arXiv:1207.2411v2 [math.NA] 30 Apr 2013

Complexity Analysis of Accelerated MCMC Methods for Bayesian Inversion

Viet Ha Hoang[1], Christoph Schwab[2] and Andrew M. Stuart[3]

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

E-mail: `vhhoang@ntu.edu.sg`
[2] Seminar for Applied Mathematics, ETH, 8092 Zurich, Switzerland

E-mail: `christoph.schwab@sam.math.ethz.ch`
[3] Mathematics Institute, Warwick University, Coventry CV4 7AL, UK

E-mail: `a.m.stuart@warwick.ac.uk`

Abstract. The Bayesian approach to inverse problems, in which the posterior probability distribution on an unknown field is sampled for the purposes of computing posterior expectations of quantities of interest, is starting to become computationally feasible for partial differential equation (PDE) inverse problems. Balancing the sources of error arising from finite dimensional approximation of the unknown field, the PDE forward solution map and the sampling of the probability space under the posterior distribution is essential for the design of efficient computational Bayesian methods for PDE inverse problems.

We study Bayesian inversion for a model elliptic PDE with unknown diffusion coefficient. We provide complexity analyses of several Markov Chain-Monte Carlo (MCMC) methods for the efficient numerical evaluation of expectations under the Bayesian posterior distribution, given data δ. Particular attention is given to bounds on the overall work required to achieve a prescribed error level ε. Specifically, we first bound the computational complexity of "plain" MCMC, based on combining MCMC sampling with linear complexity multilevel solvers for elliptic PDE. Our (new) work versus accuracy bounds show that the complexity of this approach can be quite prohibitive. Two strategies for reducing the computational complexity are then proposed and analyzed: first, a sparse, parametric and deterministic generalized polynomial chaos (gpc) "surrogate" representation of the forward response map of the PDE over the entire parameter space, and, second, a novel Multi-Level Markov Chain Monte Carlo (MLMCMC) strategy which utilizes sampling from a multilevel discretization of the posterior and of the forward PDE.

For both of these strategies we derive asymptotic bounds on work versus accuracy, and hence asymptotic bounds on the computational complexity of the algorithms. In particular we provide sufficient conditions on the regularity of the unknown coefficients of the PDE, and on the approximation methods used, in order for the accelerations of MCMC resulting from these strategies to lead to complexity reductions over "plain" MCMC algorithms for Bayesian inversion of PDEs.

1. Introduction

Many inverse problems arising in PDEs require determination of uncertain parameters u from finite dimensional observation data δ. We assume u and δ to be related by

$$\delta = \mathcal{G}(u) + \vartheta . \tag{1}$$

Here u, which we assume to belong to an (infinite-dimensional) function space, is an unknown or uncertain input (e.g. an uncertain coefficient) to a differential equation and $\mathcal{G}$ is the "forward" mapping taking one instance (one realization) of input u into a finite and noisy set of observations. We model these observations mathematically as continuous linear functionals on the solution of the governing partial differential equation. In (1), the parameter ϑ represents noise arising when observing. For problems such as these the Bayesian formulation [17, 29] is an attractive and natural one, because it allows for explicit incorporation of the statistical properties of the observational noise, because it admits the possibility of quantifying uncertainty in the solution and because it allows for clear mathematical modelling of the prior information required to account for the under-determined nature of the inversion. However the Bayesian approach to such inverse problems requires the probing of probability measures in infinite dimensional spaces and is hence a substantial computational task in which the space of the unknown parameter, the forward map and the probability space must all be approximated finite-dimensionally. Choosing appropriate approximation techniques, and balancing the resulting errors so as to optimize the overall computational cost per unit error, is thus a significant problem with potential ramifications for the resolution of many inverse problems arising in applications. The purpose of this paper is to provide a detailed and comprehensive theoretical study of this problem in computational complexity.

A key point to appreciate concerning our work herein is to understand it in the context of the existing statistics literature concerning the complexity of MCMC methods. This statistics literature is focussed on the error stemming from the central limit theorem estimate of the convergence of sample path averages, and additionally on issues such as the "burn-in" time for the Markov chain to reach stationarity. Such analyses remain central in analyzing the computational complexity of MCMC methods for inverse problems, but must be understood, additionally, in the context of errors stemming from the finite dimensional approximation of the unknown function and the the forward model. The purpose of this paper is to provide what we believe is the first complete, rigorous analysis of computational complexity of MCMC methods for inverse problems which balances the errors resulting from both the central limit theorem and finite dimensional approximation.

The analysis is necessarily quite complex as it involves balancing errors from many different sources. Thus we address a specific inverse problem, namely determination of the diffusion coefficient in an elliptic PDE from measurements of the solution. We approximate the probability space by means of an MCMC independence sampler, and we approximate the parameter space by means of Karhunen-Loeve expansion and generalized polynomial chaos, and we approximate the forward problem by finite

element type methods. However the ideas are generic and, with substantial work, will be transferable to other inverse problems, MCMC samplers and approximation methodologies. In our analysis, we place uniform prior measures on the unknown function u. This simplifies certain steps in the present analysis; however, our results will also apply to priors whose densities have bounded supports. Log-normal priors as used, for example, in [18] and the references therein, require modifications in various parts of our analysis; this shall be dealt with elsewhere. Finally, we concentrate on the independence sampler as it is, currently, the only function-space MCMC method for which an analysis of convergence rates which accounts for sampling and discretization errors is available; extensions to other MCMC algorithms are conceivable, for example by continuing the analyses intiated in [14, 31].

1.1. Overview of Paper

The methods we study incur two principal sources of error. First, the *sampling error* arising from estimating expectations conditional on given data δ by sample averages of M realizations of the unknown u, drawn from the posterior measure ρ^δ. The error in doing so will decay as $M^{-\frac{1}{2}}$ as the number M of draws of u tends to ∞. Second, the *discretization error* arising from approximation of the system response for each draw of u, i.e. the error of approximating $\mathcal{G}(u)$. For expository purposes, and to cover a wide range of discretization techniques, we let N_{dof} denote the total number of degrees of freedom that are to be computed for evaluation of the Bayesian estimate. We assume that the discretization error decays as N_{dof}^{-a}, with rate $a > 0$ ‡. We also assume that the work per step of MCMC scales as N_{dof}^b, with $b > 0$; thus the total work necessary for numerical realization of M draws in the MCMC with discretized forward model scales as $M N_{\mathrm{dof}}^b$. If (as we show in the present paper for the independence sampler) the constant in the root mean square MCMC error bound of order $\mathcal{O}(M^{-\frac{1}{2}})$ is independent of $N_{\mathrm{dof}} \to \infty$ then a straightforward calculation shows that the work to obtain root mean square error ε will grow asymptotically as $\varepsilon \to 0$, as $\varepsilon^{-2-b/a}$. The ratio b/a is thus crucial to the overall computational complexity of the algorithm.

In this paper, we develop three ideas to speed up MCMC-based Bayesian inversion in systems governed by PDEs. The first idea, which underlies the preceding expository calculation concerning complexity, is that MCMC methods can be constructed whose convergence rate is indeed independent of the number of degrees of freedom N_{dof} used in the approximation of the forward map; the key idea here is to use MCMC algorithms which are defined on (infinite dimensional) function spaces, as overviewed in [7], and to use Galerkin projections of the forward map into finite-dimensional subspaces which employ Riesz bases in these function spaces. We term this the *Plain MCMC Method*. The second idea is that sparse, deterministic parametric representations of the forward map $u \mapsto \mathcal{G}(u)$, as analyzed in [3, 27, 4], can significantly reduce b by reducing computational complexity per step of the Markov chain, as the sparse approximation

---

‡ logarithmic corrections also occur, and will be made explicit in later sections

of $\mathcal{G}$ can be precomputed, prior to running the Markov chain, and simply needs to be evaluated at each step of the chain; this idea has been succesfully used in practice in [21, 19, 20]. We term the resulting algorithm the *gpc-MCMC Finite Element Method*. The third idea is that the representation of the forward map can be truncated adaptively at different discretization levels of the physical system of interest. Then, we propose a multilevel Monte Carlo acceleration of the MCMC method, in the spirit of the work of Giles for Monte Carlo [9], and prove that this allows further improvement of the computational complexity. The idea of extending Giles' work to Monte Carlo Markov Chain, to obtain what we term *MLMCMC Methods*, is actively being considered by several other research groups at present and we point in particular to [13], where the ideas are developed in the context of conditioned diffusions, and [18] where uncertainty quantification in subsurface flow is studied.

To give the reader some ideas of what will come ahead, the following table summarizes the complexity required to obtain the approximation error ε in mean square with respect to the probability space generated by the Markov chains for the three methods studied §

	Plain MCMC	gpc-MCMC	MLMCMC
Assume:	Assumptions 1(i),(ii), 7	Assumption 9	Assumptions 1(i),(ii), 7
Number of degrees of freedom	$O(\varepsilon^{-d-2})$	$O(\varepsilon^{-1/\tau})$	$O(\varepsilon^{-d})$
Number of floating point operations	$O(\varepsilon^{-d-2-1/q})$	$O(\varepsilon^{-\max(\alpha/\tau, 2+1/\sigma)})$	$O(\varepsilon^{-d-1/q})$

Here, $d \geq 2$ denotes the dimension of the spatial variable in the elliptic inverse problem and q determines the rate of decay in the sequence of coefficients of the Karhúnen-Loève expansion of the unknown coefficient in Assumption 1(ii). The parameters τ, α and σ concern the approximation properties of the gpc solver and are detailed in Assumption 9; τ quantifies the rate of convergence and α the polynomial scaling of the floating point complexity with respect to the total number of degrees of freedom that we assume for the gpc solver, whilst σ relates the number of degrees of freedom to the number of terms in the gpc approximation. The existence of methods which realize the assumptions made is discussed in detail in Appendices D and E; in particular gpc-based adaptive forward solvers of infinite-dimensional, parametric PDEs whose overall complexity scales polynomially in the number of degrees of freedom in the discretization of PDE of interest have recently become available; we refer to [27, 12, 3] and the references there. The three main results of the paper, substantiating the displayed table, are then Theorems 13, 18 and 21. Our results show that the MLMCMC FEM can achieve a given accuracy in work equal asymptotically, as the accuracy $\varepsilon \to 0$, to that of performing a single step of "plain" MCMC FEM; they also show that the gpc-accelerated MCMC FEM is superior to the "plain" MCMC FEM when the compression afforded by sparse gpc FE approximations of the parametric forward solution operator is large, ie. when the

§ ignoring multiplicative logarithmic factors for clarity of exposition

parameter τ is close to $1/d$ and when α is close to 1 – parameter regimes which do hold in many cases (see, e.g, [27] and the references there for stochastic Galerkin discretizations of the parametric forward problem).

The paper is organized as follows. Section 2 is devoted to the specific elliptic inverse problem which we study for illustrative purposes, and includes Bayesian formulation, a discussion of approximation techniques for the unknown field and the forward map, and properties of the independence MCMC sampler that we study. In section 3 we analyze the *Plain MCMC Method*, in Section 4 the *gpc-MCMC Finite Element Method* and in Section 5 we study the *MLMCMC Method*; these sections contain Theorems 13, 18 and 21 respectively. Section 6 contains our concluding remarks. The paper also includes five appendices which are devoted to mathematical developments which are necessary for the theoretical advances made in this paper, but which employ ideas already in the literature. In particular Appendix A concerns Lipschitz dependence of the forward problem on the unknown function u, whilst Appendix B is devoted to the formulation of Bayesian inverse problems on measure spaces and Appendix C to the properties of independence samplers in this setting. Appendices D and E concern the finite element and polynomial chaos approximation properties, and demonstrate how the assumptions made concerning them may be verified.

1.2. Overview of Notation

Because of the many different approximations used in this paper, together with the mixture of probability and PDE, the notation can be quite complicated. For this reason we devote this subsection to an overview of the notation, in order to help the reader.

To be concrete we assume that the observational noise appearing in (1) is a single realization of a centred Gaussian $N(0, \Sigma)$. If a prior measure ρ is placed on the unknown u then a Bayesian formulation of the inverse problem leads to the problem of probing the posterior probability measure ρ^δ given by

$$\frac{d\rho^\delta}{d\rho}(u) \propto \exp(-\Phi(u; \delta)) , \tag{2}$$

where

$$\Phi(u; \delta) = \frac{1}{2}|\delta - \mathcal{G}(u)|_\Sigma^2 . \tag{3}$$

Here $|\cdot|_\Sigma = |\Sigma^{-\frac{1}{2}} \cdot|$ with $|\cdot|$ the Euclidean norm. In Section 2 and Appendix B we verify that the formulae (2) and (3), which represent Bayes' formula in the infinite dimensional setting, hold for the model elliptic inverse problem which is under consideration here.

Our approximation of the unknown field u will be performed through truncation of the coefficients and Finite Element Galerkin discretizations will be used to approximate the forward solution map $\mathcal{G}$.

On the probability side we use the following notation: $\mathcal{B}^k$ denotes the sigma algebra of Borel subsets of $\mathbb{R}^k$. For a probability space $(\Omega, \mathcal{A}, \rho)$ consisting of the set Ω of elementary events, a sigma algebra $\mathcal{A}$ and a probability measure ρ, and a separable

Hilbert space H with norm $\|\cdot\|_H$ and for a summability exponent $0 < p \leq \infty$ we denote by $L^p(\Omega, \rho; H)$ the Bochner space of strongly measurable mappings from Ω to H which are p-summable.

We denote by $\mathbb{E}^\mu$ the expectation with respect to a probability measure μ on the space U containing the unknown function u. In the following we will finite-dimensionalize both the space U, in which the unknown function u lies, and the space containing the response of the forward model. The parameter J denotes the truncation level of the coefficient expansion (5) used for the unknown function, and the parameter l denotes the spatial finite element discretization level introduced in section 3. The parameters N and $\mathcal{L}$ denote the cardinality of the set of the chosen active gpc coefficients and the set of finite element discretization levels for these coefficients which is introduced in section 4. We employ multilevel FE discretizations on mesh levels $l = 0, 1, ..., L$ of the forward problem together with multilevel approximations of the Bayesian posterior which are indexed by $l' = 0, 1, \dots, L$, and combine these judiciously with *a discretization level dependent sample size $M_{ll'}$* within the MCMC method. The measure μ will variously be chosen as the prior ρ, the posterior ρ^δ, and various approximations of the posterior such as $\rho^{J,l,\delta}$.

We denote by $\mathcal{P}_{u^{(0)}}$, $\mathcal{P}_{u^{(0)}}^{J,l}$ and $\mathcal{P}_{u^{(0)}}^{N,\mathcal{L}}$ probability measures on the probability space generated by the MCMC processes detailed in the following, when conditioned on the initial point $u^{(0)}$ with the acceptance probability for the Metropolis-Hastings Markov chain being α defined in (9), $\alpha^{J,l}$ defined in (26), and $\alpha^{N,\mathcal{L}}$ defined in (31) respectively for the problems on the full, infinite dimensional space and its truncations. We then denote by $\mathcal{E}_{u^{(0)}}$, $\mathcal{E}_{u^{(0)}}^{J,l}$ and $\mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}$ expectation with respect to $\mathcal{P}_{u^{(0)}}$, $\mathcal{P}_{u^{(0)}}^{J,l}$ and $\mathcal{P}_{u^{(0)}}^{N,\mathcal{L}}$ respectively.

If the initial point $u^{(0)}$ of these Markov chains is distributed with respect to an initial probability measure μ on U, then we denote the probability measure on the probability space that describes these Markov chains by $\mathcal{P}^\mu$, $\mathcal{P}^{\mu,J,l}$ and $\mathcal{P}^{\mu,N,\mathcal{L}}$, and the corresponding expectation accordingly by $\mathcal{E}^\mu$, $\mathcal{E}^{\mu,J,l}$ and $\mathcal{E}^{\mu,N,\mathcal{L}}$. As these notations are somewhat involved, for the convenience of the reader we collect them in the following table.

