie. $z_j \sim \mathcal{N}(0, 1)$ in $\mathbb{R}$. To describe the resulting random field, we endow $\mathbb{R}^{\mathbb{N}}$ with the product σ algebra

$$\Theta = \bigotimes_{j=1}^{\infty} \mathcal{B}(\mathbb{R})$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra in $\mathbb{R}$. In the measurable space $(\mathbb{R}^{\mathbb{N}}, \Theta)$, we introduce the gaussian product probability measure (see, eg., [7, Chapter 2] for detailed construction)

$$\gamma = \bigotimes_{j=1}^{\infty} \mathcal{N}(0, 1).$$

For the random diffusion coefficient K in (2.1) to be well-defined, we restrict z in (2.1) to the set

$$\Gamma = \{z = (z_1, z_2, \ldots) \in \mathbb{R}^{\mathbb{N}} : \quad \sum_{j=1}^{\infty} |z_j| b_j < \infty\} .$$

This set is $(\mathbb{R}^{\mathbb{N}}, \Theta)$-measurable and there holds $\gamma(\Gamma) = 1$ (see, e.g., [26, p. 153] or [21, Lemma 2.28]) . For every $z \in \Gamma$, the following quantities are well defined.

$$K_{\max}(z) := \operatorname{esssup}_{x \in D} K_*(x) + \exp\left(\|\bar{K}\|_{L^\infty(D)} + \sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} |z_j|\right), \tag{2.2}$$

and

$$K_{\min}(z) := \operatorname{essinf}_{x\in D} K_*(x) + \exp\left(\operatorname{essinf}_{x\in D} \bar{K}(x) - \sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)}|z_j|\right) . \tag{2.3}$$

Since $\gamma(\Gamma) = 1$, for γ-a.e. draw of $z \in \mathbb{R}^{\mathbb{N}}$, $K_{\min}(z), K_{\max}(z) \in L^\infty(\mathrm{D})$ and $z \mapsto K_{\min}(z)$, $z \mapsto K_{\max}(z)$ are (Γ, Θ)-measurable.

A typical instance of the abstract setting outlined above, has ψ_j the Karhúnen-Loève eigenfunctions of a compact, self-adjoint covariance operator $\mathcal{C} \in \mathcal{L}(L^2(\mathrm{D}))$ with kernel function $k : \mathrm{D} \times \mathrm{D} \to \mathbb{R}$. Then, the realizations of K in (2.1) are γ-a.s. Hölder continuous in $\overline{\mathrm{D}}$. We refer to [8] and the references there for a more detailed discussion.

2.2. **Forward Parabolic Initial Boundary Value Problem.** Let $V = H^1_0(\mathrm{D})$ and $H = L^2(\mathrm{D})$. We denote by $(\cdot,\cdot)$ the duality pairing between V' and V, extended by continuity from the inner product in H.

We assume given deterministic source term $f \in L^2((0,T);V')$ and initial data $g \in H$. For $z \in \Gamma$, we consider the parabolic problem

$$\frac{\partial P}{\partial t} - \nabla\cdot(K(x,z)\nabla P) = f, \quad P(0,x,z) = g. \tag{2.4}$$

We show in the following lemma that P is uniformly bounded in $L^2((0,T);V)$ with respect to each $z \in \Gamma$.

Lemma 2.2. *Under Assumption 2.1, for each $z \in \Gamma$, the solution to* (2.4) *is bounded in $L^2((0,T);V)$:*

$$\|P(z)\|_{L^2((0,T);V)} \le \|f\|_{L^2((0,T);V')} + \frac{\sqrt{\|f\|^2_{L^2((0,T);V')} + 2K_{\min}(z)\|g\|^2_H}}{2K_{\min}(z)}.$$

The proof follows the standard procedure (see, e.g. [25]). We include the argument in Appendix B for completeness. The result in Lemma 2.2 also implies that $\frac{\partial P}{\partial t}$ is a linear functional in the space D; that is, $P(z) \in H^1([0,T];V')$.

To obtain a numerical approximation of the Bayesian posterior measure, conditional on given, noisy observation data δ, we use the truncation of expansion (2.1) and numerically approximate the resulting, finite-parametric forward equation by backward Euler time-discretization and by a standard, Lagrangean FEM in the spatial domain D.

A key point in the design of the multi-level MCMC algorithm and, in particular, of selection of the algorithmic steering parameters which ensure a good error vs. work bounds are a-priori discretization error bounds incurred from the three approximations of the foward problem: parameter truncation to $J \in \mathbb{N}$ many terms, implicit backward Euler timestepping with size $k > 0$, and FE discretization in the spatial domain D with regular, Lagrangean FEM of meshwidth h.

Establishing the asserted error bounds requires regularity for the solution P. In particular, we will need that $P(z) \in H^1((0,T);V)$. For this regularity, it makes sense to identify the pointwise value of P in V for each $t \in (0,T)$. We therefore assume this regularity at the onset. We remark that if this regularity is not satisfied, one cannot identify the pointwise values of P with respect to $t \in (0,T)$. The Bayesian inverse problem could then still be defined with point-values of P

with respect to t in (2.6) being replaced by integrals with respect to the time variable over subintervals on $(0,T)$. The existence and well-posedness of the posterior measure still hold. However, in that case our method of proof does not provide error estimates for the approximations of the posterior measure in terms of the approximating parameters.

Assumption 2.3. *There holds $f\in H^1((0,T);V')$ with $f(0,\cdot)\in H$ and $g\in H^2(\mathrm{D})$. Furthermore, there are constants $C>0$ and $s>1$ such that*

$$K_*,\bar K\in W^{1,\infty}(\mathrm{D}),\qquad \forall j\in\mathbb{N}:\ \|\psi_j\|_{W^{1,\infty}(\mathrm{D})}\le Cj^{-s}.$$

We denote by $q=s-1$. For conciseness, we define by $b_j=\|\psi_j\|_{L^\infty(\mathrm{D})}$ and $\bar b_j=\|\psi_j\|_{W^{1,\infty}(\mathrm{D})}$. We define the prior probability space

$$U=\{z=(z_1,z_2,\ldots)\in\mathbb{R}^{\mathbb{N}}:\ \sum_{j=1}^\infty |z_j|\bar b_j<\infty\}.$$

The set U has γ measure 1. The σ algebra Θ in U is defined as the restriction of the σ algebra $\otimes_{j=1}^\infty\mathcal{B}(\mathbb{R})$ in $\mathbb{R}^{\mathbb{N}}$ to U. The prior probability γ in U is defined as the restriction of the tensorized measure $\otimes_{j=1}^\infty\mathcal{N}(0,1)$ to U. In addition, we have $K(z)\in W^{1,\infty}(\mathrm{D})$ and that

$$\nabla K(z,x)=\nabla K_*(x)+\exp\left(\bar K(x)+\sum_{j=1}^\infty z_j\psi_j(x)\right)\left(\nabla\bar K(x)+\sum_{j=1}^\infty z_j\nabla\psi_j(x)\right).$$

The time derivative of P satisfies the problem

$$\frac{\partial}{\partial t}(\frac{\partial P}{\partial t})-\nabla\cdot(K\nabla\frac{\partial P}{\partial t})=\frac{\partial f}{\partial t},\quad \frac{\partial P}{\partial t}(0)=f(0,\cdot)+\nabla\cdot(K\nabla g). \tag{2.5}$$

As $\frac{\partial f}{\partial t}\in L^2((0,T);V')$ and the initial condition $f(0,\cdot)+\nabla\cdot(K\nabla g)\in H$, problem (2.5) is well-posed. We thus have $P\in H^1((0,T);V)$.

Let $\ell_i\in V'$ for $i=1,\ldots,N$ where $N\in\mathbb{N}$. Let $\tau_i\in(0,T)$ $(i=1,\ldots,N)$. We consider the forward function

$$\mathcal{G}(z)=\big(\ell_1(P(\tau_1,\cdot,z)),\ldots,\ell_N(P(\tau_N,\cdot,z))\big)\in\mathbb{R}^{\mathbb{N}}. \tag{2.6}$$

We consider the noisy observation of $\mathcal{G}$

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

where the additive observation noise ϑ is assumed to be centered, gaussian, i.e. $\vartheta\sim\mathcal{N}(0,\Sigma)$ in $\mathbb{R}^N$; here, the $N\times N$ covariance matrix Σ is positive definite. Our purpose is to find the posterior probability measure $\gamma^\delta=\mathbb{P}(z|\delta)$. Let Φ be the inverse covariance weighted, observation-prediction misfit function

$$\Phi(z;\delta)=\frac12|\delta-\mathcal{G}(z)|_\Sigma^2.$$

3. Existence and Well-posedness of the Posterior measure

We will show that

$$\frac{d\gamma^\delta}{d\gamma}\propto\exp(-\Phi(z;\delta)), \tag{3.1}$$

and therefore that the Bayesian posterior γ^δ is well-posed with respect to δ. To this end, we denote $Z(\delta)=\int_U\exp(-\Phi(z;\delta))d\gamma(z)$ the normalising constant of (3.1). We have the following existence results.

Proposition 3.1. *Under Assumption 2.3, the solution P is measurable when viewed as a map from U to $H^1((0,T);V)$. This implies the existence of the posterior measure γ^δ which is determined by*

$$\frac{d\gamma^\delta}{d\gamma} \propto \exp\left(-\Phi(z;\delta)\right).$$

Proof First we show that P as a map from U to $H^1((0,T);V)$ is measurable. Let $z,z'\in U$. We then have

$$\frac{\partial}{\partial t}(P(z)-P(z')) - \nabla\cdot(K(z)\nabla(P(z)-P(z'))) = \nabla\cdot((K(z)-K(z'))\nabla P(z)$$

with $P(0,\cdot,z)-P(0,\cdot,z')=0$. Thus

$$\begin{aligned}\int_0^T \left(\frac{\partial}{\partial t}(P(z)-P(z')),(P(z)-P(z'))\right)dt& \\ +\int_0^T\int_D K(z)\nabla(P(z)-P(z'))\cdot\nabla(P(z)-P(z'))dxdt& \\ =-\int_0^T\int_D (K(z)-K(z'))\nabla P(z)\cdot\nabla(P(z)-P(z'))dxdt.&\end{aligned}$$

From this we have

$$\begin{aligned}&K_{\min}(z)\|P(z)-P(z')\|^2_{L^2((0,T);V)}\\ &\qquad\le \|K(z)-K(z')\|_{L^\infty(\mathrm{D})}\|P(z)\|_{L^2((0,T);V)}\|\nabla(P(z)-P(z'))\|_{L^2((0,T);V)}\end{aligned}$$

so

$$\|P(z)-P(z')\|_{L^2((0,T);V)} \le \frac{1}{K_{\min}(z)}\|K(z)-K(z')\|_{L^\infty(\mathrm{D})}\|P(z)\|_{L^2((0,T);V)}. \tag{3.2}$$

Similarly, we have

$$\frac{\partial}{\partial t}\frac{\partial}{\partial t}(P(z)-P(z'))-\nabla\cdot(K(z)\nabla\frac{\partial}{\partial t}(P(z)-P(z'))) = \nabla\cdot\left(K(z)-K(z'))\nabla\frac{\partial P(z)}{\partial t}\right).$$