Acceptance probability		α in (9)	$\alpha^{J,l}$ in (26)	$\alpha^{N,\mathcal{L}}$ in (31)
starting with $u^{(0)}$	Probability of the Markov chain	$\mathcal{P}_{u^{(0)}}$	$\mathcal{P}_{u^{(0)}}^{J,l}$	$\mathcal{P}_{u^{(0)}}^{N,\mathcal{L}}$
	Expectation	$\mathcal{E}_{u^{(0)}}$	$\mathcal{E}_{u^{(0)}}^{J,l}$	$\mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}$
starting with distribution μ	Probability of the Markov chain	$\mathcal{P}^\mu$	$\mathcal{P}^{\mu,J,l}$	$\mathcal{P}^{\mu,N,\mathcal{L}}$
	Expectation	$\mathcal{E}^\mu$	$\mathcal{E}^{\mu,J,l}$	$\mathcal{E}^{\mu,N,\mathcal{L}}$

Finally, in Section 5, we will work with the probability measure $\mathbf{P}_L$, on the probability space corresponding to a sequence of independent Markov chains created by the multilevel-MCMC procedure, and with $\mathbf{E}_L$, the expectation with respect to this probability measure. The definition of these measures will be given at the beginning of

Section 5.

2. Elliptic Inverse Problem and Approximations

In this section, we formulate the elliptic problem of interest in Section 2.1, formulate the resulting Bayesian inverse problem in Section 2.2, describe the independence sampler used to probe the posterior distribution in Section 2.3 and convergence rate estimates of the finite element and gpc methods in Sections 2.4 and 2.5, respectively. The results in this section all follow from direct application of well-known theories of inverse problems, MCMC and discretization techniques for PDEs; however, to render the exposition self-contained, proofs and references are provided in Appendices A–E.

2.1. Forward Problem

Let D be a bounded Lipschitz domain in $\mathbb{R}^d$. For $f \in L^2(D)$, we consider the elliptic problem

$$- \nabla \cdot (K(x)\nabla P(x)) = f(x) \text{ in } D, \quad P = 0 \text{ on } \partial D . \tag{4}$$

Throughout we assume that the domain D is a convex polyhedron with plane sides. The coefficient $K \in L^\infty(D)$ in (4) is parametrized by the series expansion

$$K(x) = \bar{K}(x) + \sum_{j \geq 1} u_j \psi_j(x), \quad x \in D \tag{5}$$

where the u_j are normalized to satisfy $\sup_j |u_j| \leq 1$ and where the summation may be either infinite or finite. In the next subsection we formulate the problem of determining the (uncertain) function K, or equivalently the sequence $\{u_j\}_{j \geq 1}$, from a finite noisy set of observations comprised of linear functionals of the solution P. Where it is notationally helpful to do so, we will write $K(x, u)$ and $P(x, u)$ for the coefficient and solution of (4) respectively.

The following sparsity assumptions on K in (5), which we will use in various combinations throughout the paper, imply the bounded invertibility of the parametric forward map $U \ni u \mapsto P(\cdot, u) \in V$. They also imply sparsity of gpc representations of this map and allow for controlling various approximation errors that arise in the sequel.

Assumption 1 *The functions $\bar{K}$ and ψ_j in (5) are in $L^\infty(D)$ and:*

(i) there exists a positive constant κ such that

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

where $\bar{K}_{\min} = \operatorname{essinf}_x \bar{K}(x) > 0$;

(ii) the functions $\bar{K}$ and ψ_j in (5) are in $W^{1,\infty}(D)$ and there exist positive constants C and q such that for all $J \in \mathbb{N}$, the sequence $\{\psi_j\}$ in the parametric representation

(5) of the uncertain coefficient in (4) satisfies

$$\sum_{j \geq 1} \|\psi_j\|_{W^{1,\infty}(D)} < \infty, \quad \sum_{j > J} \|\psi_j\|_{L^\infty(D)} < CJ^{-q};$$

(iii) the coefficients ψ_j are arranged in decreasing order of magnitude of $\|\psi_j\|_{L^\infty(D)}$ and there are constants $s > 1$ and $C > 0$ such that

$$\forall j \in \mathbb{N}: \quad \|\psi_j\|_{L^\infty(D)} \leq Cj^{-s};$$

(iv) for all $j \in \mathbb{N}$, $\psi_j \in W^{1,\infty}(D)$ and there exists a constant $C > 0$ such that

$$\forall j \in \mathbb{N}: \quad \|\nabla \psi_j\|_{L^\infty(D)} \leq Cj^{-s'} \quad \text{for some} \quad 1 < s' \leq s .$$

Sparsity of the unknown diffusion coefficient K, which is expressed in terms of the decay rate for the coefficients ψ_j of the expansion (5) in Assumption 1 (iii) and (iv) holds when the covariance of the random coefficient $K(x, \omega)$ is sufficiently smooth, as shown in Bieri et al [3]. With the decay rate in Assumption 1(iii), the constant q in (ii) can be chosen as $q = s - 1$.

We denote by $U = [-1, 1]^{\mathbb{N}}$ the set of all sequences $u = (u_j)_{j \geq 1}$ of coordinates u_j taking values in $[-1, 1]$ and note that this is the unit ball in $\ell^\infty(\mathbb{N})$. We equip the parameter domain U with the product sigma algebra $\Theta = \bigotimes_{j=1}^\infty \mathcal{B}([-1, 1])$. Due to Assumption 1(i), for any $u \in U$ the series (5) converges in $L^\infty(D)$. Furthermore, it also follows from this assumption that there exist finite positive constants $K_{\min}$ and $K_{\max}$ such that, for all $(x, u) \in D \times U$,

$$K_{\min} := \bar{K}_{\min}/(1+\kappa) \leq K(x) \leq K_{\max} := \text{esssup}_x \bar{K}(x) + \kappa \bar{K}_{\min}/(1+\kappa). \quad (6)$$

We let $V = H_0^1(D)$, whilst V^* denotes its dual space. We equip V with the norm $\|P\|_V = \|\nabla P\|_{L^2(D)}$. By (5), $K(x)$ is bounded below uniformly with respect to $(x, u) \in D \times U$ and, for every $u \in U$,

$$\begin{aligned} K_{min}\|P(\cdot, u)\|_V^2 &= K_{min}(\nabla P(\cdot, u), \nabla P(\cdot, u)) \leq (K(\cdot, u)\nabla P(\cdot, u), \nabla P(\cdot, u)) \\ &= (f, P(\cdot, u)) \leq \|f\|_{V^*}\|P(\cdot, u)\|_V , \end{aligned}$$

where $(\cdot, \cdot)$ denotes the inner product in $L^2(D)$ and $(L^2(D))^d$. It follows that

$$\sup_{u \in U} \|P(\cdot, u)\|_V \leq \frac{\|f\|_{V^*}}{K_{min}} . \quad (7)$$

In fact we have the following, proved in Appendix A.

Proposition 2 *Under Assumption 1(i), the solution $P : U \mapsto V = H_0^1(D)$ is Lipschitz when viewed as a mapping from the unit ball in $\ell^\infty(\mathbb{N})$ to V. It is in particular measurable, as a mapping from the measurable space (U, Θ) to $(V, \mathcal{B}(V))$.*

□

2.2. Bayesian Elliptic Inverse Problem

We now define the Bayesian inverse problem. For $\mathcal{O}_i \in V^*$, $i = 1, \ldots, k$, which denote k continuous, linear "observation" functionals on V, we define a map $\mathcal{G} : U \to \mathbb{R}^k$ as

$$U \ni u \mapsto \mathcal{G}(u) := (\mathcal{O}_1(P(\cdot, u)), \mathcal{O}_2(P(\cdot, u)), \ldots, \mathcal{O}_k(P(\cdot, u))) \in \mathbb{R}^k . \qquad (8)$$

In (1), by ϑ we denote an observational noise which is statistically independent of the input u and which we assume to be Gaussian, ie. with distribution $N(0, \Sigma)$ in $\mathbb{R}^k$, with positive definite covariance matrix Σ. We model the noisy observed data δ for $\mathcal{G}(u)$ by

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

For Bayesian inversion, we place a prior measure on u by assuming that $u_j : \Omega \to [-1, 1]$ comprises a sequence of independent random variables $u_j : \Omega \to [-1, 1]$ in the series expansion (5). On the measurable space (U, Θ) defined above we define the countable product probability measure

$$\rho = \bigotimes_{j \geq 1} \frac{du_j}{2},$$

where du_j is the Lebesgue measure on $[-1, 1]$. As u_j are uniformly distributed on $[-1, 1]$, the measure ρ is the law of the random vector $u = (u_1, u_2, \ldots)$ in U. As the random variables $u_j(\omega)$ were assumed independent, the probability measure on realizations of random vectors $u \in U$ is a product measure: for $S = \prod_{j \geq 1} S_j$,

$$\rho(S) = \prod_{j \geq 1} \mathbb{P}(\{\omega : u_j \in S_j\}) .$$

Combining the prior and likelihood, the posterior measure on u given δ can be explicitly determined, using the general theory from Appendix B, and a stability/well-posedness estimate; the proof of the following result for the model problem is provided there.

Proposition 3 *Let Assumption 1(i) hold. The conditional probability measure $\rho^\delta(du) = \mathbb{P}(du|\delta)$ on U is absolutely continuous with respect to $\rho(du)$ and satisfies*

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

Furthermore, for every $r > 0$ for every δ, δ' such that $|\delta|_\Sigma, |\delta'|_\Sigma \leq r$, there exists $C = C(r) > 0$ such that

$$d_{\mathrm{Hell}}(\rho^\delta, \rho^{\delta'}) \leq C(r)|\delta - \delta'|_\Sigma$$

where d_{Hell} denotes the Hellinger distance of the measures $\rho^\delta, \rho^{\delta'}$ (see, eg., equation (51) in Appendix B).

Remark 4 *The proof of the preceding proposition shows that $|\mathcal{G}(u)|$ is uniformly bounded for u in U. As a consequence there exists $\Phi^\star(r) > 0$ which is a uniform bound on $\Phi(u; \delta)$ for all $|\delta| \leq r$ and for all $u \in U$. This bound is also uniform with respect to truncation of the infinite series (5) for K, since this corresponds to a particular choice of some of the coefficients of $u \in U$, and with respect to finite element approximation*

of the solution of (4), since the uniform upper bound on $|\mathcal{G}(u)|$ will be preserved under Galerkin projections of the elliptic problem (4) into finite-dimensional subspaces $V_h \subset V$ (or, more generally, under any stable discretization of the forward problem of interest).

2.3. Independence Sampler

To approximate expectations with respect to the posterior measure ρ^δ constructed in the previous section we will use MCMC methods and, in particular, the independence sampler. To this end we define, for any $u, v \in U$,

$$\alpha(u, v) = 1 \wedge \exp(\Phi(u, \delta) - \Phi(v, \delta)) . \tag{9}$$

The Markov chain $\{u^{(k)}\}_{k=1}^\infty \subset U$ is then constructed as follows: given the current state $u^{(k)}$, we draw a proposal $v^{(k)}$ independently of $u^{(k)}$ from the prior measure ρ appearing in (2). Let $\{w^{(k)}\}_{k \geq 1}$ denote an i.i.d sequence with $w^{(1)} \sim \mathcal{U}[0, 1]$ and with $w^{(k)}$ independent of both $u^{(k)}$ and $v^{(k)}$. We then determine the next state $u^{(k+1)}$ via the formula

$$u^{(k+1)} = \mathbf{1}(\alpha(u^{(k)}, v^{(k)}) \geq w^{(k)})v^{(k)} + \left(1 - \mathbf{1}(\alpha(u^{(k)}, v^{(k)}) \geq w^{(k)})\right)u^{(k)} . \tag{10}$$

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 $1 - \alpha(u^{(k)}, v^{(k)})$. The following is proved in Appendix C.

Proposition 5 *Let Assumption 1(i) hold. Equation (10) defines a Markov chain $\{u^{(k)}\}_{k=0}^\infty$ which is reversible with respect to ρ^δ. If $p(u, A)$ denotes the transition kernel for the Markov chain, and $p^n(u, A)$ its nth iterate, then there is $R \in (0, 1)$ such that, for all $n \in \mathbb{N}$,*

$$\|p^n(u, A) - \rho^\delta\|_{\mathrm{TV}} \leq 2(1 - R)^n .$$

For every bounded, continuous function $g : U \to \mathbb{R}$, there holds, $\mathcal{P}_{u^{(0)}}$ almost surely,

$$\frac{1}{M} \sum_{k=1}^M g(u^{(k)}) = \mathbb{E}^{\rho^\delta}[g(u)] + c\xi_M M^{-\frac{1}{2}} \tag{11}$$

where ξ_M is a sequence of random variables which converges weakly as $M \to \infty$ to $\xi \sim N(0, 1)$ and where c is a positive constant which is independent of M (it depends only on $\Phi^\star(r)$ in Remark 4 and on $\sup_{u \in U} |g(u)|$). Furthermore, we have the mean square error bound: there exists $C > 0$ such that for every bounded $g : U \mapsto \mathbb{R}$ and every $M \in \mathbb{N}$

$$\left(\mathcal{E}^\rho\left[\left|\mathbb{E}^{\rho^\delta}[g(u)] - \frac{1}{M} \sum_{k=1}^M g(u^{(k)})\right|^2\right]\right)^{1/2} \leq C \sup_{u \in U} |g(u)| M^{-1/2} .$$

Remark 6 *The proof of this proposition uses fairly standard methods from the theory of Markov chains. In our context a key observation regarding the proof is that all constants – in particular the constant R – depend only on the upper bound on the negative log likelihood Φ, and on the supremum of g. Hence, if we can show for finite element and Karhúnen-Loève approximations of the forward map $\mathcal{G}$ in the physical domain D that*

these approximations are such that the potential Φ is stable under these approximations, then the conclusions of the preceding theorem will hold with constants that are uniformly bounded with respect to all approximation parameters.

Note also that our choice of uniform priors and the independence sampler means that the issue of "burn-in" does not play a significant role in our analysis; in particular the total variation metric convergence bound is independent of initialization. When generalizing our work to other priors and other MCMC methods "burn-in" effects may become more pronounced in the analysis.

2.4. Finite Element Approximation of the Forward Problem

Assumptions 1(i),(ii) are imposed throughout what follows regarding the finite element method. Thus, from (5), we deduce that $K(\cdot, u) \in W^{1,\infty}(D)$ for all $u \in U$. We describe an approximation of the forward problem based on finite element representation of the solution P of (4), together with truncation of the series (5). We start by discussing the finite element approximation. Recalling that the domain D is a bounded Lipschitz polyhedron with plane sides, we denote by $\{\mathcal{T}^l\}_{l=1}^{\infty}$ the nested sequence of simplices which is defined inductively as follows: first we subdivide D into a regular family $\mathcal{T}^0$ of simplices T; for $l \geq 1$, the regular simplicial mesh $\mathcal{T}^l$ in D is obtained by recursive subdivision, i.e. each simplex in $\mathcal{T}^{l-1}$ is divided into 2^d subsimplices. Then, the meshwidth $h_l := \max\{\text{diam}(T) : T \in \mathcal{T}^l\}$ of $\mathcal{T}^l$ is $h_l = 2^{-l}h_0$. Based on these triangulations, we define a nested multilevel family of spaces of continuous, piecewise linear functions on $\mathcal{T}^l$ as

$$V^l = \{u \in V : u|_T \in \mathcal{P}_1(T) \ \forall T \in \mathcal{T}^l\},$$

where $\mathcal{P}_1(T)$ denotes the set of linear polynomials in the simplex $T \in \mathcal{T}^l$. Approximating the solution of the parametric, deterministic problem (4) from the finite element spaces V^l introduces a *discretization error* which is well-known to be bounded by the *approximation property* of the V^l: there exists a positive constant $C > 0$ which is independent of l such that for all $P \in H^2(D) \cap H_0^1(D)$ and for every $0 < h_l \leq 1$ holds

$$\inf_{Q \in V^l} \|P - Q\|_V \leq Ch_l\|P\|_{H^2(D)}, \tag{12}$$

where $h_l = O(2^{-l})$ and where the constant $C > 0$ depends only on $\mathcal{T}^0$.

We now discuss the effect of dimensional truncation, ie. of truncating the infinite series (5) for the unknown diffusion coefficient K in problem (4) after J terms, as

$$K^J(x, u) = \bar{K}(x) + \sum_{j=1}^{J} u_j \psi_j(x) \, x \in D, \, u \in U \, . \tag{13}$$

We thus consider the parametric, deterministic family of approximate elliptic problems

$$-\nabla \cdot (K^J(\cdot, u)\nabla P^J(\cdot, u)) = f, \quad P^J = 0 \text{ on } \partial D \, . \tag{14}$$

Under Assumptions 1(i) and (ii), (49) shows that there exists a constant $C > 0$ such that for all $J \in \mathbb{N}$ and all $u \in U$

$$
\begin{aligned}
\|P(\cdot, u) - P^J(\cdot, u)\|_V &\leq C\|P(\cdot, u)\|_V \|K(\cdot, u) - K^J(\cdot, u)\|_{L^\infty(D)} \\
&\leq C\|P(\cdot, u)\|_V J^{-q} \leq \frac{C}{K_{min}} J^{-q}\|f\|_{V^*} .
\end{aligned}
\tag{15}
$$

We consider the finite element approximation of the truncated problem (14): given $J, l \in \mathbb{N}$, find $P^{J,l}(\cdot, u) \in V^l$ such that for all $\phi \in V^l$

$$
\int_D K^J(x, u)\nabla P^{J,l}(x, u) \cdot \nabla \phi(x)dx = \int_D f(x)\phi(x)dx .
\tag{16}
$$

We make the following assumption on the complexity of solving this discrete equation the justification of which is provided in Appendix D.

Assumption 7 *For $J \in \mathbb{N}$ as in (13), the solution $P^J(\cdot, u)$ of (14) is uniformly bounded in $W := H^2(D) \cap H_0^1(D)$ with respect to $J \in \mathbb{N}$ and to $u \in U$. The matrix of the Galerkin approximated problem (16) has $O(l^{d-1}2^{dl})$ non-zero entries and has a uniformly bounded condition number for all J, l and u. There exists $C > 0$ such that, for all $J \in \mathbb{N}$ and for all $u \in U$, the FE error for the approximating problem (16) satisfies*

$$
\|P^J(\cdot, u) - P^{J,l}(\cdot, u)\|_V \leq C2^{-l}\|P^J(\cdot, u)\|_W .
\tag{17}
$$

From this assumption, we obtain the following error bound.

Proposition 8 *Consider the approximation of the elliptic problem (4) via finite element solution of the truncated problem (14), under Assumptions 1 (i), (ii) and Assumption 7. Then there exists a constant $C > 0$ such that for every $J, l \in \mathbb{N}$ and for every $u \in U$ it holds that the Finite Element solutions $P^{J,l}(\cdot, u)$ are uniformly V-stable in the sense that, for every $J, l \in \mathbb{N}$ and for every $u \in U$ there holds*

$$
\sup_{J,l\in\mathbb{N}} \sup_{u\in U} \|P^{J,l}(\cdot, u)\|_V \leq \frac{C}{K_{min}}\|f\|_{V^*} .
\tag{18}
$$

Moreover, there exists $C > 0$ such that for every $u \in U$ holds the error bound

$$
\|P(\cdot, u) - P^{J,l}(\cdot, u)\|_V \leq C(2^{-l}\|P^J(\cdot, u)\|_W + J^{-q}\|P(\cdot, u)\|_V) .
\tag{19}
$$

2.5. Sparse Tensor gpc-Finite Element Surrogate of the Parametric Forward Problem

By (7), the solution $P(\cdot, u)$ of problem (4) is uniformly bounded in V by $\|f\|_{V^*}/K_{min}$. Therefore, from Proposition 2, we deduce that $P(\cdot, \cdot) \in L^2(U, \rho; V)$. Thus the parametric solution admits a polynomial chaos type representation in $L^2(U, \rho; V)$. To define it, we denote by $L_n(u_n)$ the Legendre polynomial of degree n, normalized such that

$$
\frac{1}{2}\int_{-1}^1 |L_n(\xi)|^2 d\xi = 1 .
$$

By $\mathcal{F}$ we denote the set of all sequences $\nu = (\nu_1, \nu_2, \ldots)$ of nonnegative integers such that only finitely many ν_j are nonzero. We define

$$
L_\nu(u) = \prod_{j\geq 1} L_{\nu_j}(u_j) .
\tag{20}
$$

Since $L_0 \equiv 1$, for each $\nu \in \mathcal{F}$ the product contains only finitely many nontrivial factors. The set $\{L_\nu : \nu \in \mathcal{F}\}$ forms an orthonormal basis for $L^2(U, \rho)$. We can therefore expand $P(\cdot, u)$ into the Legendre expansion

$$P(\cdot, u) = \sum_{\nu \in \mathcal{F}} P_\nu(\cdot) L_\nu(u),$$

where $P_\nu := \int_U P(\cdot, u) L_\nu(u) d\rho(u) \in V$. By the $L^2(U, \rho)$ orthonormality of the set $\{L_\nu : \nu \in \mathcal{F}\}$, Parseval's equation in the Bochner space $L^2(U, \rho; V)$ takes the form

$$\forall P \in L^2(U, \rho; V) : \quad \|P\|^2_{L^2(U,\rho;V)} = \sum_{\nu \in \mathcal{F}} \|P_\nu\|^2_V .$$

We now define the sparse tensor surrogate forward maps where we only use a finite (sparse) subset of the gpc mode index set $\mathcal{F}$. For any integer N, choose an index set $\Lambda_N \subset \mathcal{F}$ of gpc coefficients $P_\nu \in V$ which are to be included into the surrogate map, subject to the constraint $\#(\Lambda_N) \leq N$, and a second set $\mathcal{L}(\Lambda_N) := (l_\nu)_{\nu \in \Lambda_N} \subset \mathbb{N}^N$ of Finite-Element discretization levels for the FE approximation of the active P_ν, $\nu \in \Lambda_N$. Then we consider surrogate forward maps $P_{N,\mathcal{L}}$ for the response P which are of the form

$$P_{N,\mathcal{L}}(x, u) = \sum_{\nu \in \Lambda_N} P_{\nu,\mathcal{L}}(x) L_\nu(u), \quad P_{\nu,\mathcal{L}} \in V^{l_\nu} . \tag{21}$$

We wish to find the sets Λ_N and $\mathcal{L}(\Lambda_N)$ that give the best (or the quasi-best) approximations among all finite subsets $\Lambda \subset \mathcal{F}$ and of $\mathcal{L}(\Lambda)$ subject to a constraint on their *combined* cardinality $N_{dof} = \sum_{\nu \in \Lambda_N} l_\nu$. Convergence rates, in terms of N_{dof}, of such approximations of parametric forward maps have been derived recently, for example, in [5], [6], [15], [16], [27], [25]. In these references, the following assumption has been verified for various problem classes.

Assumption 9 *There are positive constants σ, τ, α and β such that for each integer N, with a total budget of $N_{\text{dof}} = O(N^{\sigma/\tau})$ degrees of freedom, a subset $\Lambda_N \subset \mathcal{F}$ of cardinality not exceeding an absolute multiple of N and such that, for every $\nu \in \Lambda_N$, $|\nu| = O(\log N)$ and a surrogate gpc FE approximation $P_{N,\mathcal{L}}$ of the parametric forward solution P of (14) of the form (21) with rate of convergence*

$$\|P - P_{N,\mathcal{L}}\|_{L^2(U,\rho;V)} \leq C N_{dof}^{-\tau},$$

can be found in $O(N_{dof}^\alpha (\log N_{dof})^\beta)$ float point operations.

For the model elliptic inverse problem of determining K given by (5) from linear functionals defined on solutions of (4), subject to the sparsity conditions in Assumption 1, the preceding assumption is verified in Appendix E; there also bounds for τ, the rate σ and for the exponents $\alpha > 0$ and $\beta \geq 0$ in Assumption 9 are derived, in terms of the sparsity assumptions on the unknown coefficient K in (13).

3. Plain MCMC

We study the computational complexity of the Plain MCMC method (10) to generate samples from the posterior probability measure ρ^δ determined in the previous section.

The main results of this analysis are summarized in Theorem 13. The complexity analysis of the MCMC algorithm is of independent interest as it utilizes the emerging idea of MCMC methods with dimension independent mixing properties [7, 14, 31]. Furthermore, the results in the present section will be the foundation for several accelerations of the Plain MCMC algorithm which will be presented in Sections 4 and 5 ahead. In order to obtain a constructive version of the MCMC algorithm, we will approximate solutions of the forward problem (4) by applying the Finite Element Method in the physical domain D to its parametric version (4) and by truncation of the expansion of the diffusion coefficient K given by (5).

Assumptions 1(i),(ii) are imposed throughout what follows. From Assumption 1(ii) and from (5), we deduce that $K(\cdot, u) \in W^{1,\infty}(D)$ for all $u \in U$.

3.1. FE Approximation of the Posterior Measure

We denote the vector of observables from the Galerkin discretized, gpc-parametrized approximate forward solution map by

$$\mathcal{G}^{J,l}(u) = (\mathcal{O}_1(P^{J,l}(u)), \ldots, \mathcal{O}_k(P^{J,l}(u))) : U \mapsto \mathbb{R}^k \qquad (22)$$

and define the corresponding Bayesian potential

$$\Phi^{J,l}(u; \delta) = \frac{1}{2}|\delta - \mathcal{G}^{J,l}(u)|_\Sigma^2 . \qquad (23)$$

We define an approximate conditional posterior probability measure $\rho^{J,l,\delta}$ on the measurable space (U, Θ) as

$$\frac{d\rho^{J,l,\delta}}{d\rho} \propto \exp(-\Phi^{J,l}(u; \delta)) .$$

Then, the measure $\rho^{J,l,\delta}$ is an approximation of the Bayesian posterior ρ^δ which, due to the discretization and the truncation of the expansion (5) incurs an approximation error. We now show that this error in the posterior measure is bounded in the Hellinger metric with respect to J and l in the same way as the forward error in Proposition 8. The proof of the following Proposition is based on a generalization of the method introduced in [8].

Proposition 10 *Let Assumptions 1(i),(ii) and 7 hold. If the domain D is convex and if $f \in L^2(D)$ then there exists a positive constant c depending only on the data δ such that for every J and l there holds*

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

Proof We denote the normalizing constants as

$$Z(\delta) = \int_U \exp(-\Phi(u; \delta))d\rho(u), \quad Z^{J,l}(\delta) = \int_U \exp(-\Phi^{J,l}(u; \delta))d\rho(u) .$$

We then estimate

$$2d_{\text{Hell}}(\rho^\delta, \rho^{J,l,\delta})^2$$
$$= \int_U \left(Z(\delta)^{-1/2} \exp\left(-\frac{1}{2}\Phi(u;\delta)\right) - (Z^{J,l}(\delta))^{-1/2} \exp\left(-\frac{1}{2}\Phi^{J,l}(u;\delta)\right) \right)^2 d\rho(u)$$
$$\leq I_1 + I_2,$$

where we defined

$$I_1 := \frac{2}{Z(\delta)} \int_U \left(\exp\left(-\frac{1}{2}\Phi(u;\delta)\right) - \exp\left(-\frac{1}{2}\Phi^{J,l}(u;\delta)\right) \right)^2 d\rho(u),$$

$$I_2 := 2|Z(\delta)^{-1/2} - Z^{J,l}(\delta)^{-1/2}|^2 \int_U \exp(-\Phi^{J,l}(u;\delta)) d\rho(u) .$$

We estimate I_1 and I_2. To bound I_1, given data δ, for every $u \in U$ holds

$$\left| \exp\left(-\frac{1}{2}\Phi(u;\delta)\right) - \exp\left(-\frac{1}{2}\Phi^{J,l}(u;\delta)\right) \right| \leq \frac{1}{2}|\Phi(u;\delta) - \Phi^{J,l}(u;\delta)|$$
$$\leq C(2|\delta| + |\mathcal{G}(u)| + |\mathcal{G}^{J,l}(u)|)|\mathcal{G}(u) - \mathcal{G}^{J,l}(u)| . \tag{24}$$

Moreover, by Proposition 8, there exists a constant $C > 0$ independent of J and of l such that, for all $u \in U$, there holds

$$|\mathcal{G}(u) - \mathcal{G}^{J,l}(u)| \leq C \max\{\|\mathcal{O}_i\|_{V^*}\}\|P(\cdot, u) - P^{J,l}(\cdot, u)\|_V$$
$$\leq C(2^{-l}\|P^J(\cdot, u)\|_W + J^{-q}\|P(\cdot, u)\|_V) .$$

By (7) and Proposition 29, $\|P(\cdot, u)\|_V$ and $\|P^J(\cdot, u)\|_W$ are uniformly bounded with respect to $u \in U$. Therefore, there exists a constant $C(\delta) > 0$ depending only on the data δ such that for all $J \in \mathbb{N}$

$$I_1 \leq C(\delta)\mathbb{E}^\rho(2^{-l}\|P^J(\cdot, u)\|_W + J^{-q}\|P(\cdot, u)\|_V)^2$$
$$\leq C(\delta)(J^{-2q}\|f\|_{V^*}^2 + 2^{-2l}\|f\|_{L^2(D)}^2) .$$

To estimate term I_2, we observe that there is a positive constant $C > 0$ such that for every $J, l \in \mathbb{N}$ holds

$$|Z(\delta)^{-1/2} - Z^{J,l}(\delta)^{-1/2}|^2 \leq C(Z(\delta)^{-3} \vee Z^{J,l}(\delta)^{-3})|Z(\delta) - Z^{J,l}(\delta)|^2 .$$

We note that

$$|Z(\delta) - Z^{J,l}(\delta)| \leq \int_U |\exp(-\Phi(u;\delta)) - \exp(-\Phi^{J,l}(u;\delta))| d\rho(u)$$
$$\leq \int_U |\Phi(u;\delta) - \Phi^{J,l}(u;\delta)| d\rho(u) .$$

Therefore, as $Z(\delta)$ and $Z^{J,l}(\delta)$ are uniformly bounded below for all δ, analysis similar to that for I_1 shows that

$$I_2 \leq C(\delta)(2^{-2l} + J^{-2q})\|f\|_{L^2(D)}^2.$$

Thus

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

$\square$

3.2. Computational Complexity of Plain MCMC

Given $J, l \in \mathbb{N}$, and data δ, we use the MCMC method (10) to sample the probability measure $\rho^{J,l,\delta}$. In so doing we create a method for approximating integrals of functions $g : U \to \mathbb{R}$ with respect to ρ^δ. We use the following notation for the empirical measure generated by the Markov chain designed to sample $\rho^{J,l,\delta}$:

$$E_M^{\rho^{J,l,\delta}}[g] := \frac{1}{M} \sum_{k=1}^{M} g(u^{(k)}), \tag{25}$$

where the Markov chain $\mathcal{C}_{J,l} = \{u^{(k)}\}_{k \in \mathbb{N}_0} \subset \mathbb{R}^J$ is started in the restriction of $u^{(0)}$ to $\mathbb{R}^J$. It depends on the discretization level l and the truncation level J since it is generated from the process (10) with the acceptance probability (9) being replaced by

$$\alpha^{J,l}(u,v) = 1 \wedge \exp(\Phi^{J,l}(u; \delta) - \Phi^{J,l}(v; \delta)), \quad (u,v) \in U \times U . \tag{26}$$

Given $M \in \mathbb{N}$ we wish to estimate the MC sampling error

$$\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{J,l,\delta}}[g] . \tag{27}$$

We develop in the following two types of error bounds as $M \to \infty$ for (27): a probabilistic error bound for $\mathcal{P}_{u^{(0)}}^{J,l}$ almost every realization of the Markov chain and a mean square bound. We refer to the table in Section 1.2 for the notation in the next result.