The initial condition is $\frac{\partial}{\partial t}(P(0,\cdot,z)-P(0,\cdot,z'))=\nabla\cdot((K(z)-K(z'))\nabla g)$. Thus

$$\begin{aligned}&K_{\min}(z)\|\frac{\partial}{\partial t}(P(z)-P(z'))\|^2_{L^2((0,T);V)}\\ &\le \|K(z)-K(z')\|_{L^\infty(\mathrm{D})}\|\nabla\frac{\partial P(z)}{\partial t}\|_{L^2((0,T);H)}\|\nabla\frac{\partial}{\partial t}(P(z)-P(z'))\|_{L^2((0,T);H)}\\ &\qquad\qquad + \|\nabla\cdot(K(z)-K(z'))\nabla g\|^2_H.\end{aligned}$$

Therefore

$$\begin{aligned}&\|\frac{\partial}{\partial t}(P(z)-P(z'))\|_{L^2((0,T);V)}\\ &\qquad\le \frac{c}{K_{\min}(z)}\|K(z)-K(z')\|_{L^\infty(\mathrm{D})}\|\nabla\frac{\partial P(z)}{\partial t}\|_{L^2((0,T);H)}\\ &\qquad\qquad + \frac{c}{K_{\min}(z)^{1/2}}\|\nabla\cdot(K(z)-K(z'))\nabla g\|_H.\end{aligned} \tag{3.3}$$

Assume that K only contains a finite number of J terms z_j for $j=1,\dots,J$, i.e. $z\in\mathbb{R}^J$. Then equations (3.2) and (3.3) show that P as a map $\mathbb{R}^J\ni z\to H^1((0,T);V)$ is continuous. Thus as a map from $\mathbb{R}^J\ni z\to C([0,T];V)$ it is also continuous, and is measurable. Let $X \in \mathcal{B}(C([0,T];V))$. The preimage of $P^{-1}(X)$ in $\mathbb{R}^J$ belongs to $\mathcal{B}(\mathbb{R}^J)$.

For $z \in U$, we consider the J-term truncated coefficient

$$K^J(x,z) := K_*(x) + \exp(\bar{K}(x) + \sum_{j=1}^{J} z_j \psi_j(x)).$$

We consider the truncated equation

$$\frac{\partial P^J}{\partial t}(z) - \nabla \cdot (K^J(z)\nabla P^J(z)) = f,\ \ P^J(0,\cdot,z) = g. \tag{3.4}$$

Consider P^J as a map $U \ni z \to H^1((0,T);V)$, we then deduce that this map is measurable. From Lemma B.2, we deduce that for $z \in U$,

$$\lim_{J\to\infty} \|P^J(z) - P(z)\|_{H^1((0,T);V)} = 0$$

so

$$\lim_{J\to\infty} \|P^J(z) - P(z)\|_{C([0,T];V)} = 0.$$

As P is a pointwise limit of a sequence of measurable maps, $P(z)$ as a map $U \ni z \to C([0,T];V)$ is measurable. Therefore $\mathcal{G} : U \ni z \to \mathbb{R}^N$ is measurable. Thus (3.1) holds (see Theorem 2.1 of [9]). □

Next, we show the well-posedness.

Proposition 3.2. *The posterior probability measure γ^δ is determined from formula* (3.1)*. It is well-posed and locally Lipschitz continuous with respect to the observation data. In particular, for observation $\delta, \delta' \in \mathbb{R}^N$ such that $|\delta|, |\delta'| < r$, there is a constant $c(r)$ such that*

$$d_{Hell}(\gamma^\delta, \gamma^{\delta'}) < c(r)|\delta - \delta'|. \tag{3.5}$$

Proof We show inequality (3.5). We note that

$$\begin{aligned}
&2d_{Hell}(\gamma^\delta,\gamma^{\delta'})^2\\
&= \int_U \left(\sqrt{\frac{d\gamma^\delta}{d\gamma}} - \sqrt{\frac{d\gamma^{\delta'}}{d\gamma}}\right)^2 d\gamma(z)\\
&= \int_U \left[Z(\delta)^{-\frac12}\exp\left(-\frac12\Phi(z;\delta)\right) - Z(\delta')^{-\frac12}\exp\left(-\frac12\Phi(z,\delta')\right)\right]^2 d\gamma(z)\\
&\le 2\int_U \left(Z(\delta)^{-\frac12}\exp\left(-\frac12\Phi(z;\delta)\right) - Z(\delta)^{-\frac12}\exp\left(-\frac12\Phi(z,\delta')\right)\right)^2 d\gamma(z)\\
&\ \ + 2\int_U \left(Z(\delta)^{-\frac12}\exp\left(-\frac12\Phi(z,\delta')\right) - Z(\delta')^{-\frac12}\exp\left(-\frac12\Phi(z,\delta')\right)\right)^2 d\gamma(z)\\
&= I_1 + I_2
\end{aligned} \tag{3.6}$$

where

$$I_1 = \frac{2}{Z(\delta)}\int_U \left(\exp(-\frac12\Phi(z;\delta)) - \exp(-\frac12\Phi(z,\delta'))\right)^2 d\gamma(z), \tag{3.7}$$

and

$$I_2 = 2|Z(\delta)^{-1/2} - Z(\delta')^{-1/2}|^2\int_U \exp(-\Phi(z,\delta'))d\gamma(z). \tag{3.8}$$

We have that

$$\left|\exp(-\frac{1}{2}\Phi(z;\delta)) - \exp(-\frac{1}{2}\Phi(z,\delta'))\right| \le c(|\delta| + |\delta'| + |\mathcal{G}(z)|)|\delta - \delta'|.$$

From Lemma B.1 in Section B, we have that

$$\|P(\cdot,\cdot,z)\|_{C([0,T];V)} \le c\exp(\sum_{j=1}^{\infty}|z_j|(b_j + \bar{b}_j)), \tag{3.9}$$

so

$$|\mathcal{G}(z)| \le c\exp(\sum_{j=1}^{\infty}|z_j|(b_j + \bar{b}_j)).$$

Thus

$$\left|\exp(-\frac{1}{2}\Phi(z;\delta)) - \exp(-\frac{1}{2}\Phi(z,\delta'))\right| \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j + \bar{b}_j))|\delta - \delta'|.$$

Therefore

$$I_1 \le c|\delta - \delta'|^2 \int_U \exp(c\sum_{j=1}^{\infty}|z_j|(b_j + \bar{b}_j))d\gamma(z) \le c|\delta - \delta'|^2$$

due to Lemma A.1. Similarly, there exists a constant $c > 0$ (depending on Γ, N, r) such that

$$\forall \delta, \delta' \in B_r(0): \quad I_2 \le c|\delta - \delta'|^2.$$

$\square$

4. Numerical Approximation

For each choice of $z \in U$, we consider the numerical approximation of the parametric parabolic problem (2.4). It involves three approximations:

(i) truncating the KL expansion (2.1) to a finite number $J \in \mathbb{N}$ of terms,
(ii) discretizing the time derivative by implicit, backward-Euler time-stepping,
(iii) discretizing the resulting sequence of elliptic boundary value problems by a standard, continuous piecewise affine Lagrangean FEM in the spatial domain D.

We now estimate the impact of each of these approximations on the overall accuracy of the forward map.

4.1. *J*-term truncation of affine-parametric input.

We approximate the parametric diffusion coefficient in (2.4) by a finite number J of terms in (2.1), i.e. for $J \in \mathbb{N}$ we define

$$K^J(x,z) = K_*(x) + \exp\left(\bar{K}(x) + \sum_{j=1}^{J} z_j\psi_j(x)\right), \quad x \in \mathrm{D},\ z \in \mathbb{R}^J .$$

We consider the corresponding J-term truncated forward equation

$$\frac{\partial P^J}{\partial t}(z) - \nabla\cdot(K^J(z)\nabla P^J(z)) = f, \quad P^J(0,\cdot,z) = g. \tag{4.1}$$

Under Assumption 2.3, the equation for $\frac{\partial P^J}{\partial t}$ reads formally

$$\frac{\partial}{\partial t}\frac{\partial P^J}{\partial t}(z) - \nabla\cdot(K^J(z)\nabla\frac{\partial P^J}{\partial t}(z)) = \frac{\partial f}{\partial t} \;\text{ with }\; \frac{\partial P^J}{\partial t}(0,\cdot,z) = f(0,\cdot) + \nabla\cdot(K^J(\cdot,z)\nabla g) .$$

This initial-boundary value problem is well posed due to our assumptions $\frac{\partial f}{\partial t} \in L^2((0,T);V')$ and $f(0,\cdot) + \nabla\cdot(K^J\nabla g) \in H$. Thus, for every $J \in \mathbb{N}$, there exists a unique solution $P^J \in H^1((0,T);V)$ of (4.1). For a given parameter realization $z \in \mathbb{R}^J$, and for $J \in \mathbb{N}$, we consider the corresponding approximate forward map

$$\mathcal{G}^J(z) = \big(\ell_1(P^J(\tau_1,\cdot,z)),\ldots,\ell_N(P^J(\tau_N,\cdot,z))\big) \in \mathbb{R}^N.$$

The corresponding data-to-observation misfit functional is

$$\Phi^J(z;\delta) = \frac{1}{2}|\delta - \mathcal{G}^J(z)|_\Sigma^2. \tag{4.2}$$

For given data δ and for the J-truncated forward problem, we define the corresponding Bayesian posterior measure

$$\frac{d\gamma^{J,\delta}}{d\gamma} \propto \exp(-\Phi^J(z;\delta)), \tag{4.3}$$

with the normalizing constant $Z^{J,\delta} = \int_{\mathbb{R}^J} \exp(-\Phi^J(z;\delta)) > 0$.

We have the following result on well-posedness of the BIP corresponding to J-term truncated, parametric inputs.

Proposition 4.1. *There exists a constant $c(r) > 0$ such that, for every $J \in \mathbb{N}$ and for every data $\delta \in B_r(0)$, the approximated posterior measure $\gamma^{J,\delta}$ satisfies, with $q = s-1$,*

$$d_{Hell}(\gamma^\delta, \gamma^{J,\delta}) \le cJ^{-q}.$$

We provide the proof of this theorem in Appendix B.

4.2. **Space- and Time-discretization.** We next consider the impact of backward Euler time discretization and of the P_1-Lagrangean Finite Element Method for discretizing the truncated equation (4.1) in the spatial domain D.

We consider a partition of $[0,T]$ by $0 = t_0 < t_1 < t_2 < \ldots < t_M = T$ such that $t_j - t_{j-1} = k$ for all $j = 1,\ldots,M$. We consider in D a nested sequence $\{\mathcal{T}^l\}_{l=0}^\infty$ of regular, simplicial triangulations of D; each simplex $T \in \mathcal{T}^l$ is obtained by dividing each simplex in $\mathcal{T}^{l-1}$ into 2^d congruent subsimplices, ie., in 4 triangles when $d = 2$ or into 8 tedrahedra when $d = 3$. The sequence $\{\mathcal{T}^l\}_{l=0}^\infty$ of triangulations of D generated in this way is uniformly shape-regular. We define a nested sequence $\{V^l\}_{l\ge1}$ of finite-dimensional spaces of continuous, piecewise affine Lagrangean finite element functions

$$V^l = \{w \in V : \ w|_T \in \mathbb{P}^1(T)\ \forall\, T \in \mathcal{T}^l\}, \tag{4.4}$$

where $\mathbb{P}^1(T)$ is the set of linear polynomials in T. The mesh size of each $V^l \subset H_0^1(\mathrm{D})$ is $h_l = O(2^{-l})$. For notational simplicity, we denote h_l by h in the presentation below and in the proof in Appendix C.