Proposition 11 *Let Assumptions 1(i),(ii) and 7 hold. Let $g : U \to \mathbb{R}$ be a bounded continuous function on U with respect to the supremum norm. Then, for every initial condition $u^{(0)}$ and for $\mathcal{P}_{u^{(0)}}^{J,l}$-almost every realization of the Markov chain holds the error bound*

$$\left| \mathbb{E}^{\rho^\delta} g(u) - E_M^{\rho^{J,l,\delta}}[g] \right| \le c_1 M^{-\frac{1}{2}} + c_2(J^{-q} + 2^{-l})$$

where $c_1 \le c_3|\xi_M|$; ξ_M is a random variable (on the probability space generating the randomness within the Markov chain) which converges weakly as $M \to \infty$ to $\xi \sim N(0,1)$ and c_2, c_3 are non-random constants independent of M, J and l.

Moreover, there exists a constant c_4 (which is deterministic and depends only on the data δ, and which is, in particular, independent of M, J and l) such that

$$\left(\mathcal{E}^{\rho,J,l} \left[\left| \mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{J,l,\delta}}[g] \right|^2 \right] \right)^{1/2} \le c_4(M^{-1/2} + J^{-q} + 2^{-l}). \tag{28}$$

Proof As g is bounded, we have from Proposition 10 and properties of the Hellinger metric (specifically, from (2.7) in [8]) for every $u \in U$ that

$$|\mathbb{E}^{\rho^\delta} g(u) - \mathbb{E}^{\rho^{J,l,\delta}} g(u)| \le \bar{c}(g) d_{\text{Hell}}(\rho^\delta, \rho^{J,l,\delta}) \le \bar{c}(g) C(\delta)(J^{-q} + 2^{-l}) . \tag{29}$$

Here, $C(\delta)$ is as in Proposition 10 and $\bar{c}(g)$ depends on the supremum of $g(u)$ over U, but is independent of J, l. By Proposition 5 (and Remarks 6 and 4) we deduce the existence of a constant $C > 0$, independent of M, J and l, such that there holds, as $M \to \infty$, $\mathcal{P}_{u^{(0)}}^{J,l}$ almost surely,

$$|\mathbb{E}^{\rho^{J,l,\delta}} g - \frac{1}{M} \sum_{k=1}^{M} g(u^{(k)})| \le C|\xi_M| M^{-1/2}$$

where ξ_M converges weakly as $M \to \infty$ to $\xi \sim N(0,1)$. Combining this with (29) gives the first assertion.

To prove the mean square error bound (28), we define

$$\bar{g}(u) := g(u) - \mathbb{E}^{\rho^{J,l,\delta}}[g],$$

as in the proof of Proposition 5. We note that $\bar{g}$ depends on J, l via dependence on the approximate posterior measure $\rho^{J,l,\delta}$, but we do not indicate this dependence explicitly. However we will use uniform boundedness of $\bar{g}$ with respect to these parameters in an essential way in what follows.

Due to the invariance of the stationary measures $\rho^{J,l,\delta}$, we may write

$$\frac{1}{M}\mathcal{E}^{\rho^{J,l,\delta},J,l}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] = \mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=1}^{M}\sum_{j=k+1}^{M}\mathcal{E}^{\rho^{J,l,\delta},J,l}[\bar{g}(u^{(k)})\bar{g}(u^j)]$$

$$= \mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathcal{E}^{\rho^{J,l,\delta},J,l}[\bar{g}(u^{(0)})\bar{g}(u^{(j)})]$$

$$= \mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})\mathcal{E}^{J,l}_{u^{(0)}}[\bar{g}(u^{(j)})]]$$

$$\leq \mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2]$$

$$+ 2\frac{1}{M}\sum_{k=0}^{M-1}\sup|\bar{g}|\sum_{j=1}^{M-k}\mathbb{E}^{\rho^{J,l,\delta}}[|\mathcal{E}^{J,l}_{u^{(0)}}g(u^{(j)}) - \mathbb{E}^{\rho^{J,l,\delta}}[g]|]$$

$$\leq \mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2] + 4\frac{1}{M}\sum_{k=0}^{M-1}\sup|\bar{g}|^2\sum_{j=1}^{M-k}(1-R)^j.$$

In the last line we have used the estimate on the total variation contraction from Proposition 5 noting, as in Remark 6, that because $\sup_{u \in U}\|P^{J,l}(u)\|_V$ is bounded uniformly with respect to the (discretization) parameters J and l, the constant $0 < R < 1$ is independent of the parameters J and l. Since $\sup_{J,l}\mathbb{E}^{\rho^{J,l,\delta}}[\bar{g}(u^{(0)})^2]$ is bounded independently of J and of l, we deduce that

$$\sup_{J,l,M \in \mathbb{N}} M\,\mathcal{E}^{\rho^{J,l,\delta},J,l}\Big[\Big|\frac{1}{M}\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] < \infty\,.$$

Next, we note that

$$\mathcal{E}^{\rho,J,l}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] = \int_U \mathcal{E}^{J,l}_{u^{(0)}}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]d\rho(u^{(0)})$$

$$= \int_U \mathcal{E}^{J,l}_{u^{(0)}}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]\frac{d\rho}{d\rho^{J,l,\delta}}(u^{(0)})d\rho^{J,l,\delta}(u^{(0)})$$

$$\leq \mathcal{E}^{\rho^{J,l,\delta},J,l}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]Z^{J,l}(\delta)\sup_{u \in U}\exp(\Phi^{J,l}(u;\delta))\,.$$

As $Z^{J,l}(\delta) \leq 1$ and $\sup_{u \in U} |\Phi^{J,l}(u; \delta)|$ are bounded uniformly with respect to J and l, we get the conclusion after using the bound from Proposition 10 on the Hellinger distance between ρ^δ and $\rho^{J,l,\delta}$. $\qquad\square$

We consider the case where $g(u) = \ell(P(\cdot, u))$ with $\ell \in V^*$ being a bounded linear functional on V. As there exists a constant $C > 0$ such that for all $J, l \in \mathbb{N}$, $u \in U$, there holds

$$|\mathbb{E}^{\rho^\delta}[\ell(P(\cdot, u))] - \mathbb{E}^{\rho^\delta}[\ell(P^{J,l}(\cdot, u))]| \leq C(J^{-q} + 2^{-l})$$

and

$$|\mathbb{E}^{\rho^\delta}[\ell(P^{J,l}(\cdot, u))] - \mathbb{E}^{\rho^{J,l,\delta}}[\ell(P^{J,l}(\cdot, u))]| \leq C(J^{-q} + 2^{-l}),$$

we have

$$|\mathbb{E}^{\rho^\delta}[\ell(P(\cdot, u))] - \mathbb{E}^{\rho^{J,l,\delta}}[\ell(P^{J,l}(\cdot, u))]| \leq C(J^{-q} + 2^{-l}) .$$

We therefore perform an MCMC algorithm to approximate $\mathbb{E}^{\rho^{J,l,\delta}}[\ell(P^{J,l}(\cdot, u))]$. As $\ell(P^{J,l}(\cdot, u))$ and $\Phi^{J,l}(u; \delta)$ depend only on the finite set of coordinates $\{u_1, \ldots, u_J\}$ in expansion (5), we perform the Metropolis-Hastings MCMC method on this set with proposals being drawn from the restriction of the prior measure ρ to this finite set.

Proposition 12 *Let Assumptions 1 (i), (ii) and 7 hold, and assume that $g(u) = \ell(P(\cdot, u))$ where ℓ is a bounded linear functional in V^*. Then the approximate evaluation of the sample average $\frac{1}{M} \sum_{k=1}^{M} \ell(P^{J,l}(\cdot, u^{(k)}))$ by the Plain MCMC method with M realizations of the chain, with Finite Element discretization in the domain D at mesh level l as described above, and with J-term truncated coefficient representation (13), requires $\mathcal{O}(l^{d-1}2^{dl}MJ)$ floating point operations.*

Proof The approximate computation of each of the $O(l^{d-1}2^{dl})$ non-zero entries of the stiffness matrix of (16), requires at most $O(J)$ operations to compute the coefficients K^J at the quadrature points. ‖ Therefore the cost of constructing the stiffness matrix is $O(l^{d-1}2^{dl}J)$. Since the condition number of this stiffness matrix is assumed to be uniformly bounded, the conjugate gradient method for the approximate solution of the linear system resulting from the Finite Element discretization with an accuracy comparable to the order of the discretization error requires at most $O(l^{d-1}2^{dl})$ floating point operations. The total cost for solving the approximated forward problem at each step of the Markov chain requires at most $O(l^{d-1}2^{dl}J)$ floating point operations. The numerical evaluation of $\ell(P^{J,l}(u^{(k)}))$ requires at most $O(2^{dl})$ floating point operations. Since we generate M draws of the chain, the assertion follows. $\qquad\square$

Theorem 13 *Let Assumptions 1 (i), (ii) and 7 hold. For $g(u) = \ell(P(u))$ where ℓ is a bounded linear functional in V^*, with probability $p_{N_{\text{dof}}}(t)$ the conditional expectation $\mathbb{E}^{\rho^\delta}g(u)$ can be approximated using N_{dof} degrees of freedom to approximate the forward*

‖ There is an implicit assumption here that the basis functions can be evaluated at a point with $\mathcal{O}(1)$ cost.

PDE and $t^2 N_{\mathrm{dof}}^{2/d}$ MCMC steps (with a total of $t^2 N_{\mathrm{dof}}^{1+2/d}$ degrees of freedom), incurring an error of $O(N_{\mathrm{dof}}^{-1/d})$, and using at most

$$ct^2 \log(N_{\mathrm{dof}})^{d-1} N_{\mathrm{dof}}^{1+(2+1/q)/d}$$

floating point operations for a positive constant c where, for all t,

$$\lim_{N_{\mathrm{dof}} \to \infty} p_{N_{\mathrm{dof}}}(t) \to \int_{-c't}^{c't} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx \ ,$$

for a positive constant c' independent of N_{dof} and t.

In mean square with respect to the measure $\mathcal{P}^{\rho,J,l}$, $\mathbb{E}^{\rho^\delta}[g(u)]$ can be approximated with an error $O(N_{dof}^{-1/d})$, using at most $N_{\mathrm{dof}}^{1+2/d}$ number of degrees of freedom in total, and $O(\log(N_{\mathrm{dof}})^{d-1} N_{\mathrm{dof}}^{1+(2+1/q)/d})$ floating point operations.

Proof We first prove the probabilistic convergence result. We invoke the error estimate in Proposition 11, and choose the parameters M, J and l so as to balance the bounds $M^{-1/2}$, J^{-q} and 2^{-l}, taking into account the fact that the coefficient of $M^{-\frac{1}{2}}$ is only known through its asymptotic normality. We select $J = 2^{l/q}$ and $M = t^2 N_{dof}^{2/d}$ where $t = c_3|\xi_M|$, with N_{dof} denoting the number of degrees of freedom at each step being $N_{dof} = O(2^{dl})$; the constant c_3 and the random variable ξ_M is as in Proposition 11. Then the total number of floating point operations required as $l \to \infty$ is not larger than $O(t^2 l^{d-1} 2^{(d+2+1/q)l})$. We then arrive at the conclusion. The mean square convergence is proved in a similar manner. $\square$

4. Sparse gpc-MCMC

We again study computational complexity of the MCMC method defined by (10) to sample the posterior probability distribution ρ^δ. We adapt the approach in the previous section to use a computational method which effects a reduction in computational cost by precomputing the parametric dependence of the forward model, which enters the likelihood. The main results are summarized in Theorem 18. This method is introduced, and used in practice, in the series of papers [21, 19, 20]. The major cost in MCMC methods is the repeated solution of the forward equations, with varying coefficients from the MCMC sampler of ρ^δ. The complexity of these repeated forward solves can be drastically reduced by precomputing an approximate, deterministic parametric representation of the system's response which is valid for *all* possible realizations of $u \in U$. Specifically, we precompute a sparse tensor finite element approximation of the parametric, deterministic forward problem by an approximate polynomial chaos representation of the solution's dependence on u and by discretization of the forward solutions' spatial dependence from a multilevel hierarchy of Finite Element spaces in D. This precomputed "surrogate" of the parametric response map is then evaluated when running M steps of the chain to estimate expectations with respect to the (approximate due to the use of the surrogate) posterior measure.

As we shall show, this strategy is particularly effective if the observations consist only of continuous, linear functionals of the system's response. In this case, only scalar coefficients of the forward map's gpc expansion need to be stored and evaluated. We use this to reduce the cost per step of the MCMC method. We continue to work under Assumption 1(i),(ii) and, furthermore, we add Assumption 1(iii) throughout the remainder of the paper.

4.1. Approximation of the Posterior Measure

For the solution $P_{N,\mathcal{L}}$ in Assumption 9, we define the parametric, deterministic forward map based on the N-term truncated gpc expansion and Finite Element projected surrogate forward map, as specified in Section 2.5

$$\mathcal{G}^{N,\mathcal{L}}(u) = (\mathcal{O}_1(P_{N,\mathcal{L}}(\cdot, u)), \ldots, \mathcal{O}_k(P_{N,\mathcal{L}}(\cdot, u))), \tag{30}$$

and the corresponding approximate Bayesian potential

$$\Phi^{N,\mathcal{L}}(u; \delta) = \frac{1}{2}|\delta - \mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma^2 .$$

The conditional measure $\rho^{N,\mathcal{L},\delta}$ on the measurable space (U, Θ) is defined as

$$\frac{d\rho^{N,\mathcal{L},\delta}}{d\rho} \propto \exp(-\Phi^{N,\mathcal{L}}(u; \delta)) .$$

We then have the following approximation result.

Proposition 14 *Let Assumptions 1(i) and 9 hold. Then there is a constant $C(\delta)$ which only depends on the data δ such that, for every N and $\mathcal{L}$ as in Assumption 9,*

$$d_{\text{Hell}}(\rho^\delta, \rho^{N,\mathcal{L},\delta}) \le C(\delta)N_{\text{dof}}^{-\tau} .$$

Proof The proof for this proposition is similar to the proof of Proposition 10, differing only in a few details; hence we highlight only the differences. These are due to estimates on the forward error from Assumptions 9 being valid only in the mean square sense whilst Proposition 8 holds pointwise for $u \in U$. Nonetheless, at the point in the estimation of I_1 and I_2 where the forward error estimate is used, it is possible to use a mean square forward error estimate instead of a pointwise forward error estimate. From Assumption 9, we deduce that there is a positive constant c such that:

$$\rho\{u :\ |\mathcal{G}(u) - \mathcal{G}^{N,\mathcal{L}}(u)| > 1\} \le cN_{dof}^{-2\tau} .$$

As $\|P(u)\|_V$ is uniformly bounded for all u, there is a constant $c_1(\delta) > 0$ such that $|\delta - \mathcal{G}(u)|_\Sigma < c_1(\delta)$. Choose a constant $c_2(\delta) > 0$ sufficiently large. If $|\delta - \mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma > c_2(\delta)$, then

$$|\mathcal{G}^{N,\mathcal{L}}(u) - \mathcal{G}(u)|_\Sigma \ge |\delta - \mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma - |\delta - \mathcal{G}(u)|_\Sigma > c_2(\delta) - c_1(\delta) > 1 .$$

Let $U_1 \subset U$ be the set of $u \in U$ such that $|\delta - \mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma > c_2(\delta)$. We have that $\rho(U_1) \le cN_{dof}^{-2\tau}$. Thus,

$$\frac{1}{Z(\delta)} \int_{U_1} \left| \exp\left(-\frac{1}{2}\Phi(u; \delta)\right) - \exp\left(-\frac{1}{2}\Phi^{N,\mathcal{L}}(u; \delta)\right) \right| d\rho(u) \le C(\delta)N_{dof}^{-2\tau} .$$

When $u \notin U_1$, $|\delta - \mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma \leq c_2(\delta)$ so there is a constant $c_3(\delta)$ so that $|\mathcal{G}^{N,\mathcal{L}}(u)| \leq c_3(\delta)$. An argument similar to that for (24) shows that

$$\left| \exp\left(-\frac{1}{2}\Phi(u;\delta)\right) - \exp\left(-\frac{1}{2}\Phi^{N,\mathcal{L}}(u;\delta)\right) \right| \leq$$
$$C(2|\delta| + |\mathcal{G}(u)| + |\mathcal{G}^{N,\mathcal{L}}(u)|)|\mathcal{G}(u) - \mathcal{G}^{N,\mathcal{L}}(u)|.$$

Therefore

$$I_1 = \frac{1}{Z(\delta)}\int_U \left| \exp\left(-\frac{1}{2}\Phi(u;\delta)\right) - \exp\left(-\frac{1}{2}\Phi^{N,\mathcal{L}}(u;\delta)\right) \right|^2 d\rho(u)$$
$$\leq C(\delta)N_{dof}^{-2\tau} +$$
$$c\int_U (2|\delta| + |\mathcal{G}(u)| + c_3(\delta))^2|\mathcal{G}(u) - \mathcal{G}^{N,\mathcal{L}}(u)|^2 d\rho(u)$$
$$\leq C(\delta)N_{dof}^{-2\tau} + C(\delta)\int_U \|P(\cdot,u) - P_{N,\mathcal{L}}(\cdot,u)\|_V^2 d\rho(u)$$
$$\leq C(\delta)N_{dof}^{-2\tau} .$$

To show that $I_2 < C(\delta)N_{dof}^{-2\tau}$ we still need to verify that

$$Z^{N,\mathcal{L}}(\delta) = \int_U \exp(-\Phi^{N,\mathcal{L}}(u;\delta))d\rho(u)$$

is uniformly bounded from below by a positive bound for all N and $\mathcal{L}$. As $P_{N,\mathcal{L}}$ is uniformly bounded in $L^2(U,\rho;V)$,

$$\int_U |\mathcal{G}^{N,\mathcal{L}}(u)|d\rho(u) \leq c\int_U \|P_{N,\mathcal{L}}(u)\|_V d\rho(u) \leq c.$$

Fixing $r > 0$ sufficiently large, the ρ measure of the set $u \in U$ such that $|\mathcal{G}^{N,\mathcal{L}}(u)| > r$ is bounded by c/r. Therefore the measure of the set of $u \in U$ such that $|\mathcal{G}^{N,\mathcal{L}}(u)| \leq r$ is bounded from below by $1 - c/r$. Thus we have proved that for every realization δ of the data, there exists $C(\delta) > 0$ such that

$$Z^{N,\mathcal{L}}(\delta) \geq \int_U \exp\left(-\frac{1}{2}(|\delta|_\Sigma + |\mathcal{G}^{N,\mathcal{L}}(u)|_\Sigma)^2\right)d\rho(u) > C(\delta) > 0 .$$

$\square$

Let $(u^{(k)})_k$ be the Markov chain generated by the sampling process (10) with the acceptance probability being replaced by

$$\alpha^{N,\mathcal{L}}(u,v) = 1 \wedge \exp(\Phi^{N,\mathcal{L}}(u,\delta) - \Phi^{N,\mathcal{L}}(v,\delta)) . \tag{31}$$

We denote by

$$E_M^{\rho^{N,\mathcal{L},\delta}}[g] = \frac{1}{M}\sum_{k=1}^M g(u^{(k)}) .$$

We then have the following error for the gpc-FE surrogate based MCMC method (for notation we refer to Section 1.2)

Proposition 15 *Let Assumptions 1(i) and 9 hold and let g be a bounded continuous function from U to $\mathbb{R}$. Then*

$$|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]| \le c_6 M^{-1/2} + c_7 N_{\mathrm{dof}}^{-\tau}, \tag{32}$$

$\mathcal{P}\rho^{N,\mathcal{L},\delta,N,\mathcal{L}}$ *almost surely, where $c_6 \le c_8|\xi_M|$ where ξ_M is a random variable which converges weakly as $M \to \infty$ to $\xi \sim N(0,1)$; the constants c_7 and c_8 are deterministic and do not depend on M, N and N_{dof}.*

There exists a deterministic positive constant c_9 such that the gpc-MCMC converges in the mean square with the same rate of convergence

$$\left(\mathcal{E}^{\rho,N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]\right)^{1/2} \le c_9(M^{-1/2} + N_{dof}^{-\tau}) .$$

Proof Using (31), a random draw from ρ has probability larger than $\exp(-\Phi^{N,\mathcal{L}}(v;\delta))$ of being accepted. Therefore the transition kernel of the Markov chain generated by (10) with the acceptance probability (31) satisfies

$$p(u, A) \ge \int_A \exp(-\Phi^{N,\mathcal{L}}(v;\delta))d\rho(v).$$

Using Theorem 16.2.4 of [22], we deduce that the nth iteration of the transition kernel satisfies

$$\|p^n(u,\cdot) - \rho^{N,\mathcal{L},\delta}\|_{\mathrm{TV}} \le 2\left(1 - \int_U \exp(-\Phi^{N,\mathcal{L}}(v;\delta))d\rho(v)\right)^n .$$

From the proof of Proposition 14, we have

$$\int_U \exp(-\Phi^{N,\mathcal{L}}(v;\delta))d\rho(v) \ge \exp(-c_2(\delta)^2/2) + cN_{dof}^{-2\tau} .$$

Thus, we can choose a constant $R < 1$ independent of the approximating parameters N and $\mathcal{L}$ so that for all $n \in \mathbb{N}$ holds

$$\|p^n(u,\cdot) - \rho^{N,\mathcal{L},\delta}\|_{\mathrm{TV}} \le 2(1-R)^n .$$

In a similar manner as for Proposition 11, we deduce the probabilistic bound. For the mean square bound, similar to the proof of Proposition 11, we have

$$\mathcal{E}^{\rho^{N,\mathcal{L}},N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right] \le C(M^{-1/2} + N_{dof}^{-\tau})^2 .$$

Let $U' := \{u \in U : |\mathcal{G}^{N,\mathcal{L}}(u) - \mathcal{G}(u)| > 1\}$. We deduce that there exists a constant $c > 0$ independent of $\mathcal{L}$, N_{dof}, N such that $\rho(U') \le cN_{dof}^{-2\tau}$ and such that we may estimate

$$\mathcal{E}^{\rho,N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]$$
$$= \int_{U'} \mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]d\rho(u^{(0)})$$
$$+ \int_{U\setminus U'} \mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]d\rho(u^{(0)})$$
$$\le CN_{dof}^{-2\tau} + \int_{U\setminus U'} \mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]d\rho(u^{(0)})$$
$$\le CN_{dof}^{-2\tau} +$$
$$\int_{U\setminus U'} \mathcal{E}_{u^{(0)}}^{N,\mathcal{L}}\left[\left|\mathbb{E}^{\rho^\delta}[g] - E_M^{\rho^{N,\mathcal{L},\delta}}[g]\right|^2\right]Z^{N,\mathcal{L}}(\delta)\exp(\Phi^{N,\mathcal{L}}(u;\delta))d\rho^{N,\mathcal{L},\delta}(u^{(0)}) .$$

On $U \setminus U'$, $\sup_{u \in U} |\mathcal{G}^{N,\mathcal{L}}(u)|$ is uniformly bounded with respect to all N and $\mathcal{L}$. From this, we arrive at the conclusion. $\qquad\square$

Remark 16 *In Proposition 15, g is assumed to be a bounded continuous function from U to $\mathbb{R}$. The gpc-FE surrogate accelereated MCMC is of particular interest in the case where g is given by $\ell \circ P$ where ℓ is a bounded, linear functional on V, i.e. $\ell \in V^*$. From Assumption 9 and the fact that*

$$\frac{d\rho^\delta}{d\rho}(u) = \frac{1}{Z(\delta)} \exp(-\Phi(u;\delta)) ,$$

we deduce that

$$|\mathbb{E}^{\rho^\delta}[\ell(P(\cdot,u))] - \mathbb{E}^{\rho^\delta}[\ell(P_{N,\mathcal{L}}(\cdot,u))]| \leq C(\delta,\ell)N_{dof}^{-\tau} .$$

On the other hand, from Proposition 14, we have (cf. [8, Eq. (2.7)])

$$|\mathbb{E}^{\rho^\delta}[\ell(P_{N,\mathcal{L}}(\cdot,u))] - \mathbb{E}^{\rho^{N,\mathcal{L},\delta}}[\ell(P_{N,\mathcal{L}}(\cdot,u))]| \leq C(\delta,\ell)N_{dof}^{-\tau} .$$

Therefore, by the triangle inequality,

$$|\mathbb{E}^{\rho^\delta}[\ell(P(\cdot,u))] - \mathbb{E}^{\rho^{N,\mathcal{L},\delta}}[\ell(P_{N,\mathcal{L}}(\cdot,u))]| \leq C(\delta,\ell)N_{dof}^{-\tau} .$$

We wish to approximate $\mathbb{E}^{\rho^{N,\mathcal{L},\delta}}[\ell(P_{N,\mathcal{L}}(\cdot,u))]$ with a Markov Chain-Monte Carlo algorithm. In doing so, the following difficulty may arise: although $\ell(P(\cdot,u))$ is uniformly bounded with respect to $u \in U$, $\sup_{u \in U} \ell(P_{N,\mathcal{L}}(\cdot,u))$ may not be uniformly bounded with respect to N and $\mathcal{L}$. However, we can still apply Proposition 15 by using a cut-off argument: to this end, we define the continuous bounded function $\tilde{g}(u) : U \to \mathbb{R}$ by truncation, i.e.

$$\tilde{g}(u) := \begin{cases} \ell(P_{N,\mathcal{L}}(\cdot,u)) & if & |\ell(P_{N,\mathcal{L}}(\cdot,u))| \leq \sup_{u \in U} |\ell(P(\cdot,u))| + 1 , \\ \sup_{u \in U} |\ell(P(\cdot,u))| + 1 & if & \ell(P_{N,\mathcal{L}}(\cdot,u)) > \sup_{u \in U} |\ell(P(\cdot,u))| + 1 , \\ -\sup_{u \in U} |\ell(P(\cdot,u))| - 1 & if & \ell(P_{N,\mathcal{L}}(\cdot,u)) < -\sup_{u \in U} |\ell(P(\cdot,u))| - 1 . \end{cases}$$

Define $U' := \{u \in U : |\ell(P(\cdot,u)) - \ell(P_{N,\mathcal{L}}(\cdot,u))| > 1\}$. From Assumption 9, we find that $\rho(U') < C(\ell,\delta)N_{dof}^{-2\tau}$. It follows then that there exists a constant $c > 0$ depending on the data δ, but independent of N and of $\mathcal{L}$ such that

$$|\mathbb{E}^{\rho^{N,\mathcal{L},\delta}}[\ell(P_{N,\mathcal{L}}(\cdot,u)) - \tilde{g}(u)]| \leq \int_{U'} |\ell(P_{N,\mathcal{L}}(\cdot,u)) - \tilde{g}(u)| d\rho^{N,\mathcal{L},\delta}(u)$$

$$\leq C(\delta) \int_U I_{U'}(u)(|\ell(P_{N,\mathcal{L}}(\cdot,u))| + c)d\rho(u)$$

$$\leq C(\delta)\rho(U')^{1/2}(\|\ell(P_{N,\mathcal{L}}(\cdot,u))\|_{L^2(U,\rho;\mathbb{R})} + c) \leq C(\delta)N_{dof}^{-\tau} .$$

Therefore, we may run the MCMC algorithm on $\mathbb{E}^{\rho^{N,\mathcal{L},\delta}}[\tilde{g}(u)]$.

At each step of the MCMC algorithm, we need to compute $\ell(P_{N,\mathcal{L}}(\cdot,u^{(k)}))$ which, *for linear functionals $\ell(\cdot)$, is equal to* $\sum_{\nu \in \Lambda} \ell(P_{\nu,\mathcal{L}})L_\nu(\cdot,u^{(k)})$. Because the parametric solution of the elliptic problem can be precomputed before the MCMC is run, and then needs only to be *evaluated* at each state of the MCMC method, significant savings can be obtained. We illustrate this, using the ideas of the previous Remark 16, to guide the choice of test functions.

Proposition 17 *Let Assumptions 1(i) and 9 hold and let $g(u) = \ell(P(\cdot, u))$ where $\ell \in V^*$. Then the total number of floating point operations required for performing M steps in the Metropolis-Hastings method as $N, M \to \infty$ is bounded by $O(N_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta} + MN \log N)$.*

Proof By Assumption 9 and with the notation as in that assumption, the cost of solving one instance of problem (62) is bounded by $O(N_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta})$. At each MCMC step, we need to evaluate the observation functionals

$$\mathcal{O}_i(P_{N,\mathcal{L}}(u^{(k)})) = \sum_{\nu \in \Lambda_N} \mathcal{O}_i(P_{\nu,\mathcal{L}}) L_\nu(u^{(k)}) . \tag{33}$$

We note in passing that the storage of the parametric gpc-type representation of the forward map (33) requires only one real per gpc mode, *provided that only functionals of the forward solution are of interest.* We now estimate the complexity of computing one draw of the forward map (33). For $\nu \in \mathcal{F}$, each multivariate Legendre polynomial $L_\nu(u^k)$ can be evaluated with $O(|\nu|)$ float point operations. As $|\nu| = O(\log N)$, computing the observation functionals $\mathcal{O}_i(P_{N,\mathcal{L}})$ requires $O(N \log N)$ floating point operations. Thus we need $O(N_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta} + MN \log N)$ floating point operations to perform M steps of the Metropolis-Hastings method with sampling of the surrogate, sparse gpc-Finite Element representation of the forward map.

□

Theorem 18 *Let Assumptions 1(i) and 9 hold. For $g(u) = \ell(P(\cdot, u))$ with given $\ell \in V^*$, with probability $p_{N_{\mathrm{dof}}}(t)$ the conditional expectation $\mathbb{E}^{\rho^\delta}[g(u)]$ can be approximated with N_{dof} degrees of freedom, incurring an error of $O(N_{\mathrm{dof}}^{-\tau})$ using at most*

$$cN_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta} + ct^2 N_{\mathrm{dof}}^{2\tau + \tau/\sigma} \log(N_{\mathrm{dof}})$$

many floating point operations, where

$$\lim_{N_{\mathrm{dof}} \to \infty} p_{N_{\mathrm{dof}}}(t) \to \int_{-c't}^{c't} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx,$$

for some constants c, c' independent of N_{dof}.

In the mean square with respect to the measure $\mathcal{P}^{\rho, N, \mathcal{L}}$, $\mathbb{E}^{\rho^\delta}[g(u)]$ can be approximated with N_{dof} degrees of freedom, with an error $N_{dof}^{-\tau}$ using at most

$$O(N_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta} + N_{\mathrm{dof}}^{2\tau + \tau/\sigma} \log(N_{\mathrm{dof}}))$$

floating point operations.

Proof We relate the number of MCMC realizations M with the total number of degrees of freedom N_{dof} by equating the terms in the error bound (32). To this end, we choose $M = t^2 N_{\mathrm{dof}}^{2\tau}$ where $t = c_8 |\xi_M|$; the constant c_8 and the random variable ξ_M are as in Proposition 15 . With $N = O(N_{\mathrm{dof}}^{\tau/\sigma})$, the number of floating point operations required in Proposition 17 is bounded by

$$O\left(N_{\mathrm{dof}}^{\alpha}(\log N_{\mathrm{dof}})^{\beta} + t^2 N_{\mathrm{dof}}^{2\tau + \tau/\sigma} \log N_{\mathrm{dof}}\right) .$$

As ξ_M converges weakly to the normal Gaussian variable, we deduce the limit for the probability density $p_{N_{\text{dof}}}(t)$ of the random variable t. The proof for the mean square approximation is similar. □

Remark 19 *In studying complexity of the Plain MCMC method in Theorem 13, the discretized parametric PDE (4) is to be solved once at every step of the MCMC process, using N_{dof} degrees of freedom, with $O(N_{\text{dof}}^{2/d})$ steps required (the multiplying constant depends on a random variable when we consider the realization-wise error). Ignoring* log *factors, the error resulting from discretization and running the MCMC on the discretized PDE can be bounded in terms of the total number of floating point operations N_{fp} by $O(N_{fp}^{-1/(d+2+1/q)})$. In Theorem 18 the discretized forward PDE is solved for every realization before running the MCMC process. The rate of convergence of the MCMC process in terms of the total number of floating point operations used is $O(N_{fp}^{-\min(\tau/\alpha, 1/(2+1/\sigma))})$. This can be significantly smaller than the rate of convergence in Theorem 13 when α is close to 1. To see this we need to dig into the manner in which Assumptions 9 are typically verified which, in turn, requires Assumptions 1(iii),(iv) (which together imply Assumptions 1(ii), see discussion in Appendix E). For example, with the decay rate of $\|\psi_j\|_\infty$ in Assumption 1(iii), the summability constant p in Assumption 30 can be any constant that is greater than $1/s$. Therefore, the gpc approximation rate σ in Proposition 31 can be chosen as any positive constant smaller than $s - 1/2$. On the other hand, the J-term approximation rate q of the unknown input K in Assumption 1(ii) is bounded by $s - 1$. As*

$$2 + \frac{1}{s - 1/2} < d + 2 + \frac{1}{s - 1},$$

we therefore can choose σ so that

$$\frac{1}{2 + 1/\sigma} > \frac{1}{d + 2 + 1/q}.$$

As shown in [5], when $(\|P_\nu\|_{H^2(D)})_\nu \in \ell^p(\mathcal{F})$, τ can be chosen as $1/d$. Thus, when α is sufficiently close to 1, the complexity of the sparse gpc-MCMC approach is superior to that of the Plain MCMC approach in the previous section.