For $z \in U$, we consider the backward Euler finite element scheme: Find $P_m^{J,h,k}(z) \in V^l$ for $m = 1,\ldots,M$ such that

$$\int_{\mathrm{D}} \frac{P_m^{J,h,k}(x,z) - P_{m-1}^{J,h,k}(x,z)}{k}\phi dx + \int_{\mathrm{D}} K^J(x,z)\nabla P_m^{J,h,k}(x,z)\cdot\nabla\phi(x)dx = \int_{\mathrm{D}} f(t_m,x)\phi(x)dx, \tag{4.5}$$

$\forall \phi \in V^l$ with $P_0^{J,h,k} = g^l$ where g^l is an approximation of g in V^l. As $g \in H^2(\mathrm{D})$, we choose g^l so that $\|g - g^l\|_V \le ch$. For simplicity, we assume that $\tau_i \in \{t_1, t_2, \dots, t_M\}$. If $\tau_i = t_m$, we denote $P_m^{J,h,k}$ by $P_{\tau_i}^{J,h,k}$. We define

$$\mathcal{G}^{J,h,k}(z) = \big(\ell_1(P_{\tau_1}^{J,h,k}(z)), \dots, \ell_N(P_{\tau_N}^{J,h,k}(z))\big).$$

Let

$$\Phi^{J,h,k}(z;\delta) = \frac{1}{2}|\delta - \mathcal{G}^{J,h,k}(z)|_\Sigma^2.$$

We define the approximate Bayesian posterior measure resulting from the discretization of the forward model, including J-term truncation of the coefficient expansion (2.1), by its density

$$\frac{d\gamma^{J,h,k,\delta}}{d\gamma} \propto \exp(-\Phi^{J,h,k}(z;\delta)). \tag{4.6}$$

The following result provides a bound on the difference between the true Bayesian posterior and its approximation resulting from the discretization of the forward problem.

Proposition 4.2. *Assume that* D *is a bounded, convex polytope, and that* $f \in H^1((0,T);L^2(\mathrm{D}))$ *with* $f(0,\cdot) \in V$ *and* $g \in H_0^3(D)$. *Under Assumption 2.3, there exists a constant* $c > 0$ *such that for all* $J \in \mathbb{N}$, *and all* $h, k > 0$ *there holds*

$$d_{Hell}(\gamma^\delta, \gamma^{J,h,k,\delta}) \le c(J^{-q} + h + k).$$

We prove this proposition in Appendix C. This shows the corresponding error for approximating the posterior probability measure by truncating the coefficient and solving the truncated equation using Backward Euler timestepping in $[0,T]$ and P_1-FEM in the domain D. The effect of this approximation is observed numerically when we approximate the expectation with respect to the Bayesian posterior probability measure of a quantity of interest by the expectation with respect to the approximate posterior probability measure, obtained via the space and time-discretized forward problem. We remark that MLMC algorithms in forward UQ for parabolic stochastic PDEs, this truncation is not required; see, e.g., [3].

5. Multilevel MCMC

5.1. **Definition of the MLMCMC Algorithm.** Let $\ell \in V'$. To develop the multilevel MCMC method for approximating the posterior expectation of the quantity of interest $\ell(P(T,\cdot,z))$, for the FE mesh $h_l = O(2^{-l})$ we choose J and k so that the terms in the right hand side of the estimate in Proposition 4.2 are equivalent. In particular, we let $k = T2^{-l}$ and $J = \lceil 2^{l/q} \rceil$. We then denote the measure $\gamma^{J,h_l,k,\delta}$ by γ^l, solution $P_m^{J,h_l,k}$ by P_m^l and potential $\Phi^{J,h_l,k}$ by Φ^l. We denote $P_M^{J,h_l,k}$ which approximates $P(T,\cdot,z)$ by P_T^l on discretization level l. The multilevel MCMC method for the parabolic equation (2.4) is similar to that for elliptic problems in [15]. Relying on the detailed derivation in that reference, we shall only present the result.

To this end, we recall the definition in [15]

$$
\begin{aligned}
A_1^{ll'} &= (1-\exp(\Phi^l(z;\delta)-\Phi^{l-1}(z;\delta))Q^{l'}(z)I^l(z),\\
A_2^{ll'} &= (\exp(\Phi^{l-1}(z;\delta)-\Phi^l(z;\delta))-1)Q^{l'}(z)(1-I^l(z)),\\
A_3^{l} &= (\exp(\Phi^l(z;\delta)-\Phi^{l-1}(z;\delta))-1)I^l(z),\\
A_4^{ll'} &= Q^{l'}(z)I^l(z),\\
A_5^{l} &= (1-\exp(\Phi^{l-1}(z;\delta)-\Phi^l(z;\delta)))(1-I^l(z)),\\
A_6^{ll'} &= \exp(\Phi^l(z;\delta)-\Phi^{l-1}(z;\delta))Q^{l'}(z)I^l(z),\\
A_7^{ll'} &= Q^{l'}(z)(1-I^l(z)),\\
A_8^{ll'} &= \exp(\Phi^{l-1}(z;\delta)-\Phi^l(z;\delta))Q^{l'}(z)(1-I^l(z)),
\end{aligned}
$$

where $0\le l,l'\le L$ and

$$
\begin{aligned}
&Q^{l'} := \ell(P_T^{l'})-\ell(P_T^{l'-1}) \quad \text{when } l'\neq 0 \text{ and}\\
&Q^{l'} := \ell(P_T^{0}) \quad \text{when } l'=0;
\end{aligned}
$$

and

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

With this notation at hand, the multilevel MCMC approximation of the posterior expectation of a quantity of interest $\ell(P(T,\cdot,z))$, i.e., of $\mathbb{E}^{\gamma^\delta}[\ell(P(T,\cdot,z))]$ where $\ell\in V'$, is defined by

$$
\begin{aligned}
&E_L^{MLMCMC}[\ell(P(T,\cdot,z))]\\
&=\sum_{l=1}^{L}\sum_{l'=1}^{L'(l)}\Big[E^{\gamma^l}_{M_{ll'}}[A_1^{ll'}]+E^{\gamma^{l-1}}_{M_{ll'}}[A_2^{ll'}]+E^{\gamma^l}_{M_{ll'}}[A_3^{l}]\cdot E^{\gamma^{l-1}}_{M_{ll'}}[A_4^{ll'}+A_8^{ll'}]\\
&\qquad\qquad +E^{\gamma^{l-1}}_{M_{ll'}}[A_5^{l}]\cdot E^{\gamma^l}_{M_{ll'}}[A_6^{ll'}+A_7^{ll'}]\Big]\\
&+\sum_{l=1}^{L}\Big[E^{\gamma^l}_{M_{l0}}[A_1^{l0}]+E^{\gamma^{l-1}}_{M_{l0}}[A_2^{l0}]+E^{\gamma^l}_{M_{l0}}[A_3^{l}]\cdot E^{\gamma^{l-1}}_{M_{l0}}[A_4^{l0}+A_8^{l0}]\\
&\qquad\qquad +E^{\gamma^{l-1}}_{M_{l0}}[A_5^{l}]\cdot E^{\gamma^l}_{M_{l0}}[A_6^{l0}+A_7^{l0}]\Big]\\
&+\sum_{l'=1}^{L'(0)}E^{\gamma^0}_{M_{0l'}}[\ell(P_T^{l'}-P_T^{l'-1})]+E^{\gamma^0}_{M_{00}}[\ell(P_T^0)]\,.
\end{aligned}
$$

Here $E^{\gamma^l}_{M_{ll'}}$ is the approximation of $\mathbb{E}^{\gamma^l}$ using MCMC with $M_{ll'}$ samples with the acceptance probability

$$
\alpha^l(z,s)=1\wedge\exp(\Phi^l(z;\delta)-\Phi^l(s,\delta)). \tag{5.1}
$$

5.2. **Analysis of the MLMCMC Algorithm.** To analyze the MLMCMC algorithm, we work under the following *assumption of geometric ergodicity*. To state this assumption, we introduce some auxiliary quantities. From (3.9), there are positive constants c_1 and c_2 such that for all $z\in U$

$$
\Phi^l(z)\le c_1+c_2\exp\left(2\sum_{j=1}^{\infty}(b_j+\bar b_j)|z_j|\right).
$$

We define by

$$\kappa = \int_U \exp\left(-c_2 \exp(2\sum_{j=1}^{\infty}(b_j + \bar{b}_j)|z_j|)\right) d\gamma(z).$$

As shown in [16], κ is strictly positive. Following [16], we define the probability measure $\bar{\gamma}$ as

$$d\bar{\gamma}(z) = \frac{1}{\kappa} \exp\left(-c_2 \exp(2\sum_{j=1}^{\infty}(b_j + \bar{b}_j)|z_j|)\right) d\gamma(z), \quad z \in U\ . \tag{5.2}$$

Let $\mathcal{E}^{\bar{\gamma},l}$ denote the expectation with respect to the probability space generated by the MCMC algorithm at discretization level l, with the acceptance probability defined in (5.1), and with the initial sample $z^{(0)}$ distributed according to the probability measure $\bar{\gamma}$. To prove the convergence of the MLMCMC sampling, we work under the following *assumption of geometric ergodicity*. As in our work on the elliptic problem in [15], we assume

Assumption 5.1. *[Geometric Ergodicity] Let $C > 0$ be sufficiently large. For each l and l' in $\mathbb{N}$, denote by*

$$\mathcal{V}^{ll'}(z) = \exp\left(C\sum_{j=1}^{\infty}(b_j + \bar{b}_j)|z_j| + \frac{1}{\varepsilon}\sum_{j>J_{l-1}}(b_j + \bar{b}_j)|z_j| + \frac{1}{\varepsilon'}\sum_{j'>J_{l'-1}}(b_{j'} + \bar{b}_{j'})|z_{j'}|\right) \tag{5.3}$$

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

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

Remark 5.2. *Assume that the expansion in* (2.1) *only has a finite number J of random variables z_j, i.e.*

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

With a sufficiently large constant $C > 0$, we can choose $\mathcal{V}^{ll'}$ as

$$\mathcal{V}^{ll'}(z) = \exp\left(C\sum_{j=1}^{J} b_j|z_j|\right). \tag{5.4}$$

Remark 5.3. *The proof of Assumption 5.1 for the independence sampler follows exactly the same lines of* [15] *Appendix B. For the gPC sampler, if we assume the validity of the $L^2_{\gamma^\delta}$ spectral gap result of* [14]*, then Assumption 5.1 holds.*

For each discretization level $l \in \mathbb{N}_0$, we introduce the Markov chains

$$\mathcal{C}_l = \{z^{(k)}\}_{k\in\mathbb{N}_0} \subset \mathbb{R}^{J_l}$$

which are started with $z^{(0)} \in \mathbb{R}^{J_l}$ and then generated by the MCMC process with the acceptance probability α^l in (5.1) with $J_l = \lceil 2^{l/q} \rceil$. We denote by $\mathbf{E}_L$ the expectation with respect to the probability space generated by these Markov chains for $l = 1, \dots, L$. We have the following convergence result.