5. Multilevel MCMC

In the preceding section we showed that complexity reduction is possible in the Plain MCMC sampling of the posterior measure ρ^δ provided that all samples are determined from *one* precomputed, "surrogate" sparse tensor gpc-representation of the forward map of the discretization of the parametric, deterministic problem (4). Specifically, we proved that when the forward map $\mathcal{G}(u)$ consists of *continuous, linear functionals $\mathcal{O}_i(\cdot)$ on the forward solution $U \ni u \mapsto P(\cdot, u) \in V$*, and when the functional whose posterior expectation we seek is also linear on this space, this gpc-type approximation of the solution can reduce the complexity required. Lower efficiency results if, for example, the rate of convergence of the sparse tensor finite element solution in

(62) is moderate in terms of the total number of degrees of freedom, and/or if the complexity grows superlinearly with respect to the number of degrees of freedom. Furthermore, although an increasing number of efficient algorithms for the computation of approximate responses of the forward problem on the entire parameter space U are available (e.g. [3, 27, 11, 2, 4]) and therefore gpc-surrogates for the MCMC are available, many systems of engineering interest may not admit gpc-based representations of the parametric forward maps. Finding other, non-gpc based, methods for reducing the complexity of Plain MCMC sampling under ρ^δ is therefore of interest. We do this by using ideas from Multilevel Monte Carlo. The resulting complexity is summarized in Theorem 21. We give sufficient conditions on the approximation methods and on the basis functions ψ_j appearing in (5) such that complexity reduction is possible by performing a multilevel sampling procedure where a number of samples depending on the discretization parameters is used for problem (4).

5.1. Derivation of the MLMCMC

For given, fixed $\ell \in V^*$, a bounded linear functional on V, we aim at estimating $\mathbb{E}^{\rho^\delta}[\ell(P(\cdot, u))]$ where P is the solution of problem (4). For each level $l = 1, 2, \ldots, L$, we assume that problem (4) is discretized with the truncation of the Karhúnen-Loève expansion after J terms with $J = J_l$ as defined in (13) and with a finite element discretization meshwidth h_l. The multilevel FE-discretization of the forward problem (4) and the truncation (13) induces a corresponding hierarchy of approximations $\rho^{J_l, l, \delta}$ of the posterior measure ρ^δ.

Following [9, 1, 23, 10] the MLMCMC will be based on sampling a telescopic expansion of the discretization error with a level-dependent sample size. Contrary to [9, 1, 23], however, we introduce now two multilevel discretization hierarchies, one for the parametric forward solutions $\{P^{J_l, l}\}_{l=0}^L$ (where the level corresponds to mode truncation of the coefficient and to discretization for the approximate solution of the parametric forward problem) and a second one $\{\rho(J_{l'}, l', \delta)\}_{l'=0}^L$ for the posterior measure. The 'usual' telescoping argument as in [9] or in [1, 23] together with the single-level error bound in Proposition 11 alone does not allow for obtaining a convergence for the present problem.

As in Section 3, we work under Assumptions 1 (i), (ii) and 7. We recall the sequence of discretization levels in the FE discretizations in the physical domain D, as in Assumption 7, and the input truncation dimension J in Assumption 1(ii). We then derive the Multilevel MCMC-FEM as follows. First, we note that, by Assumption 7, using the uniform boundedness of the P^J in W as in the preceding section, there exists $C > 0$ independent of L such that

$$|\mathbb{E}^{\rho^\delta}[\ell(P(\cdot, u))] - \mathbb{E}^{\rho^\delta}[\ell(P^{J_L, L}(\cdot, u))]| \leq C \sup_{u \in U} \|P(\cdot, u) - P^{J_L, L}(\cdot, u)\|_V \leq C 2^{-L} . \qquad (34)$$

With the convention that expectation with respect to $\rho^{J_{-1}, -1, \delta}$ denotes integration with respect to the measure which assigns zero mass to all subsets of the probability space,

we write

$$\mathbb{E}^{\rho^{J_L,L,\delta}}[\ell(P(\cdot,u))] = \sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}}[\ell(P(\cdot,u))] - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P(\cdot,u))]\right) \quad (35)$$

$$= \sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P(\cdot,u))]. \quad (36)$$

Analogously we may write, for any $L' \leq L$ (omitting the arguments of P and its approximations for brevity of notation)

$$\mathbb{E}^{\rho^{J_L,L,\delta}}[\ell(P^{J_{L'},L'})] = \sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P^{J_{L'},L'})]. \quad (37)$$

With the convention that $P^{J_{-1},-1} := 0$ we have for each l and L'

$$\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P^{J_{L'},L'})] = \sum_{l'=0}^{L'}\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})]. \quad (38)$$

Finally, a computable multilevel approximation will be obtained on running, for each level $l = 0, 1, ..., L$ of truncation resp. Galerkin projection, simultaneously a suitable number of realizations of a Markov chain $\mathcal{C}_l$ to approximate the expecations

$$\sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P^{J_{L'},L'})]$$

$$= \sum_{l=0}^{L}\sum_{l'=0}^{L'}\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})] \quad (39)$$

by sample averages of $M_{ll'}$ many realizations, upon choosing $L'(l)$ judiciously. To derive a computable MLMCMC estimator we observe that, for any measurable functions $Q : U \to \mathbb{R}$ which is integrable with respect to the approximate posterior measures $\rho^{J_l,l,\delta}$,

$$\left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}\right)[Q]$$

$$= \frac{1}{Z^{J_l,l}}\int_U \exp(-\Phi^{J_l,l}(u;\delta))Q(u)d\rho(u) - \frac{1}{Z^{J_{l-1},l-1}}\int_U \exp(-\Phi^{J_{l-1},l-1}(u;\delta))Q(u)d\rho(u)$$

$$= \frac{1}{Z^{J_l,l}}\int_U \exp(-\Phi^{J_l,l}(u;\delta))\left(1 - \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta))\right)Q(u)d\rho(u)$$

$$+ \left(\frac{Z^{J_{l-1},l-1}}{Z^{J_l,l}} - 1\right)\frac{1}{Z^{J_{l-1},l-1}}\int \exp(-\Phi^{J_{l-1},l-1}(u;\delta))Q(u)d\rho(u).$$

We note further that

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

Thus an approximation for $Z^{J_{l-1},l-1}/Z^{J_l,l} - 1$ can be found by running the MCMC with respect to the approximate posterior $\rho^{J_l,l,\delta}$ to sample the potential difference

$\exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)) - 1$. We define the Multilevel Markov Chain Monte Carlo estimator $E_L^{MLMCMC}[\ell(P)]$ of $\mathbb{E}^{\rho^\delta}[\ell(P)]$ as

$$E_L^{MLMCMC}[\ell(P)] =$$

$$\sum_{l=0}^{L}\sum_{l'=0}^{L'(l)} E_{M_{ll'}}^{\rho^{J_l,l,\delta}}\left[\left(1 - \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta))\right)(\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1}))\right]$$

$$+E_{M_{ll'}}^{\rho^{J_l,l,\delta}}\left[\exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)) - 1\right] \cdot E_{M_{ll'}}^{\rho^{J_{l-1},l-1,\delta}}\left[\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})\right] \ .$$

Up to this point, the choice of the index $L'(l)$ and the sample sizes $M_{ll'}$ are still at our disposal. Choices for them will be made based on an error-versus-work analysis of this estimator that we now present.

5.2. Error Analysis

To perform the error analysis of the MLMCMC approximation we decompose the error into three terms as follows.

Proposition 20 *There holds*

$$\mathbb{E}^{\rho^\delta}[\ell(P)] - E_L^{MLMCMC}[\ell(P)] = I_L + II_L + III_L \tag{40}$$

where

$$I_L := \mathbb{E}^{\rho^\delta}[\ell(P)] - \mathbb{E}^{\rho^{J_L,L,\delta}}[\ell(P)], \quad II_L = \sum_{l=0}^{L}(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}})[\ell(P) - \ell(P^{J_{L'(l)},L'(l)})]$$

and

$$III_L := \sum_{l=0}^{L}\sum_{l'=0}^{L'(l)}\left\{\mathbb{E}^{\rho^{J_l,l,\delta}}\left[(1 - \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)))](\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1}))\right]\right.$$

$$-E_{M_{ll'}}^{\rho^{J_l,l,\delta}}\left[(1 - \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)))](\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1}))\right]\right\}$$

$$+\left\{\mathbb{E}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)) - 1] \cdot \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})]\right.$$

$$-E_{M_{ll'}}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)) - 1] \cdot E_{M_{ll'}}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})]\right\}$$

respectively.

Proof From equation (36) we have

$$\mathbb{E}^{\rho^\delta}[\ell(P)] - \mathbb{E}^{\rho^{J_L,L,\delta}}[\ell(P)] = \mathbb{E}^{\rho^\delta}[\ell(P)] - \sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}}[\ell(P)] - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P)]\right) \ . \tag{41}$$

from which it follows that

$$\mathbb{E}^{\rho^\delta}[\ell(P)] - \mathbb{E}^{\rho^{J_L,L,\delta}}[\ell(P)] = \mathbb{E}^{\rho^\delta}[\ell(P)]$$

$$-\sum_{l=0}^{L}\left(\mathbb{E}^{\rho^{J_l,l,\delta}}[\ell(P^{J_{L'},L'})] - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{L'},L'})]\right) - II_L \ .$$

Rearranging and using (38) gives

$$
\mathbb{E}^{\rho^\delta}[\ell(P)] = \quad I_L + II_L + \sum_{l=0}^{L} \left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}} \right) [\ell(P^{J_{L'},L'})]
$$

$$
= I_L + II_L + \sum_{l=0}^{L} \sum_{l'=0}^{L'} \left(\mathbb{E}^{\rho^{J_l,l,\delta}} - \mathbb{E}^{\rho^{J_{l-1},l-1,\delta}} \right) [\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})] .
$$

The claimed expression follows from the definition of the MLMCMC method which computes the MCMC sample path averages of the terms on the right hand side in (39).
□

We next derive an error bound by estimating the three terms in the error (40) separately. Throughout we choose $J_l = 2^{\lceil l/q \rceil}$. For the first term I_L, we obtain from Proposition 10, the bound

$$
|I_L| \leq C(\delta)(J_L^{-q} + 2^{-L}) \leq C(\delta) 2^{-L} . \tag{42}
$$

Likewise we obtain, using Assumption 7 in addition to Proposition 10,

$$
|II_L| \leq C(\delta) \sum_{l=0}^{L} (J_\ell^{-q} + 2^{-l}) 2^{-L'(l)} \leq C(\delta) \sum_{l=0}^{L} 2^{-(l+L'(l))} . \tag{43}
$$

We now estimate III_L. Since $J_l = 2^{\lceil l/q \rceil}$ we have that

$$
\sup_{u \in U} |\ell(P^{J_{l'},l'}) - \ell(P^{J_{l'-1},l'-1})| \leq C 2^{-l'} .
$$

Further

$$
\sup_{u \in U} |1 - \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta))|
$$

$$
\leq \sup_{u \in U} |\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)|(1 + \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_{l-1},l-1}(u;\delta)))
$$

$$
\leq C 2^{-l} .
$$

As in Section 3, for each discretization level $l \in \mathbb{N}$, we introduce the Markov chains $\mathcal{C}_l = \{u^{(k)}\}_{k \in \mathbb{N}_0} \subset \mathbb{R}^{J_l}$ which are started in the restriction of $u^{(0)}$ to $\mathbb{R}^{J_l}$ and then generated by (10) with the J_l-term truncated parametric coefficient expansions (13) and forward problems (16) which are Galerkin-discretized at mesh level l, ie. with the acceptance probability $\alpha(u,v)$ in (9) replaced by

$$
\alpha^{J_l,l}(u,v) = 1 \wedge \exp(\Phi^{J_l,l}(u;\delta) - \Phi^{J_l,l}(v;\delta)) , \quad (u,v) \in U \times U . \tag{44}
$$

For each discretization level $l = 1, 2, ...$, the chains $\mathcal{C}_l$ are pairwise independent. For every fixed discretization level L, we denote by $\mathbf{C}_L = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_L\}$ the collection of Markov chains obtained from the different discretizations, and by $\mathbf{P}_L$ the product probability measure on the probability space generated by the collection of these L chains. For each fixed discretization level L, this measure describes the law of the collection of chains $\mathbf{C}_L$:

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

Let $\mathbf{E}_L$ be the expectation over all realizations of the collection $\mathbf{C}_L$ of chains $\mathcal{C}_l$ with respect to the product measure $\mathbf{P}_L$. We have, by Assumption 7 and by Proposition 5,

$$\mathbf{E}_L\left[\left\{\mathbb{E}^{\rho^{J_l,l,\delta}}\left[(1-\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta)))](\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1}))\right]\right.\right.$$
$$\left.\left.-E_{M_{ll'}}^{\rho^{J_l,l,\delta}}\left[(1-\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta)))](\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1}))\right]\right\}^2\right]$$
$$\leq CM_{ll'}^{-1}2^{-2(l+l')}\ .$$

We also have, again using Assumption 7 and the last estimate in Proposition 5,

$$\mathbf{E}_L\left[\left\{\mathbb{E}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))-1]\cdot\mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1})]\right.\right.$$
$$\left.\left.-E_{M_{ll'}}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))-1]\cdot E_{M_{ll'}}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1})]\right\}^2\right]$$
$$\leq C\,\mathbf{E}_L\left[\left\{\mathbb{E}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))-1]-E_{M_{ll'}}^{\rho^{J_l,l,\delta}}[\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))-1]\right\}^2\right]$$
$$\cdot\sup_{u\in U}|\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1})|^2+$$
$$C\sup_{u\in U}|\exp(\Phi^{J_l,l}(u;\delta)-\Phi^{J_{l-1},l-1}(u;\delta))-1|^2$$
$$\cdot\mathbf{E}_L\left[\left\{\mathbb{E}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1})]-E_{M_{ll'}}^{\rho^{J_{l-1},l-1,\delta}}[\ell(P^{J_{l'},l'})-\ell(P^{J_{l'-1},l'-1})]\right\}^2\right]$$
$$\leq CM_{ll'}^{-1}2^{-2(l+l')}\ .$$

Hence

$$\mathbf{E}_L[|III_L|^2]\leq CL\sum_{l=0}^{L}L'(l)\sum_{l'=0}^{L'(l)}M_{ll'}^{-1}2^{-2(l+l')}\ . \tag{45}$$

To achieve a bound on the error (40) which is $O(L^m2^{-L})$ for some $m\in\mathbb{R}_+$, we choose

$$L'(l):=L-l,\quad\text{and}\quad M_{ll'}:=2^{2(L-(l+l'))}\ . \tag{46}$$

We then have

$$\mathcal{E}[|III_L|^2]\leq CL\sum_{l=0}^{L}(L-l)^22^{-2L}\leq CL^42^{-2L}\ .$$

Theorem 21 *For $d=2,3$, under the assumption that $P\in L^\infty(U,\rho;H^2(D)\bigcap V)$, and with the choices (46) we have that*

$$\mathbf{E}_L[|\mathbb{E}^{\rho(\delta)}[P]-E_L^{MLMCMC}[P]|]\leq C(\delta)L^22^{-L}\ . \tag{47}$$

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 $d=3$. Assuming the availability of a linear scaling multilevel iterative solver for the linear systems of equations arising from Galerkin discretization, the total number of floating point operations in the approximate evaluation of the conditional expectation under the posterior (for one instance of data δ) 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 level by N, the error in (47) is bounded by $O((\log N)^{3/2}N^{-1/2})$ for $d=2$ and by $O((\log N)^2N^{-1/3})$ for $d=3$.

The total number of floating point operations to realize the MLMCMC 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 $d = 3$.

Proof At each step we solve the truncated forward equation (14) for the truncation levels J_l and $J_{l'}$ in expansion (13) and the Finite Element discretization levels l and l', respectively. Assuming approximate solution of the discretized problems with termination at the discretization error $O(2^{-l})$, resp. $O(2^{-l'})$, and a linear scaling iterative solver with the choices (46) the total work is bounded by an absolute multiple of the total number of degrees of freedom, which in turn is bounded, for $d \geq 2$, by

$$\sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} M_{ll'}(2^{dl} + 2^{dl'}) = 2^{2L} \sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} \left(2^{(d-2)l} \cdot 2^{-2l'} + 2^{-2l} \cdot 2^{(d-2)l'}\right)$$

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

Therefore for $d = 2$, the number of degrees of freedom 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 the proof of Proposition 12, we infer that the number of floating point operations at each step is not larger than $O(l^{d-1} 2^{dl} J_l + l'^{d-1} 2^{dl'} J_{l'})$ so that, using again (46), the total number of floating point operations required to evaluate the MLMCMC estimator is bounded by

$$\lesssim \sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} M_{ll'}(l^{d-1} 2^{dl} J_l + l'^{d-1} 2^{dl'} J_{l'})$$

$$\lesssim 2^{2L} \sum_{l=0}^{L} \sum_{l'=0}^{L'(l)} (l^{d-1} 2^{(d-2+1/q)l} 2^{-2l'} + l'^{d-1} 2^{(d-2+1/q)l'} 2^{-2l})$$

$$\lesssim L^{d-1} 2^{2L} \left(\sum_{l=0}^{L} 2^{(d-2+1/q)l} + \sum_{l=0}^{L} 2^{(d-2+1/q)(L-l)} 2^{-2l}\right)$$

$$\lesssim L^{d-1} 2^{(d+1/q)L}.$$

$\square$

Remark 22 *In the preceding work analysis we assumed the availability of linear scaling, iterative solvers for the approximate solution of the Finite Element discretizations (16) of the parametric, elliptic problems (14). These include Richardson or conjugate gradient iteration with diagonal preconditioning using Riesz bases for the Finite Element spaces V^l as presented in Appendix D. However, if such bases are not available, our results will also apply to standard Finite Element discretizations, provided that linear scaling*

multilevel solvers, such as multigrid, are used to solve (16), for each increment of the Markov chain. Our work versus accuracy analysis will also apply in these cases, with identical conclusions.

6. Conclusions

Rewriting the convergence bounds in Theorems 13, 18 and 21 in terms of the error, to achieve an error ε in the mean square with respect to the probability measure that generates the randomness of the Markov Chain ie., $\mathcal{P}^{\rho,J,l}$, $\mathcal{P}^{\rho,N,\mathcal{L}}$ and $\mathbf{P}_L$, respectively we bound the complexity as follows (ignoring the logarithmic factors for conciseness):

- for the Plain MCMC procedure: $O(\varepsilon^{-d-2})$ degrees of freedom and $O(\varepsilon^{-d-2-1/q})$ floating point operations;

- for the gpc-MCMC procedure: $O(\varepsilon^{-1/\tau})$ degrees of freedom and $O(\varepsilon^{-\max(\alpha/\tau,2+1/\sigma)})$ floating point operations;

- and for the MLMCMC procedure: the essential optimal complexity of $O(\varepsilon^{-d})$ degrees of freedom and $O(\varepsilon^{-d-1/q})$ floating point operations.

Therefore, the complexity of the gpc-MCMC is superior to that of the Plain MCMC procedure when in Assumption 9 τ is close to $1/d$ and α is close to 1, which holds in many cases (see, e.g. [27] and the references therein). We also have shown that the asymptotic accuracy vs. work of the MLMCMC is always superior to that of the Plain MCMC. We note that the complexity of the MLMCMC is of the same order as that for solving one single realization of equation (4) with the same level of truncation for the coefficient, which indicates some form of optimality of the MLMCMC method proposed here.

We have considered the MLMCMC estimation of functionals $\ell(\cdot)$ of the solution; in doing so, we used only minimal regularity $\ell(\cdot) \in V^*$ of these functionals, leading to the convergence rates (34). Higher regularity $\ell(\cdot) \in L^2(D)$ will entail in (34) the Finite Element convergence rate 2^{-2L}, via a classical Aubin-Nitsche duality argument. Likewise, we only used the lowest order Finite Element methods. In the present analysis, we focused on first order Finite Element discretizations as stronger regularity assumptions on f, on $\ell(\cdot)$ and on the ψ_j will *not* imply corresponding higher rates of convergence for gpc-MCMC and ML-MCMC, due to the (maximal) order $1/2$ afforded by the MC method.

A further aspect pertaining to the overall complexity is the following: in [28] an entirely deterministic approach to the solution of the Bayesian inverse problem is presented, and in [26] corresponding numerical experiments are presented. These methods offer convergence rates in terms of the number M of samples which are superior to the rate $1/2$ of MC methods such as those analyzed here. Nonetheless, MCMC methods will remain popular because of their data-adaptive nature; the present results indicate how the use of gpc and multilevel ideas may be used to attain significant speedups of MCMC methods.

We also observe that we have concentrated on a very special MCMC method, namely the independence sampler. This has been dictated by the need to use MCMC methods which scale-independently of dimension and *for which there is a complete analysis of the convergence resulting chain* demonstrating this fact. Whilst there are now several dimension-independent MCMC methods [7, 31] the independence sampler is the only one for which the required analysis of the convergence properties is sufficiently developed for our theory; we anticipate further theoretical developments for different MCMC methods, and different inverse problems, in the future. However we do note that the independence sampler will work well when the negative log likelihood Φ does not vary significantly, although it will be inefficient in general. For problems where the prior is Gaussian or log-normal Gaussian, appropriate MCMC methods may be found in [7]. However for these more general methods the analysis of the Markov chain based on the methods of [22], as we have used for the independence sampler here, is not appropriate and more sophisticated arguments are required, as presented in [14].

Appendix A: Lipschitz Dependence of the Forward Map on Parameters

Proof of Proposition 2 The existence of a solution P to (4), and the bound (7), follows from standard application of Lax-Milgram theory. We have, for every $\phi \in V$, $u, u' \in U$,

$$\int_D K(x, u)(\nabla P(x, u) - \nabla P(x, u')) \cdot \nabla \phi(x) dx$$
$$= \int_D (K(x, u') - K(x, u)) \nabla P(x, u') \cdot \nabla \phi(x) dx . \tag{48}$$

Again using (6), i.e. that $K(x, u)$ is bounded below uniformly with respect to $(x, u) \in D \times U$, it follows that there exists $C > 0$ such that for all $u \in U$

$$\|P(\cdot, u) - P(\cdot, u')\|_V \leq C \|P(\cdot, u')\|_V \|K(\cdot, u') - K(\cdot, u)\|_{L^\infty(D)} . \tag{49}$$

Due to (7), it follows from (49) that there exists a constant $C > 0$ such that

$$\forall u \in U : \quad \|P(\cdot, u) - P(\cdot, u')\|_V \leq C \|K(\cdot, u') - K(\cdot, u)\|_{L^\infty(D)} . \tag{50}$$

From (5) and Assumption 1(i) it follows with $C > 0$ as in (50) that

$$\|P(\cdot, u) - P(\cdot, u')\|_V \leq C \sum_{\geq 1} |u_j - u'_j| \|\psi_j\|_{L^\infty(D)}$$

$$\leq C \|u - u'\|_{\ell^\infty(\mathbb{N})} \sum_{j \geq 1} \|\psi_j\|_{L^\infty(D)}$$

$$\leq C \frac{\kappa}{1 + \kappa} \bar{K}_{\min} \|u - u'\|_{\ell^\infty(\mathbb{N})} .$$

This establishes the desired Lipschitz continuity,

Appendix B: Bayesian inverse problems in measure spaces

On a measurable space (U, Θ) where Θ is a σ-algebra consider a measurable map $\mathcal{G} : U \to (\mathbb{R}^k, \mathcal{B}^k)$. The data δ is assumed to be an observation of $\mathcal{G}$ subject to an

unbiased observation noise ϑ:

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

We assume that ϑ is a centred Gaussian with law $N(0, \Sigma)$. Let ρ be a prior probability measure on (U, Θ). Our purpose is to determine the conditional probability $\mathbb{P}(u|\delta)$ on (U, Θ). The following result holds.

Proposition 23 *Assume that* $\mathcal{G} : U \rightarrow \mathbb{R}^k$ *is measurable. The posterior measure* $\rho^\delta(du) = \mathbb{P}(du|\delta)$ *given data* δ *is absolutely continuous with respect to the prior measure* $\rho(du)$ *and has the Radon-Nikodym derivative (2) with* Φ *given by (3).*

This result is established in Cotter et al.[8] and Stuart [29]. Though the setting in [8] and [29] is in a Banach space X, the proofs of Theorem 2.1 in [8] and Theorem 6.31 of [29] hold for any measurable spaces as long as the mapping $\mathcal{G}$ is measurable.

To study the well-posedness of the posterior measures, that is continuity with respect to changes in the observed data, we use the Hellinger distance, as in Cotter et al. [8], which is defined as

$$d_{\text{Hell}}(\mu, \mu') = \left(\frac{1}{2} \int_U \left(\sqrt{\frac{d\mu}{d\rho}} - \sqrt{\frac{d\mu'}{d\rho}} \right)^2 d\rho \right)^{1/2} \tag{51}$$

for any two measures μ and μ' on U which are absolutely continuous with respect to a common reference measure ρ¶. In [8], it is proved that when U is a Banach space, if the prior measure ρ is Gaussian, and under the conditions that Φ grows polynomially with respect to u, and is locally Lipschitz with respect to u fixing y and with respect to y fixing u, in the second case with a Lipschitz constant which also grows polynomially in u, then the posterior measure given the data δ, i.e. ρ^δ, is locally Lipschitz in the Hellinger distance d_{Hell}:

$$d_{\text{Hell}}(\rho^\delta, \rho^{\delta'}) \leq c|\delta - \delta'| \,,$$

where (recall) $|\cdot|$ denotes the Euclidean distance in $\mathbb{R}^k$. The Fernique theorem plays an essential role in the proofs, exploiting the fact that polynomially growing functions are integrable with respect to Gaussians. In this section, we extend this result to measurable spaces under more general conditions than those in Assumption 2.4 of Cotter et al. [8]; in particular we do not assume a Gaussian prior. The following assumption concerning the local boundedness, and local Lipschitz dependence of Φ on δ, will be crucial in our argument. Its validity for the model problem (4), with (5) and Assumption 1 will be verified in the ensuing proof of Proposition 26.

Assumption 24 *Let* ρ *be a probability measure on the measure space* (U, Θ). *The Bayesian potential function* $\Phi : U \times \mathbb{R}^k \rightarrow \mathbb{R}$ *satisfies:*

(i) (local boundedness) for each $r > 0$, *there is a constant* $\Phi^\star(r) > 0$ *and a set* $U(r) \subset U$ *of positive* ρ *measure such that for all* $u \in U(r)$ *and for all* δ *such that* $|\delta|_\Sigma \leq r$

$$0 \leq \Phi(u; \delta) \leq \Phi^\star(r) \,. \tag{52}$$

¶ Note that *any* such common reference measure will deliver the same value for the Hellinger distance

(ii) *(local Lipschitz continuity of Bayesian Potential Φ on data δ) there is a mapping $G : \mathbb{R} \times U \mapsto \mathbb{R}$ such that for each $r > 0$, $G(r, \cdot) \in L^2(U, \rho)$; and for every $|\delta|_\Sigma, |\delta'|_\Sigma \leq r$ it holds that*

$$|\Phi(u; \delta) - \Phi(u; \delta')| \leq G(r, u)|\delta - \delta'|_\Sigma .$$

Under Assumption 24, the definition (2) of the posterior measure ρ^δ is meaningful as we now demonstrate.

Proposition 25 *Under Assumption 24, the measure ρ^δ depends locally Lipschitz continuously on the data δ with respect to the Hellinger metric: for each positive constant r there is a positive constant $C(r)$ such that if $|\delta|_\Sigma, |\delta'|_\Sigma \leq r$, then*

$$d_{\mathrm{Hell}}(\rho^\delta, \rho^{\delta'}) \leq C(r)|\delta - \delta'|_\Sigma .$$

Proof Throughout this proof $K(r)$ denotes a constant depending on r, possibly changing from instance to instance. The normalization constant in (2) is

$$Z(\delta) = \int_U \exp(-\Phi(u; \delta))d\rho(u) . \tag{53}$$

We first show that for each $r > 0$, there is a positive constant $K(r)$ such that $Z(\delta) \geq K(r)$ when $|\delta|_\Sigma \leq r$. From (53) and Assumption 24(i) it follows that when $|\delta|_\Sigma \leq r$, there holds

$$Z(\delta) \geq \rho(U(r))\exp(-\Phi^\star(r)) > 0 . \tag{54}$$

Using the inequality $|\exp(-x) - \exp(-y)| \leq |x - y|$ which holds for $x, y \geq 0$ we find

$$|Z(\delta) - Z(\delta')| \leq \int_U |\Phi(u; \delta) - \Phi(u; \delta')|d\rho(u) .$$

From Assumption 24(ii),

$$|\Phi(u; \delta) - \Phi(u; \delta')| \leq G(r, u)|\delta - \delta'|_\Sigma .$$

As $G(r, u)$ is ρ-integrable, there is $K(r)$ such that

$$|Z(\delta) - Z(\delta')| \leq K(r)|\delta - \delta'|_\Sigma .$$

The Hellinger distance satisfies

$$2d_{\mathrm{Hell}}(\rho^\delta, \rho^{\delta'})^2 = \int_U \left(Z(\delta)^{-1/2}\exp(-\frac{1}{2}\Phi(u; \delta)) - Z(\delta')^{-1/2}\exp-\frac{1}{2}\Phi(u; \delta')\right)^2 d\rho(u)$$
$$\leq I_1 + I_2,$$

where

$$I_1 = \frac{2}{Z(\delta)}\int_U \left(\exp(-\frac{1}{2}\Phi(u; \delta)) - \exp(-\frac{1}{2}\Phi(u; \delta'))\right)^2 d\rho(u),$$

and

$$I_2 = 2|Z(\delta)^{-1/2} - Z(\delta')^{-1/2}|^2 \int_U \exp(-\Phi(u; \delta'))d\rho(u).$$

Using again $|\exp(-x) - \exp(-y)| \leq |x - y|$, we have, for constant $K(r) > 0$,

$$I_1 \leq K(r) \int_U |\Phi(u; \delta) - \Phi(u; \delta')|^2 d\rho(u)$$

$$\leq K(r) \int_U (G(r, u))^2 d\rho(u) |\delta - \delta'|_\Sigma^2 \leq K(r) |\delta - \delta'|_\Sigma^2 .$$

Furthermore,

$$|Z(\delta)^{-1/2} - Z(\delta')^{-1/2}|^2 \leq K(r)|Z(\delta) - Z(\delta')|^2 \leq K(r)|\delta - \delta'|_\Sigma^2 .$$

The conclusion follows. □

Proposition 26 *For the elliptic PDE (4), the function $\mathcal{G}$ defined by (8) and viewed as map from U to $\mathbb{R}^k$ is Lipschitz , if U is endowed with the topology of $\ell^\infty(\mathbb{N})$. Moreover, Assumption 24 holds with $U(r) = U$ and $G(r, u) = G(r)$ independent of u.*

Proof We have

$$\forall u, u' \in U : \quad |\mathcal{G}(u) - \mathcal{G}(u')| \leq C \max_i \{\|\mathcal{O}_i\|_{V^*}\} \|P(\cdot, u) - P(\cdot, u')\|_V .$$

From (50) there exists a constant $c > 0$ such that

$$\forall u, u' \in U : \quad |\mathcal{G}(u) - \mathcal{G}(u')| \leq C \|K(\cdot, u) - K(\cdot, u')\|_{L^\infty(D)} .$$

From Proposition 2, we deduce that $\mathcal{G}$ as map from $U \subset \ell^\infty(\mathbb{N})$ to $\mathbb{R}^k$ is Lipschitz .