Theorem 5.4. *Let* D *be a bounded polytope in* $\mathbb{R}^d$ *with Lipschitz boundary. Under Assumption 2.3, the hypothesis of Proposition 4.2 and under the geometric ergodicity Assumption 5.1, in space dimension* $d = 2, 3$, *with the choices*

$$L'(l) = L - l, \quad M_{ll'} = 2^{2(L-(l+l'))} \ \text{for } l \geq 1,\ l' \geq 1,$$
$$M_{0l} = M_{l0} = 2^{2L}/L^2,\ M_{00} = 2^{2L}/L^4,$$

there exists a constant $C(\ell, \delta) > 0$ *such that for every* $L \in \mathbb{N}$

$$\mathbf{E}_L\left[|\mathbb{E}^{\gamma^\delta}[\ell(P(T,\cdot,z))] - E_L^{MLMCMC}[\ell(P(T,\cdot,z))]|\right] \leq C(\ell,\delta) L^2 2^{-L} \ . \tag{5.5}$$

The proof of this theorem is along the lines of the argument in the elliptic case which we detailed in [15]. Similar to what we found there, the logarithmic factor L^2 in the estimate of Theorem 5.4 can be reduced by slightly increasing the sample size of each Markov chain with the choice

$$M_{ll'} = (l + l')^a 2^{2(L-(l+l'))}, \quad l \geq 1 \ \text{ and } l' \geq 1 \ .$$

We list the resulting asymptotic bounds in Table 1 below.

a	$M_{ll'}, l, l' > 1$	$M_{l0} = M_{0l}$	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	$(l+l')^3 2^{2(L-(l+l'))}$	$l 2^{2(L-l)}$	$2^{2L}/L$	$O(L^{1/2} 2^{-L})$
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. Relation of total error with different a

6. NUMERICAL EXPERIMENTS

We present the numerical experiments to verify the theoretical results for MLMCMC for parabolic equations with log-normal coefficients. We consider periodic boundary condition as this allows for establishing a very accurate reference solution to compare the MLMCMC results to. We consider both the independent sampler and the preconditioned Crank-Nicolson (pCN) sampler (see [22]) for the MCMC. We used Gauss Hermite quadrature to estimate the posterior expectation. At each of the quadrature point, the forward parabolic equation is solved using Fourier collocation and a highly accurate implicit Runge Kutta method.

6.1. **Independence Sampler.** We present the MLMCMC using the independence sampler first. We consider the parametric forward equation

$$\frac{\partial P}{\partial t}(t,x,z) - \nabla \cdot (K(x,z)\nabla P(t,x,z)) = f(t,x) \quad \text{for } x \in \mathrm{D} = (0,1)\times(0,1), t \in (0,1) \tag{6.1}$$

where $K(x,z) = \exp(z(\sin(2\pi x_1) + \sin(2\pi x_2)))$ and $f(x,t) = 200(\sin(2\pi x_1) + \sin(2\pi x_2))$, with the periodic boundary condition and initial condition $P(x,0) = 0$; here $x = (x_1, x_2) \in \mathrm{D}$. The observation functional is chosen to be

$$\mathcal{G}(z) = \int_D x_1 \frac{\partial P}{\partial x_1}(1,\cdot,z) + x_2 \frac{\partial P}{\partial x_2}(1,\cdot,z) dx \tag{6.2}$$

and the quantity of interest as

$$\ell(P(z)) = \int_D x_1^{1.5} \frac{\partial P}{\partial x_1}(1,\cdot,z) + x_2^{1.5} \frac{\partial P}{\partial x_2}(1,\cdot,z) dx. \tag{6.3}$$

To compute the reference solution, 1200 Gauss-Hermite quadrature points are used to estimate the expectation integral, 25×25 collocation points are used in Fourier Collocation and an implicit Runge Kutta method of order 12 with time step $\frac{1}{32}$ is used to solve the forward parabolic equation.

In Figure 1, we present the numerical result for a in Table 1 being 0. The error is computed as the average error of 64 runs of MLMCMC. We plot the error versus the finest mesh level h_L. The gradient of the best fit straight line is 0.8776. Similarly in Figures 2 and 3, we present the MLMCMC error for $a = 2$ and 4. The gradient of the best fit straight lines are 0.9306 and 0.9840 respectively. These results illustrates the theory. The slight difference in the slope for $a = 0$ compare to $a = 2, 4$ appears indicative of the significance of the logarithmic factor in the convergence error shown in Table 1.

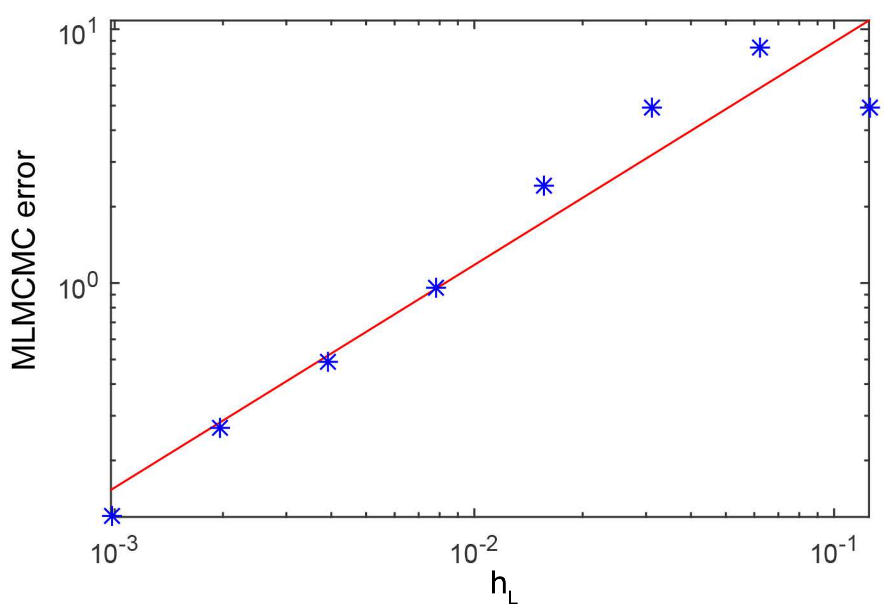

FIGURE 1. MLMCMC error for 2D parabolic equation with Gaussian prior with independence sampler, $a = 0$

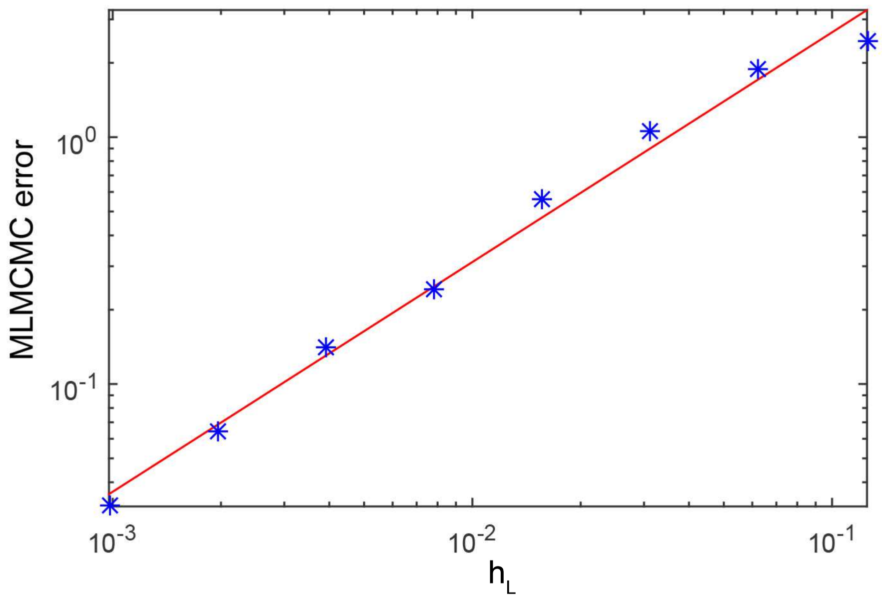

FIGURE 2. MLMCMC error for 2D parabolic equation with Gaussian prior with independence sampler, $a = 2$

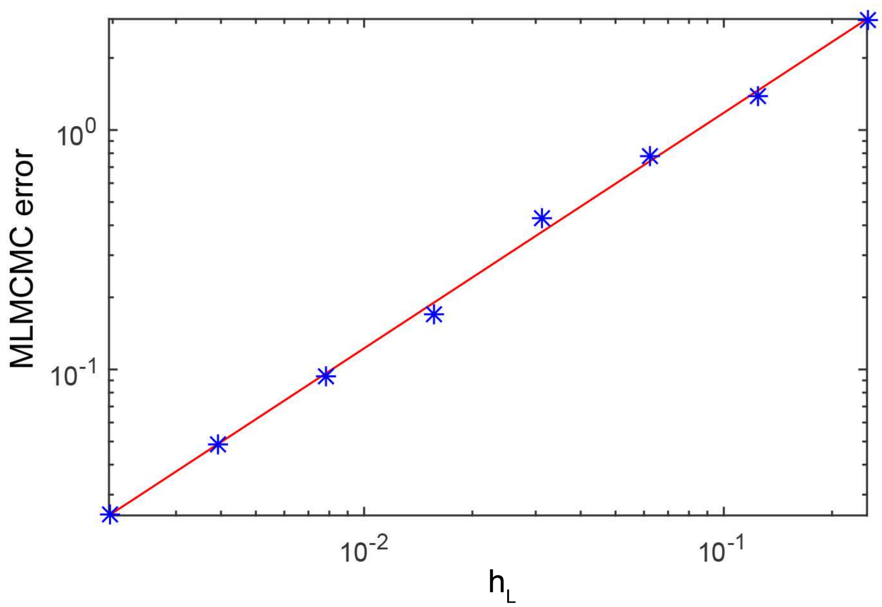

FIGURE 3. MLMCMC error for 2D parabolic equation with Gaussian prior with independence sampler, $a = 4$

6.2. **pCN Sampler.** We now present the results for the MLMCMC method with the pCN sampler (see [22]) where a proposal is chosen from the current MCMC state $z^{(k)}$ as:

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

where ξ is the standard normal variable. We consider several choices for β. We present the results for $\beta = \frac{1}{\sqrt{2}}$ first. In Figures 4 and 5, we plot the MLMCMC error versus the finest mesh level h_L for $a = 0$ and 2 respectively. The slopes of the best fit straight lines are 0.7947 and 1.0185.

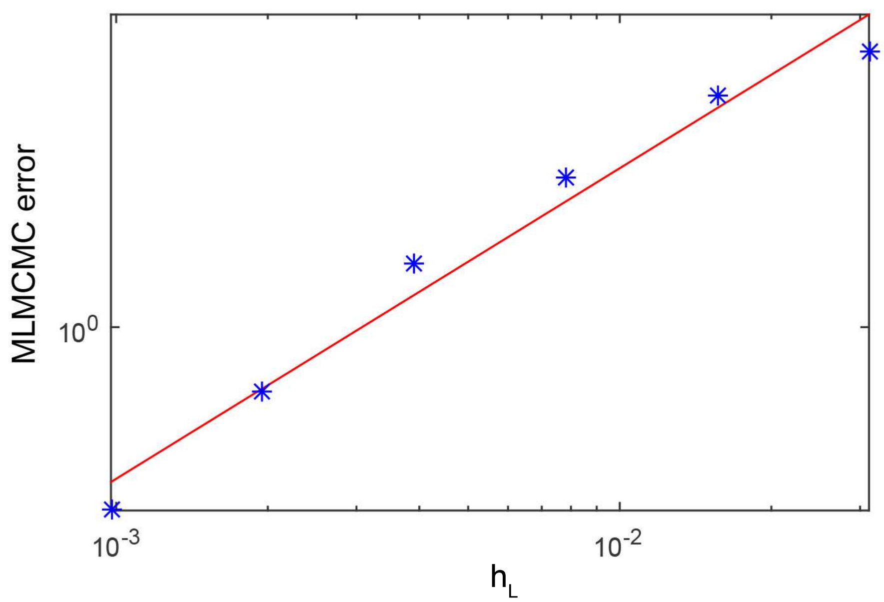

FIGURE 4. MLMCMC error for 2D parabolic equation with Gaussian prior with pCN sampler, $a = 0, \beta = \frac{1}{\sqrt{2}}$

Next, we present the results for $\beta = \frac{1}{\sqrt{10}}$. In Figures 6 and 7, we plot the MLMCMC error against finest mesh level h_L for $a = 0$ and 2 respectively. The gradient of the best fit straight lines are 0.6109 and 0.9285. In both cases, we observe that the choice of $a = 0$ results in inferior algorithm performance, in terms of error vs. accuracy as compared to the choice of $a = 2$. This appears indicative of the logarithmic factor in the error bound of the MLMCMC.

7. Conclusions

We considered first order time and space discretization; the error bounds are essentially best possible for the given regularity of the solution, and also for the data: due to the time independence of the diffusion coefficient K, one could expect e.g. high time-regularity due to "parabolic smoothing". However, due to the nonuniform ellipticity of the spatial operator, which is a consequence of our lack of lower-boundedness away from zero, it seems not obvious to establish a corresponding

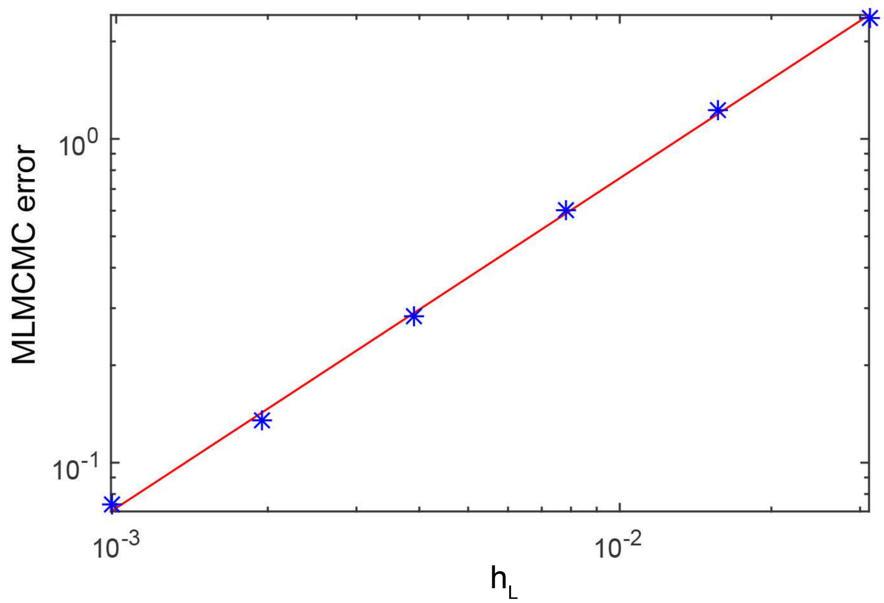

FIGURE 5. MLMCMC error for 2D parabolic equation with Gaussian prior with pCN sampler, $a = 2, \beta = \frac{1}{\sqrt{2}}$

regularity theory where constants are explicit w.r. to the coefficient parameters $z \in U$.

We imposed sufficient conditions on the data (domain, f and g) in order to allow for almost sure $H^2(\mathrm{D})$ regularity of the solutions; this in turn, allowed using standard discretization error bounds based on quasiuniform spatial and temporal stepsizes in the discretization of the forward model. We hasten to add that we expect the steering parameter choices for the MLMCMC algorithm which were derived in the present ms. will remain valid also for discretizations that allow local mesh refinement, e.g. for domains D with re-entrant corners, or mixed boundary conditions.

Finally, a corresponding multilevel error analysis for approximating posterior expectations of quantities of interest for parabolic equations with log-normal coefficients could be equally developed for other sampling methods such as Hamiltonian Monte Carlo (HMC) and Sequential Monte Carlo (SMC), or for their variants such as the "MLS2MC" algorithm recently proposed in [18].

APPENDIX A

We record the following estimates whose proofs can be found in [16].

Lemma A.1. *There exists a constant $c > 0$ such that for every $s > 0$ there holds the estimates*

$$\int_{-\infty}^{\infty} \exp(-t^2/2 + |t|s)\frac{dt}{\sqrt{2\pi}} \leq c\exp(s^2/2)\exp(s\sqrt{2/\pi}),$$

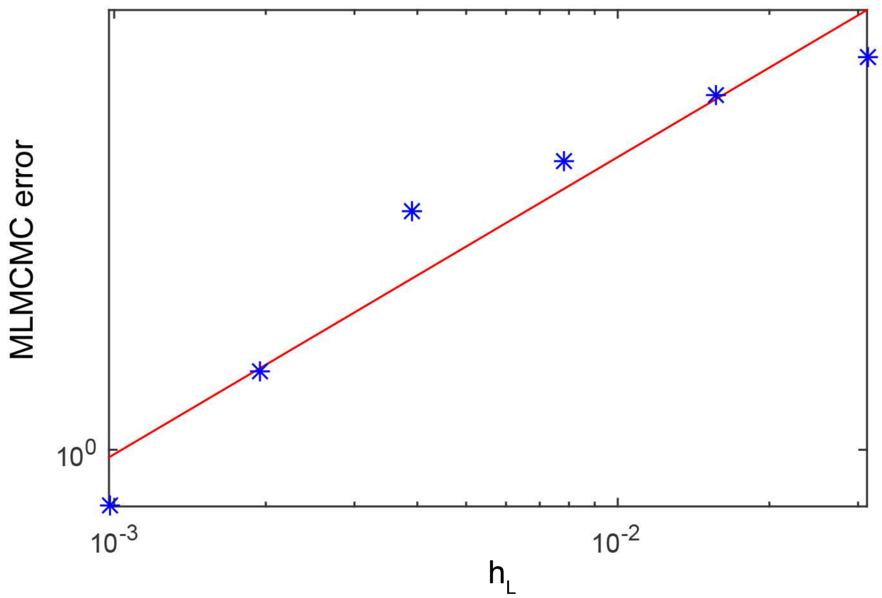

FIGURE 6. MLMCMC error for 2D parabolic equation with Gaussian prior and pCN sampler, $a = 0, \beta = \frac{1}{\sqrt{10}}$

$$\int_{-\infty}^{\infty} t^2 \exp(-t^2/2 + |t|s) \frac{dt}{\sqrt{2\pi}} \leq c \exp(s^2/2)(1 + s^2),$$

and

$$\int_{-\infty}^{\infty} |t| \exp(-t^2/2 + |t|s) \frac{dt}{\sqrt{2\pi}} \leq c \exp(s^2/2)(1 + s) \ .$$

APPENDIX B. PROOF OF PROPOSITION 4.1

We start with a stability bound of the solutions of the exact and of the J-term truncated, parametric parabolic problem.

Lemma B.1. *Under Assumption 2.3, there is a constant $c > 0$ such that*

$$\forall z \in U: \quad \|P(\cdot, \cdot, z)\|_{H^1((0,T);V)} \leq c \exp\left(\sum_{j=1}^{\infty} |z_j|(b_j + \bar{b}_j)\right).$$

Furthermore, for every $J \in \mathbb{N}$ holds the uniform stability bound

$$\forall z \in U: \quad \|P^J(\cdot, \cdot, z)\|_{H^1((0,T);V)} \leq c \exp\left(\sum_{j=1}^{\infty} |z_j|(b_j + \bar{b}_j)\right).$$

Proof From (2.4) we have

$$\int_0^T \left(\frac{\partial P}{\partial t}(z), P(z)\right) dt + \int_0^T \int_D K(z) \nabla P(z) \cdot \nabla P(z) dx dt = \int_0^T \int_D f P(z) dx dt.$$

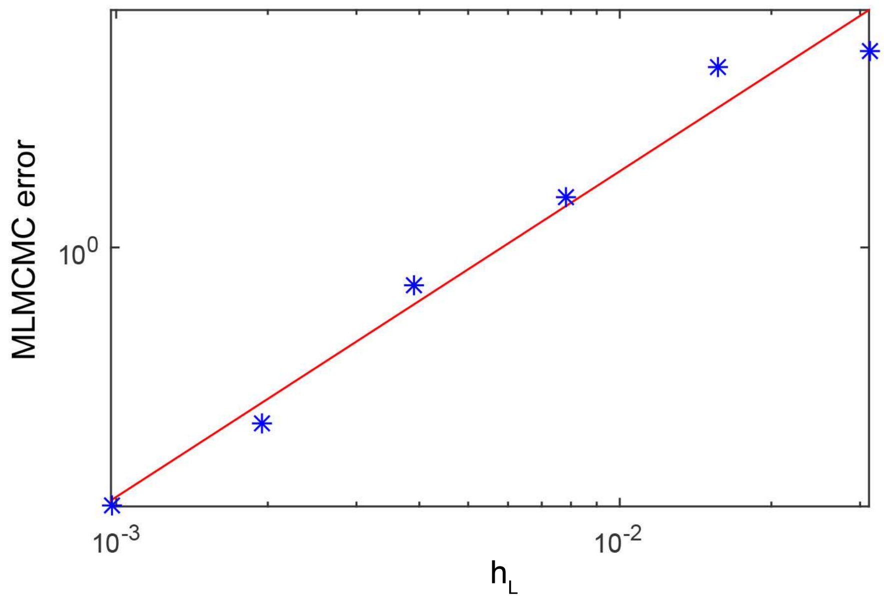

FIGURE 7. MLMCMC error for 2D parabolic equation with Gaussian prior with pCN sampler, $a = 2, \beta = \frac{1}{\sqrt{10}}$

Thus

$$\frac{1}{2}\|P(T,\cdot,z)\|_H^2 + K_{\min}(z)\|P(z)\|^2_{L^2((0,T);V)}$$
$$\leq \frac{1}{2K_{\min}(z)}\|f\|^2_{L^2((0,T);V')} + \frac{K_{\min}(z)}{2}\|P(z)\|^2_{L^2((0,T);V)} + \frac{1}{2}\|P(0,\cdot,z)\|_H^2.$$

This implies

$$\frac{K_{\min}(z)}{2}\|P(z)\|^2_{L^2((0,T);V)} \leq \frac{1}{2K_{\min}(z)}\|f\|^2_{L^2((0,T);V')} + \frac{1}{2}\|g\|_H^2$$

i.e.

$$\|P(z)\|^2_{L^2((0,T);V)} \leq \frac{1}{K^2_{\min}(z)}\|f\|^2_{L^2((0,T);V')} + \frac{1}{K_{\min}(z)}\|g\|_H^2.$$

Therefore, there is a constant $c > 0$ such that

$$\|P(z)\|_{L^2((0,T);V)} \leq c\left(\frac{1}{K_{\min}(z)}\|f\|_{L^2((0,T);V')} + \frac{1}{K_{\min}(z)^{1/2}}\|g\|_H\right).$$

Similarly, from (2.5), there is a constant $c > 0$ such that

$$\left\|\frac{\partial P}{\partial t}(z)\right\|_{L^2((0,T);V)}$$
$$\leq c\left(\frac{1}{K_{\min}(z)}\left\|\frac{\partial f}{\partial t}\right\|_{L^2((0,T);V')} + \frac{1}{K_{\min}(z)^{1/2}}(\|f(0,\cdot)\|_H + \|\nabla\cdot(K(z)\nabla g)\|_H)\right).$$

We note that

$$\nabla\cdot(K(z)\nabla g)=\nabla K(z)\cdot\nabla g+K(z)\Delta g.$$

From

$$\nabla K(z)=\nabla K_*+\exp(\bar K(x)+\sum_{j=1}^\infty z_j\psi_j)(\nabla\bar K(x)+\sum_{j=1}^\infty z_j\nabla\psi_j),$$

we have

$$\begin{aligned}\|\nabla K(z)\|_{L^\infty(D)}&\le\|\nabla K_*\|_{L^\infty(D)}+c\exp(\sum_{j=1}^\infty|z_j|b_j)(1+\sum_{j=1}^\infty z_j\bar b_j)\\&\le c(1+\exp(\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)).\end{aligned}$$

Thus

$$\|P(z)\|_{H^1((0,T);V)}\le c(1+c\exp(\sum_{j=1}^\infty|z_j|(b_j+\bar b_j))).$$

The proof for $P^J(z)$ is similar. □

Lemma B.2. *Under Assumption 2.3, there is a constant $c>0$ such that*

$$\|P(z)-P^J(z)\|_{H^1((0,T);V)}\le c\left(\sum_{j>J}|z_j|(b_j+\bar b_j)\right)\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right).$$

Proof From (2.4)

$$\frac{\partial}{\partial t}(P(z)-P^J(z))-\nabla\cdot(K(z)\nabla(P(z)-P^J(z)))=\nabla\cdot(K(z)-K^J(z))\nabla P^J(z))$$

with $P(0,\cdot,z)-P^J(0,\cdot,z)=0$. We therefore have that

$$\begin{aligned}\text{(B.1)}\quad&\int_0^T\left(\frac{\partial}{\partial t}(P(z)-P^J(z)),(P(z)-P^J(z))\right)dt\\&+\int_0^T\int_D K\nabla(P(z)-P^J(z))\cdot\nabla(P(z)-P^J(z))dxdt\\&=-\int_0^T\int_D(K(z)-K^J(z))\nabla P^J(z)\cdot\nabla(P(z)-P^J(z))dxdt.\end{aligned}$$

Thus

$$\begin{aligned}&K_{\min}(z)\|P(z)-P^J(z)\|^2_{L^2((0,T);V)}\\&\le\|K(z)-K^J(z)\|_{L^\infty(D)}\|\nabla P^J(z)\|_{L^2((0,T);H)}\|\nabla(P(z)-P^J(z))\|_{L^2((0,T);H)}.\end{aligned}$$

From this we have

$$\begin{aligned}\|P(z)-P^J(z)\|_{L^2((0,T);V)}&\le\frac{1}{K_{\min}(z)}\|K(z)-K^J(z)\|_{L^\infty(D)}\|P^J(z)\|_{L^2((0,T);V)}\\&\le c\left(\sum_{j>J}|z_j|b_j\right)\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right).\end{aligned}$$

Similarly, we have from (2.5)

$$\frac{\partial}{\partial t}\frac{\partial}{\partial t}(P(z)-P^J(z))-\nabla\cdot(K\nabla\frac{\partial}{\partial t}(P(z)-P^J(z)))=\nabla\cdot(K(z)-K^J(z))\nabla\frac{\partial P^J}{\partial t}(z)),$$

with $\frac{\partial}{\partial t}(P(z)-P^J(z))(0,\cdot,z)=\nabla\cdot((K(z)-K^J(z))\nabla g)$. We deduce

$$\begin{aligned}&K_{\min}(z)\left\|\frac{\partial}{\partial t}(P(z)-P^J(z))\right\|^2_{L^2((0,T);V)}\\ &\leq\|K(z)-K^J(z)\|_{L^\infty(D)}\left\|\nabla\frac{\partial P^J}{\partial t}(z)\right\|_{L^2((0,T);H)}\left\|\nabla\frac{\partial}{\partial t}(P(z)-P^J(z))\right\|_{L^2((0,T);H)}\\ &\qquad+\|\nabla\cdot(K(z)-K^J(z))\nabla g\|^2_H.\end{aligned}$$

We thus have

$$\begin{aligned}&\|\frac{\partial}{\partial t}(P(z)-P^J(z))\|_{L^2((0,T);V)}\\ &\qquad\leq\frac{c}{K_{\min}(z)}\|K(z)-K^J(z)\|_{L^\infty(D)}\|\nabla\frac{\partial P^J}{\partial t}(z)\|_{L^2((0,T);H)}\\ &\qquad\qquad+\frac{c}{K_{\min}(z)^{1/2}}\|\nabla\cdot(K(z)-K^J(z))\nabla g\|_H.\end{aligned}$$

Using $|\exp(x)-\exp(y)|\leq|x-y|(\exp(x)+\exp(y))$ which holds for all $x,y\in\mathbb{R}$, there exists $c>0$ such that for all $J\in\mathbb{N}$ and for every $z\in U$ holds

$$\begin{aligned}&\|K(z)-K^J(z)\|_{L^\infty(D)}\\ &\qquad=\|\exp(\bar{K}(x)+\sum_{j=1}^\infty z_j\psi_j(x))-\exp(\bar{K}(x)+\sum_{j=1}^J z_j\psi_j(x))\|_{L^\infty(D)}\\ &\qquad\leq c\exp\left(2\sum_{j=1}^\infty|z_j|b_j\right)\left(\sum_{j>J}|z_j|b_j\right).\end{aligned}$$

We note furthermore that for every $z \in U$

$$
\begin{aligned}
&\|\nabla K(z) - \nabla K^J(z)\|_{L^\infty(D)} \\
&\le \Bigg\| \exp\left(\bar{K}(x) + \sum_{j=1}^\infty z_j\psi_j(x)\right)\left(\nabla\bar{K}(x) + \sum_{j=1}^\infty z_j\nabla\psi_j(x)\right) \\
&\qquad - \exp\left(\bar{K}(x) + \sum_{j=1}^J z_j\psi_j(x)\right)\left(\nabla\bar{K}(x) + \sum_{j=1}^J z_j\nabla\psi_j(x)\right)\Bigg\|_{L^\infty(D)} \\
&\le \left\| \exp\left(\bar{K}(x) + \sum_{j=1}^\infty z_j\psi_j(x)\right) - \exp\left(\bar{K}(x) + \sum_{j=1}^J z_j\psi_j(x)\right)\right\|_{L^\infty(D)} \cdot \\
&\qquad\qquad \left\|\nabla\bar{K}(x) + \sum_{j=1}^J z_j\nabla\psi_j(x)\right\|_{L^\infty(D)} \\
&\quad + \left\|\exp\left(\bar{K}(x) + \sum_{j=1}^\infty z_j\psi_j(x)\right)\right\|_{L^\infty(D)}\left\|\sum_{j>J} z_j\nabla\psi_j(x)\right\|_{L^\infty(D)} \\
&\le c\exp\left(\sum_{j=1}^\infty |z_j|b_j\right)\left(\sum_{j>J}|z_j|b_j\right)\left(1 + \sum_{j=1}^J |z_j|\bar{b}_j\right) \\
&\qquad + c\exp\left(\sum_{j=1}^\infty |z_j|b_j\right)\left(\sum_{j>J}|z_j|\bar{b}_j\right) \\
&\le c\exp\left(\sum_{j=1}^\infty |z_j|(b_j + \bar{b}_j)\right)\left(\sum_{j>J}|z_j|(b_j+\bar{b}_j)\right).
\end{aligned}
$$

From Lemma B.1, there exists $c > 0$ such that for every $z \in U$

$$
\left\|\frac{\partial}{\partial t}(P(z) - P^J(z))\right\|_{L^2((0,T);V)} \le c\exp\left(c\sum_{j=1}^\infty |z_j|(b_j+\bar{b}_j)\right)\left(\sum_{j>J}|z_j|(b_j+\bar{b}_j)\right).
$$

□

We now prove Proposition 4.1.

Proof of Proposition 4.1. From Lemma B.1, there exists a constant $c > 0$ such that for every $z \in U$ holds

$$
|\mathcal{G}(z)| \le c\exp(c\sum_{j=1}^\infty |z_j|(b_j+\bar{b}_j)), \quad |\mathcal{G}^J(z)| \le c\exp(c\sum_{j=1}^\infty |z_j|(b_j+\bar{b}_j))
$$

for a constant $c > 0$. From Lemma B.2, we have

$$
|\mathcal{G}(z) - \mathcal{G}^J(z)| \le c(\sum_{j>J}|z_j|(b_j+\bar{b}_j))\exp(c\sum_{j=1}^\infty |z_j|(b_j+\bar{b}_j)).
$$

Let $Z(\delta)$ and $Z^J(\delta)$ be the normalizing constants in (3.1) and (4.3) respectively. We first show that Z^J is uniformly bounded below from 0 uniformly with respect to J. From (4.2) and Lemma B.1, we have that

$$|\Phi^J(z;\delta))| \le c(\delta)\exp(\sum_{j=1}^{J}(b_j+\bar{b}_j)).$$

From Lemma A.1, we have that

$$\int_U \Phi^J(z;\delta)d\gamma(z) < \Lambda$$

uniformly for all J. Fixing a constant $C>0$, the γ measure of the set $z\in U$ such that $\Phi^J(z;\delta)>C$ is less than Λ/C so the γ measure of the set of $z\in U$ such that $\Phi^J(z;\delta)\le C$ is more than $1-\Lambda/C$ which is positive when C is sufficiently large. Thus

$$Z^J(\delta)=\int_U \exp(-\Phi^J(z;\delta))d\gamma(z) > (1-\Lambda/C)\exp(-C)>0.$$

As in equation (3.6), we have

$$d_{Hell}(\gamma^\delta,\gamma^{J,\delta})^2 \le I_1+I_2,$$

where

$$I_1=\frac{2}{Z(\delta)}\int_U\left(\exp(-\frac{1}{2}\Phi(z;\delta))-\exp(-\frac{1}{2}\Phi^J(z;\delta))\right)^2 d\gamma(z), \tag{B.2}$$

and

$$I_2=2|Z(\delta)^{-1/2}-Z^J(\delta)^{-1/2}|^2\int_U \exp(-\Phi^J(z;\delta))d\gamma(z). \tag{B.3}$$

We then have

(B.4)

$$\begin{aligned}\left|\exp(-\frac{1}{2}\Phi(z;\delta))-\exp(-\frac{1}{2}\Phi^J(z;\delta))\right| &\le c(|\delta|+|\mathcal{G}(z)|+|\mathcal{G}^J(z)|)|\mathcal{G}(z)-\mathcal{G}^J(z)|\\ &\le c(\delta)\exp\left(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)\right)\left(\sum_{j>J}|z_j|(b_j+\bar{b}_j)\right).\end{aligned}$$

Thus

$$I_1\le c(\delta)\int_U \exp\left(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)\right)\left(\sum_{j>J}|z_j|(b_j+\bar{b}_j)\right)^2 d\gamma(z).$$

Now we show that the right hand side of this inequality is bounded by $c(\delta)J^{-2q}$. Denoting by γ_1 the standard normal probability measure in $\mathbb{R}$, and denoting $B_j :=$

$b_j+\bar b_j$, we have

$$
\begin{aligned}
&\int_U \exp\Big(c\sum_{j=1}^\infty |z_j|B_j\Big)\Big(\sum_{j=J+1}^\infty |z_j|B_j\Big)^2 d\gamma(z)\\
=\ &\int_U \exp\Big(c\sum_{j=1}^\infty |z_j|B_j\Big)\Big(\sum_{i,j=J+1}^\infty B_iB_j|z_i||z_j|\Big) d\gamma(z)\\
\le\ &\sum_{i=J+1}^\infty B_i^2\int_{-\infty}^\infty \exp(cB_i|z_i|)z_i^2 d\gamma_1(z_i)\prod_{\substack{k=1\\k\neq i}}^\infty\int_{-\infty}^\infty \exp(cB_k|z_k|)d\gamma_1(z_k)\\
&+\sum_{\substack{i,j=J+1\\i\neq j}}^\infty B_iB_j\int_{-\infty}^\infty \exp(cB_i|z_i|)|z_i|d\gamma_1(z_i)\cdot\int_{-\infty}^\infty \exp(cB_j|z_j|)|z_j|d\gamma_1(z_j)\\
&\qquad\cdot\prod_{\substack{k=1\\k\neq i,j}}^\infty\int_{-\infty}^\infty \exp(cB_k|z_k|)d\gamma_1(z_k)\ .
\end{aligned}
$$

From Lemma A.1, there exists $c>0$ (independent of B_j) such that

$$\int_{-\infty}^\infty t^2\exp(-t^2/2+|t|s)\frac{dt}{\sqrt{2\pi}}\le c\exp(s^2/2)(1+s^2),$$

and

$$\int_{-\infty}^\infty |t|\exp(-t^2/2+|t|s)\frac{dt}{\sqrt{2\pi}}\le c\exp(s^2/2)(1+s)\ .$$

We deduce that there exists $C>0$ such that for all $J\ge 1$

$$
\begin{aligned}
&\int_U \exp\Big(c\sum_{j=1}^\infty B_j|z_j|\Big)\Big(\sum_{j=J+1}^\infty B_j|z_j|\Big)^2 d\gamma(z)\\
\le\ & C\sum_{i=J+1}^\infty B_i^2(1+B_i^2)\exp\left(\sum_{k=1}^\infty c^2B_k^2/2+cB_k\sqrt{2/\pi}\right)\\
&+C\sum_{\substack{i,j=J+1\\i\neq j}}^\infty B_iB_j(1+B_i)(1+B_j)\exp\left(\sum_{k=1}^\infty c^2B_k^2/2+cB_k\sqrt{2/\pi}\right)\\
\le\ & C\left(\sum_{j=J+1}^\infty B_j\right)^2\le CJ^{-2q}.
\end{aligned}
$$

Thus $I_1\le CJ^{-2q}$. To bound I_2, we observe that

(B.5)

$$|Z(\delta)^{-\frac12}-Z^J(\delta)^{-\frac12}|^2\le C\max\{Z(\delta)^{-3},Z^J(\delta)^{-3}\}|Z(\delta)-Z^J(\delta)|^2\le C|Z(\delta)-Z^J(\delta)|^2.$$

From (B.4),

$$|Z(\delta)-Z^J(\delta)|\le\int_U|\exp\left(-\Phi(z;\delta)\right)-\exp\left(-\Phi^J(z;\delta)\right)|d\gamma(z)\le CJ^{-q}. \tag{B.6}$$

Therefore,

$$I_2\le CJ^{-q}.$$

$\square$

Appendix C. Proof of Proposition 4.2

We note the following approximation for the approximating problem (4.5).

Lemma C.1. *Let D be a convex domain. Under Assumption 2.3 and the hypothesis of Proposition 4.2, there is a constant $c>0$ such that for all $m=1,\ldots,M$, and for every $z\in U$, the solution of problem (4.5) satisfes*

$$\|P_m^{J,h,k}(z)-P^J(t_m,\cdot,z)\|_V\le c\exp\left(c\sum_{j=1}^J|z_j|(b_j+\bar{b}_j)\right)(h+k)+c\|g-g^l\|_V.$$

Proof We start with

$$\begin{aligned}\|P_m^{J,h,k}(\cdot,z)&-P^J(t_m,\cdot,z)\|_V\\&\le\|P_m^{J,h,k}(\cdot,z)-R_hP^J(t_m,\cdot,z)\|_V+\|R_hP^J(t_m,\cdot,z)-P^J(t_m,\cdot,z)\|_V\\&=\|\theta^m\|_V+\|\varphi^m\|_V\end{aligned}$$

where

$$\theta^m=P_m^{J,h,k}(\cdot,z)-R_hP^J(t_m,\cdot,z)\ ,\quad \varphi^m=R_hP^J(t_m,\cdot,z)-P^J(t_m,\cdot,z)$$

and that $R_h:V\to V^l$ is the projection satisfying

$$\int_D K^J(\cdot,z)\nabla(R_h\psi)\cdot\nabla\phi dx=\int_D K^J(\cdot,z)\nabla\psi\cdot\nabla\phi dx$$

$\forall\psi\in V$ and $\forall\phi\in V^l$. We consider

$$\begin{aligned}&\int_D\frac{\theta^m-\theta^{m-1}}{k}\phi dx+\int_D K^J(\cdot,z)\nabla\theta^m\cdot\nabla\phi dx\\&=\int_D\frac{P_m^{J,h,k}(\cdot,z)-P_{m-1}^{J,h,k}(\cdot,z)}{k}\phi dx-\int_D\frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k}\phi dx\\&\quad+\int_D K^J(\cdot,z)\nabla P_m^{J,h,k}(\cdot,z)\cdot\nabla\phi dx-\int_D K^J(\cdot,z)\nabla R_hP^J(t_m,\cdot,z)\cdot\nabla\phi dx.\end{aligned}$$

From (4.5) and the definition of R_h,

$$\begin{aligned}&\int_D\frac{\theta^m-\theta^{m-1}}{k}\phi dx+\int_D K^J(\cdot,z)\nabla\theta^m\cdot\nabla\phi dx\\&\qquad=\int_D f(t_m,\cdot)\phi dx-\int_D\frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k}\phi dx\\&\qquad\qquad-\int_D K^J(\cdot,z)\nabla P^J(t_m,\cdot,z)\cdot\nabla\phi dx.\end{aligned}$$

Adding and subtracting $\int_D \frac{\partial P^J}{\partial t}(t_m,\cdot,z)\phi dx$,

$$\begin{aligned}&\int_D \frac{\theta^m-\theta^{m-1}}{k}\phi dx + \int_D K^J(\cdot,z)\nabla\theta^m\cdot\nabla\phi dx\\ &= \int_D f(t_m,\cdot)\phi dx\\ &\quad - \int_D \frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k}\phi dx + \int_D \frac{\partial P^J}{\partial t}(t_m,\cdot,z)\phi dx\\ &\qquad - \int_D \frac{\partial P^J}{\partial t}(t_m,\cdot,z)\phi dx - \int_D K^J(\cdot,z)\nabla P^J(t_m,\cdot,z)\cdot\nabla\phi dx.\end{aligned}$$

Using (3.4),

$$\begin{aligned}&\int_D \frac{\theta^m-\theta^{m-1}}{k}\phi dx + \int_D K^J(\cdot,z)\nabla\theta^m\cdot\nabla\phi dx\\ &= \int_D f(t_m,\cdot)\phi dx\\ &\quad - \int_D \frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k}\phi dx + \int_D \frac{\partial P^J}{\partial t}(t_m,\cdot,z)\phi dx\\ &\qquad - \int_D f(t_m,\cdot)\phi dx.\end{aligned}$$

This implies that

$$\int_D \frac{\theta^m-\theta^{m-1}}{k}\phi dx + \int_D K^J(\cdot,z)\nabla\theta^m\cdot\nabla\phi dx = -\int_D \omega^m\phi dx \tag{C.1}$$

where

$$\omega^m = \frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k} - \frac{\partial P^J}{\partial t}(t_m,\cdot,z).$$

We write $\omega^m = \omega_1^m + \omega_2^m$ where

$$\begin{aligned}\omega_1^m &= \frac{R_hP^J(t_m,\cdot,z)-R_hP^J(t_{m-1},\cdot,z)}{k} - \frac{P^J(t_m,\cdot,z)-P^J(t_{m-1},\cdot,z)}{k}\\ \omega_2^m &= \frac{P^J(t_m,\cdot,z)-P^J(t_{m-1},\cdot,z)}{k} - \frac{\partial P^J}{\partial t}(t_m,\cdot,z).\end{aligned}$$

We note that

$$\int_D \nabla\theta^m\cdot\nabla\frac{\theta^m-\theta^{m-1}}{k}dx = \frac{1}{2}\left[\frac{\|\nabla\theta^m\|_H^2-\|\nabla\theta^{m-1}\|_H^2}{k} + k\left\|\nabla\left(\frac{\theta^m-\theta^{m-1}}{k}\right)\right\|_H^2\right].$$

From (C.1), under Assumption 2.3, letting $\phi = \frac{\theta^m-\theta^{m-1}}{k}$,

$$\left\|\frac{\theta^m-\theta^{m-1}}{k}\right\|_H^2 + K_{\min}\int_D \nabla\theta^m\cdot\nabla\left(\frac{\theta^m-\theta^{m-1}}{k}\right)dx \le \int_D -\omega^m\left(\frac{\theta^m-\theta^{m-1}}{k}\right)dx.$$

Thus

$$\begin{aligned}&\left\|\frac{\theta^m-\theta^{m-1}}{k}\right\|_H^2 + \frac{K_{\min}}{2}\left[\frac{\|\nabla\theta^m\|_H^2-\|\nabla\theta^{m-1}\|_H^2}{k} + k\left\|\nabla\left(\frac{\theta^m-\theta^{m-1}}{k}\right)\right\|_H^2\right]\\ &\qquad\qquad \le \frac{1}{2}\|\omega^m\|_H^2 + \frac{1}{2}\left\|\frac{\theta^m-\theta^{m-1}}{k}\right\|_H^2,\end{aligned}$$

which implies

$$\frac{K_{\min}}{2}\frac{\|\nabla\theta^m\|_H^2-\|\nabla\theta^{m-1}\|_H^2}{k}\le\frac{1}{2}\|\omega^m\|_H^2.$$

Thus

$$\|\theta^m\|_V^2\le\frac{1}{K_{\min}}k\|\omega^m\|_H^2+\|\theta^{m-1}\|_V^2,$$

and recursively,

$$\|\theta^m\|_V^2\le\frac{1}{K_{\min}}k\sum_{i=1}^m\|\omega^i\|_H^2+\|\theta^0\|_V^2.$$

Therefore

$$\|\theta^m\|_V^2\le\frac{ck}{K_{\min}}\sum_{i=1}^m\|\omega_1^i\|_H^2+\frac{ck}{K_{\min}}\sum_{i=1}^m\|\omega_2^i\|_H^2+\|\theta^0\|_V^2. \tag{C.2}$$

We have

$$\omega_1^m=(R_h-I)\frac{1}{k}\int_{t_{m-1}}^{t_m}P_t^J(s,\cdot,z)ds=\frac{1}{k}\int_{t_{m-1}}^{t_m}(R_h-I)P_t^J(s,\cdot,z)ds.$$

From the duality argument in the proof of Theorem 1.1 of Thomee [24], we have

$$\|(R_h-I)P_t^J(t,\cdot,z)\|_H\le ch\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right)\|P_t^J(t,\cdot,z)\|_V.$$

We have

$$k\omega_1^m=\int_{t_{m-1}}^{t_m}(R_h-I)P_t^J(s,\cdot,z)ds.$$

Thus

$$\begin{aligned}k^2\|\omega_1^m\|_H^2&\le c\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right)h^2\left(\int_{t_{m-1}}^{t_m}\|P_t^J(s,\cdot,z)\|_Vds\right)^2\\&\le c\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right)h^2k\int_{t_{m-1}}^{t_m}\|P_t^J(s,\cdot,z)\|_V^2ds.\end{aligned}$$

Thus for all $m=1,\dots,M$

$$k\sum_{i=1}^m\|\omega_1^i\|_H^2\le c\exp\left(c\sum_{j=1}^\infty|z_j|(b_j+\bar b_j)\right)h^2\int_0^T\|P_t^J(s,\cdot,z)\|_V^2ds.$$

For ω_2^m, we have

$$k\omega_2^m=-\int_{t_{m-1}}^{t_m}(s-t_{m-1})P_{tt}^J(s,\cdot,z)ds$$

so

$$k^2\|\omega_2^m\|_H^2\le k^2\left(\int_{t_{m-1}}^{t_m}\|P_{tt}^J(s,\cdot,z)\|_Hds\right)^2,$$

i.e.

$$\|\omega_2^m\|_H^2\le\left(\int_{t_{m-1}}^{t_m}\|P_{tt}^J(s,\cdot,z)\|_Hds\right)^2\le k\int_{t_{m-1}}^{t_m}\|P_{tt}^J(s,\cdot,z)\|_H^2ds.$$

Thus

$$k\sum_{i=1}^{m}\|\omega_2^i\|_H^2 \le ck^2\int_0^T \|P_{tt}^J(s,\cdot,z)\|_H^2 ds.$$

Using Lemma B.1 for P^J where we consider z with only the first J components being nonzero, we have that

$$\|P_t^J(z)\|_{L^2((0,T);V)} \le c\exp(c\sum_{j=1}^{J}|z_j|(b_j+\bar{b}_j)).$$

Using [12, Theorem 5, page 360] for equation (2.5), as $\frac{\partial f}{\partial t} \in L^2((0,T);H)$ and $f(0,\cdot)+\nabla\cdot(K\nabla g)\in V$, we have that

$$\|P_{tt}^J(z)\|_{L^2((0,T);H)} \le c\frac{K_{\max}(z)}{K_{\min}(z)}\left(\left\|\frac{\partial f}{\partial t}\right\|_{L^2((0,T);H)} + \|f(0,\cdot)+\nabla\cdot(K\nabla g)\|_V\right),$$

(the dependence of the multiplying constant C in the right hand side of this estimate in the statement of [12, Theorem 5 page 360] on $K_{\max}(z)/K_{\min}(z)$ can be straightforwardly verified by following the proof of this theorem on pages 361-362 of [12]). Thus

$$\|P_{tt}^J(z)\|_{L^2((0,T);H)} \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)).$$

From (2.5), we have $\frac{dP^J}{dt}(z)\in C([0,T];H)$ and

$$\left\|\frac{\partial P^J}{\partial t}(z)\right\|_{C([0,T];H)} \le c\left(\left\|\frac{\partial P^J}{\partial t}(z)\right\|_{L^2([0,T];V)} + \left\|\frac{\partial P^J}{\partial t}(z)\right\|_{H^1((0,T);V')}\right).$$

Thus

$$\left\|\frac{\partial P^J}{\partial t}(z)\right\|_{C([0,T];H)} \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)). \tag{C.3}$$

From (2.4), for every $z\in U$ and for every $0<t\le T$ there holds in $H^{-1}(\mathrm{D})$

$$-\Delta P^J(z) = \frac{1}{K^J(z)}\left(f - \frac{\partial P^J}{\partial t}(z) + \nabla K^J(z)\cdot\nabla P^J(z)\right). \tag{C.4}$$

Applying the shift theorem for the weighed Sobolev spaces, there exists a constant c (which only depends on the domain D) such that for every $0<t\le T$ and for every $z\in U$ holds

$$\|P^J(t,\cdot,z)\|_{H^2(D)}$$
$$\le \frac{c}{K_{\min}^J(z)}\left(\|f(t,\cdot)\|_H + \left\|\frac{\partial P^J}{\partial t}(t,\cdot,z)\right\|_H + \|\nabla K^J(t,\cdot,z)\|_{L^\infty(D)}\|\nabla P^J(t,\cdot,z)\|_H\right).$$

From Lemma B.1, we have

$$\sup_{t\in[0,T]}\|\nabla P^J(t,\cdot,z)\|_H \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)).$$

From (C.3),

$$\sup_{t\in[0,T]}\left\|\frac{\partial P^J}{\partial t}(t,\cdot,z)\right\|_H \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar{b}_j)).$$

Thus

$$\|P^J(t,\cdot,z)\|_{H^2(D)} \le c\exp\left(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar b_j)\right).$$

We further have $\varphi^m = (R_h - I)P^J(t_m,\cdot,z)$ so

$$\|\varphi^m\|_V \le ch\|P^J(t_m,\cdot,z)\|_{H^2(D)}.$$

Therefore for all $m = 1,\ldots,M$

$$\|P_m^{J,h,k}(z) - P^J(t_m,\cdot,z)\|_V \le c\exp(c\sum_{j=1}^{\infty}|z_j|b_j)(h+k)\cdot$$

$$\left(\|P_t^J(z)\|_{L^2((0,T);V)} + \|P_{tt}^J(z)\|_{L^2((0,T);H)} + \sup_{t\in(0,T)}\|P^J(t,\cdot,z)\|_{H^2(D)})\right) + \|g-g^L\|_V.$$

Hence, there is a constant $c>0$ such that $\forall\, z\in U$

$$\|P_m^{J,h,k}(z) - P^J(t_m,\cdot,z)\|_V \le c\exp\left(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar b_j)\right)(h+k) + \|g-g^l\|_V.$$

We are now ready to prove Proposition 4.2.

Proof of Proposition 4.2: We note that it is sufficient to show that $d_{Hell}(\gamma^{J,\delta},\gamma^{J,h,k,\delta}) \le C(h+k)$. From Lemmas B.1 and C.1, when choosing g^l such that $\|g-g^l\|_V \le ch$, we have

$$\|P_{\tau_i}^{J,h,k}(z)\|_V \le c\exp(c\sum_{j=1}^{J}|z_j|(b_j+\bar b_j))$$

so

$$|\mathcal{G}^{J,h,k}(z)| \le c\exp(c\sum_{j=1}^{\infty}|z_j|(b_j+\bar b_j)).$$

We have further from Lemma C.1 that $\forall\, z\in U$

$$|\ell_i(P^J(\tau_i,\cdot,z)) - \ell_i(P_{\tau_i}^{J,h,k}(\cdot,z))| \le c\exp\left(c\sum_{j=1}^{J}|z_j|(b_j+\bar b_j)\right)(h+k).$$

Thus

$$|\mathcal{G}^J(z) - \mathcal{G}^{J,h,k}(z)| \le c\exp\left(c\sum_{j=1}^{J}|z_j|(b_j+\bar b_j)\right)(h+k).$$

We proceed as in the proof of Proposition 4.1. Let $Z^J(\delta)$ and $Z^{J,h,k}(\delta)$ be the normalizing constants in (4.1) and (4.6) respectively. We have

$$2d_{Hell}(\gamma^{J,\delta},\gamma^{J,h,k,\delta})^2 \le I_1 + I_2,$$

where

$$I_1 = \frac{2}{Z^J(\delta)}\int_U\left(\exp(-\frac12\Phi^J(z;\delta)) - \exp(-\frac12\Phi^{J,h,k}(z;\delta))\right)^2 d\gamma(z), \tag{C.5}$$

and

$$I_2 = 2|Z^J(\delta)^{-1/2} - Z^{J,h,k}(\delta)^{-1/2}|^2\int_U \exp(-\frac12\Phi^{J,h,k}(z;\delta))d\gamma(z). \tag{C.6}$$

A proof which is similar to that for lower-bounding $Z^J(\delta)$ above shows that $Z^{J,h,k}(\delta)$ is uniformly bounded below from 0 for all J, h and k. We then obtain that there is a constant $c > 0$ such that for all J, h, k and all δ with $|\delta| \le r$ holds

$$\begin{aligned}\Big|\exp(-\frac{1}{2}\Phi^J(z;\delta)) - \exp(-\frac{1}{2}\Phi^{J,h,k}(z;\delta))\Big| \\ &\le c\left(|\delta| + |\mathcal{G}^J(z)| + |\mathcal{G}^{J,h,k}(z)|\right)|\mathcal{G}^J(z) - \mathcal{G}^{J,h,k}(z)| \\ &\le c\exp\left(c\sum_{j=1}^J |z_j|(b_j + \bar{b}_j)\right)(h+k).\end{aligned}$$

Using the first inequality of Lemma A.1 we have

$$I_1 \le c\exp\left(c\sum_{j=1}^J (b_j + \bar{b}_j)^2 + c\sum_{j=1}^J (b_j + \bar{b}_j)\right)(h+k)^2 \le c(h+k)^2.$$

Similarly, we have $I_2 \le c(h+k)^2$, implying that $d_{Hell}(\gamma^{J,\delta}, \gamma^{J,h,k,\delta}) \le c(h+k)^2$. Together with $d_{Hell}(\gamma^{\delta}, \gamma^{J,\delta}) \le cJ^{-q}$, we get the conclusion. □

References

[1] Robert J. Adler. *The geometry of random fields*. John Wiley & Sons Ltd., Chichester, 1981. Wiley Series in Probability and Mathematical Statistics.

[2] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numer.*, 28:1–174, 2019.

[3] Andrea Barth, Annika Lang, and Christoph Schwab. Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT*, 53(1):3–27, 2013.

[4] 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.

[5] 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.

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

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

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

[9] 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.

[10] 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.

[11] 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.

[12] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[13] I.G. Graham, F.Y. Kuo, J.A. Nichols, R. Scheichl, Ch. Schwab, and I.H. Sloan. Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. *Numerische Mathematik*. Published online 2nd December 2014.

[14] 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.

[15] V. H. Hoang, J. H. Quek, and Ch. Schwab. Analysis of a multilevel Markov Chain Monte Carlo Finite Element Method for Bayesian inversion of log-normal diffusion. *Inverse Problems*, 36:035021, 2020.

[16] 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.
[17] Viet Ha Hoang, Christoph Schwab, and Andrew Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8), 2013.
[18] Jonas Latz, Iason Papaioannou, and Elisabeth Ullmann. Multilevel Sequential^2 Monte Carlo for Bayesian inverse problems. *J. Comput. Phys.*, 368:154–178, 2018.
[19] Kody Law, Andrew Stuart, and Konstantinos Zygalakis. *Data assimilation*, volume 62 of *Texts in Applied Mathematics*. Springer, Cham, 2015. A mathematical introduction.
[20] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, second edition, 2009.
[21] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.
[22] Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 2010.
[23] A.L. Teckentrup, R. Scheichl, M.B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.
[24] Vidar Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.
[25] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987. Translated from the German by C. B. Thomas and M. J. Thomas.
[26] 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