We now verify Assumption 24. For the function $\mathcal{G}(u)$ we have from (8) for every $u \in U$ the bound

$$|\mathcal{G}(u)| \leq \max_i \{\|\mathcal{O}_i\|_{V^*}\} \|P(\cdot, u)\|_V .$$

From (7), $\sup\{|\mathcal{G}(u)| : u \in U\} < \infty$. We note that for given data δ, there holds

$$\forall u \in U : \quad |\Phi(u; \delta)| \leq \frac{1}{2}(|\delta|_\Sigma + |\mathcal{G}(u)|_\Sigma)^2$$

and hence, since $\sup_{u \in U} |\mathcal{G}(u)|$ is finite, the set $U(r)$ in Assumption 24(i) can be chosen as U for all r. We have, for every $u \in U$,

$$|\Phi(u; \delta) - \Phi(u; \delta')| \leq \frac{1}{2}|\langle \Sigma^{-1/2}(\delta + \delta' - 2\mathcal{G}(u)), \Sigma^{-1/2}(\delta - \delta')\rangle|$$

$$\leq \frac{1}{2}(|\delta|_\Sigma + |\delta'|_\Sigma + 2|\mathcal{G}(u)|_\Sigma)|\delta - \delta'|_\Sigma .$$

Choosing the function $G(r, u)$ in Assumption 24(ii) as

$$G(r, u) = \frac{1}{2}(2r + C),$$

for a sufficiently large constant $C > 0$ (depending only on bounds of the forward map and the observation functionals, but independent of the data δ and of u), we have shown that Assumption 24(ii) holds in the desired form. With Proposition 25 follows the assertion. □

Proof of Proposition 3 From Proposition 26, we deduce that $\mathcal{G}$ as map from $U \subset \ell^\infty(\mathbb{N})$ to $\mathbb{R}^k$ is Lipschitz and, hence, ρ-measurable. We then apply Proposition 23 to deduce the existence of ρ^δ and the formula for its Radon-Nikodym derivative with respect to ρ. Proposition 25 gives the desired Lipschitz continuity, since Proposition 26 establishes Assumption 24. □

Appendix C: Convergence Properties of the Independence Sampler

Proof of Proposition 5 We claim that (10) defines a Markov chain $\{u^{(k)}\}_{k=0}^{\infty}$ which is reversible with respect to ρ^{δ}. To see this let $\nu(du, dv)$ denote the product measure $\rho^{\delta}(du) \otimes \rho(dv)$ and $\nu^{\dagger}(du, dv) = \nu(dv, du)$. Note that ν describes the probability distribution of the pair $(u^{(k)}, v^{(k)})$ on $U \times U$ when $u^{(k)}$ is drawn from the posterior distribution ρ^{δ}, and $\nu^{\dagger}$ designates the same measure with the roles of u and v reversed. These two measures are equivalent as measures if ρ^{δ} and ρ are equivalent, which we establish below; it then follows that

$$\frac{d\nu^{\dagger}}{d\nu}(u, v) = \exp(\Phi(u; \delta) - \Phi(v; \delta)) , \quad (u, v) \in U \times U . \tag{55}$$

From Proposition 1 and Theorem 2 in [30] we deduce that (10) is a Metropolis-Hastings Markov chain which is ρ^{δ} reversible, since $\alpha(u, v)$ given by (9) is equal to $\min\{1, \frac{d\nu^{\dagger}}{d\nu}(u, v)\}$.

Equivalence of ρ^{δ} and ρ follows since the negative of the log-density is bounded from above and below, uniformly on U, because Proposition 26 shows that (52) holds with $U(r) = U$. Using (52) and (9) it follows that the proposed random draw from ρ has probability greater than $\exp(-\Phi^{\star}(r))$ of being accepted. Thus

$$p(u, A) \geq \exp(-\Phi^{\star}(r))\rho(A) \quad \forall u \in U .$$

The first result follows from [22], Theorem 16.2.4 with $X = U$. The second result follows from [22], Theorem 17.0.1. To see that the constant c in (11) can be bounded only in terms of $\Phi^{\star}(r)$ and $\sup_{u \in U} |g(u)|$, we note that it is given by

$$c^2 = \mathcal{E}^{\rho^{\delta}} |\bar{g}(u^{(0)})|^2 + 2 \sum_{n=1}^{\infty} \mathcal{E}^{\rho^{\delta}} [\bar{g}(u^{(0)}) \bar{g}(u^{(n)})] \tag{56}$$

where to ease notation we introduced the function $\bar{g}$ as $\bar{g} = g - \mathbb{E}^{\rho^{\delta}}(g)$ (and we do not tag the dependence of $\bar{g}$ on the data δ in the remainder of this proof). Note that $\sup_{u \in U} |\bar{g}(u)|$ is bounded uniformly w.r. to the data δ. The equation (56) is commonly known as the integrated autocorrelation time. Now

$$2 \sum_{n=0}^{\infty} \mathcal{E}^{\rho^{\delta}} [\bar{g}(u^{(0)}) \bar{g}(u^{(n)})] \leq 2 \sup_{u} |\bar{g}(u)| \mathbb{E}^{\rho^{\delta}} \sum_{n=0}^{\infty} |\mathcal{E}_{u^{(0)}} [\bar{g}(u^{(n)})]|$$

$$= 2 \sup_{u} |\bar{g}(u)| \mathbb{E}^{\rho^{\delta}} \sum_{n=0}^{\infty} |\mathcal{E}_{u^{(0)}} [g(u^{(n)})] - \mathbb{E}^{\rho^{\delta}} [g]|$$

$$\leq 4 \sup_{u} |\bar{g}(u)|^2 \sum_{n=0}^{\infty} \left(1 - \exp(-\Phi^{\star}(r))\right)^n .$$

For the mean square approximation, using the stationarity of the Markov chain conditioned to start in $U \ni u^{(0)} \sim \rho^{\delta}$, we have

$$\frac{1}{M} \mathcal{E}^{\rho^{\delta}} \left[\Big| \sum_{k=1}^{M} \bar{g}(u^{(k)}) \Big|^2 \right] = \mathbb{E}^{\rho^{\delta}} [\bar{g}(u^{(0)})^2] + 2 \frac{1}{M} \sum_{k=1}^{M} \sum_{j=k+1}^{M} \mathcal{E}^{\rho^{\delta}} [\bar{g}(u^{(k)}) \bar{g}(u^{j})]$$

$$= \mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathcal{E}^{\rho^\delta}[\bar{g}(u^{(0)})\bar{g}(u^{(j)})]$$

$$= \mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})^2] + 2\frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=1}^{M-k}\mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})\mathcal{E}_{u^{(0)}}[\bar{g}(u^{(j)})]]$$

$$\leq \mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})^2]$$
$$+ 2\frac{1}{M}\sum_{k=0}^{M-1}\sup_u|\bar{g}(u)|\sum_{j=1}^{M-k}\mathbb{E}^{\rho^\delta}[|\mathcal{E}_{u^{(0)}}[g(u^{(j)})] - \mathbb{E}^{\rho^\delta}[g]|]$$

$$\leq \mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})^2] + 4\frac{1}{M}\sum_{k=0}^{M-1}\sup_u|\bar{g}(u)|^2\sum_{j=1}^{M-k}\Big(1 - \exp(-\Phi^\star(r))\Big)^j$$

$$\leq \mathbb{E}^{\rho^\delta}[\bar{g}(u^{(0)})^2] + 4\sup_u|\bar{g}(u)|^2\sum_{j=1}^{\infty}\Big(1 - \exp(-\Phi^\star(r))\Big)^j,$$

which is clearly bounded uniformly with respect to M. Thus we have shown that there exists $C > 0$ such that for all M

$$\mathcal{E}^{\rho^\delta}\Big[\Big|\frac{1}{M}\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] \leq \frac{C}{M}.$$

It remains to show that the expectation $\mathcal{E}^{\rho^\delta}$ with respect to the unknown posterior ρ^δ can be replaced by an expectation with respect to the prior measure ρ.

To this end we note that

$$\mathcal{E}^{\rho}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] = \int_U \mathcal{E}_{u^{(0)}}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]d\rho(u^{(0)})$$

$$= \int_U \mathcal{E}_{u^{(0)}}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]\frac{d\rho}{d\rho^\delta}(u^{(0)})d\rho^\delta(u^{(0)})$$

$$\leq \mathcal{E}^{\rho^\delta}\Big[\Big|\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big]Z(\delta)\sup_{u\in U}\Big[\exp(\Phi(u;\delta))\Big].$$

As $Z(\delta) \leq 1$ and $\Phi(\cdot;\delta)$ is assumed to be bounded uniformly, we deduce that

$$\mathcal{E}^{\rho}\Big[\Big|\frac{1}{M}\sum_{k=1}^{M}\bar{g}(u^{(k)})\Big|^2\Big] \leq \frac{C}{M},$$

for a constant C independent of M. The conclusion then follows. □

Appendix D: Finite Element Methods

In this Appendix we prove that Assumption 7 holds if we employ a Riesz basis for the Finite Element space and when the domain and the equation's coefficients possess sufficient regularity.

We assume in the following that *the union of all Finite Element basis functions w_j^l of the spaces $V^l = \text{span}\{w_j^l : j = 1, ..., N_l\}$, $l = 0, 1, 2, ...$, constitutes a Riesz*

basis in V. We remark that such bases are available in two and three dimensional polyhedral domains (see, e.g., [24]) (the following assumption of availability of V-stable Riesz bases is made for convenience, and may also be replaced by the availability of a linear complexity, optimal preconditioning, such as a BPX preconditioner).

Assumption 27 *(Riesz Basis Property in V) For each $l \in \mathbb{N}_0$ there exists a set of indices $I^l \subset \mathbb{N}^d$ of cardinality $N_l = O(2^{ld})$ and a family of basis functions $w_k^l \in H_0^1(D)$ indexed by a multi-index $k \in I^l$ such that $V^l = \mathrm{span}\{w_k^l : k \in I^l\}$, and there exist constants c_1 and c_2 which are independent of the discretization level l such that if $w \in V^l$ is written as $w = \sum_{k \in I^l} c_k^l w_k^l \in V^l$, then*

$$c_1 \sum_{k \in I^l} |c_k^l|^2 \leq \|w\|_V^2 \leq c_2 \sum_{k \in I^l} |c_k^l|^2 .$$

Multiscale Finite Element bases entail, in general, larger supports than the standard, single scale basis functions which are commonly used in the Finite Element Method, which implies that the stiffness matrices in these bases have additional nonzero entries, as compared to $O(\dim V^l) = O(2^{dl})$ many nonzero entries of the stiffness matrices that result when one-scale bases, such as the hat functions, are used.

To bound the number of nonzero entries, we shall work under

Assumption 28 *(Support overlap) For all $l \in \mathbb{N}_0$ and for every $k \in I^l$, for every $l' \in \mathbb{N}_0$ the support intersection $\mathrm{supp}(w_k^l) \cap \mathrm{supp}(w_{k'}^{l'})$ has positive measure for at most $O(\max(1, 2^{l'-l}))$ values of k'.*

Assumption 28 implies that the number of non-zero entries in the stiffness matrix of the approximating problem (16) is at most $O(l^{d-1}2^{dl})$. To prove the error bound (17), we require the regularity $P(\cdot, u)$ belonging and being bounded in $H^2(D) \bigcap H_0^1(D)$ uniformly with respect to $u \in U$. Assumption 1(ii) implies the following regularity results.

Proposition 29 *If D is convex and $f \in L^2(D)$, and if Assumptions 1(i),(ii) hold, then, for every $u \in U$, the solution $P^J(\cdot, u)$ of (14) belongs to the space $W := H^2(D) \cap H_0^1(D)$ and there exists a positive constant $C > 0$ such that*

$$\sup_{J \in \mathbb{N}} \sup_{u \in U} \|P^J(\cdot, u)\|_W \leq C\|f\|_{L^2(D)} .$$

Proof By (6), $K^J(x, u) \geq K_{min} > 0$ and we may rewrite the PDE in (14) as

$$-\Delta P^J(x, u) = \frac{1}{K^J(x, u)}(f(x) + \nabla K^J(x, u) \cdot \nabla P^J(x, u)).$$

By our assumptions, the right hand side is uniformly bounded with respect to J and $u \in U$ in the space $L^2(D)$. As the domain D is convex, we deduce that P^J are uniformly bounded with respect to J and $u \in U$ in the space W: it holds

$$
\begin{aligned}
\sup_{u \in U} \|\Delta P^J(\cdot, u)\|_{L^2(D)} &\leq \frac{1}{K_{min}} \sup_{u \in U} \sup_{J \geq 1} \left[\|f\|_{L^2(D)} + \|K^J(\cdot, u)\|_{W^{1,\infty}(D)} \|P^J(\cdot, u)\|_V\right] \\
&\leq C < \infty
\end{aligned}
$$

due to the summability of the $W^{1,\infty}(D)$-norms of the ψ_j implied by Assumption 1(ii). The desired, uniform (w.r. to J and u) bound in the W norm then follows from the L^2 bound on $\Delta P^J(\cdot, u)$ and (7) and the convexity of the domain D. □

Appendix E: Generalized Polynomial Chaos Methods

We justify Assumption 9 in this section. For the ensuing analysis, we shall impose the following assumption on the summability of the gpc expansion of P:

Assumption 30 *There exists a constant $0 < p < 1$ such that the coefficients P_ν of the gpc expansion of P satisfy $(\|P_\nu\|_V)_\nu \in \ell^p(\mathcal{F})$.*

This assumption is valid under the provision of suitable decay of the coefficient functions ψ_j such as Assumption 1(ii). We refer to [5, 6] for details. By a classical argument ("Stechkin's Lemma"), this implies the following, so-called "best N-term approximation property".

Proposition 31 *Under Assumption 30, there exists a nondecreasing sequence $\{\Lambda_N\}_{N\in\mathbb{N}} \subset \mathcal{F}$ of subsets Λ_N whose cardinality does not exceed N, such that*

$$\left\| P - \sum_{\nu\in\Lambda_N} P_\nu L_\nu \right\|_{L^2(U,\rho;V)}^2 = \sum_{\nu\in\mathcal{F}\setminus\Lambda_N} \|P_\nu\|_V^2 \leq CN^{-2\sigma}, \tag{57}$$

where the convergence rate $\sigma = 1/p - 1/2 > 1/2$ and where the constant $C = \|(\|P_\nu\|_V)_{\nu\in\mathcal{F}}\|_{\ell^p(\mathcal{F})}^2$ is bounded independently of N.

The best N-term approximations

$$P_{\Lambda_N} := \sum_{\nu\in\Lambda_N} P_\nu L_\nu \tag{58}$$

in Proposition 31 indicate that sampling the parametric forward map with evaluation of N solutions P_ν, $\nu \in \Lambda_N$ of the parametric, elliptic problem with accuracy $N^{-\sigma}$ is possible; since $\sigma > 1/2$, this is superior to what can be expected from N MC samples. There are, however, two obstacles which obstruct the practicality of this idea: first, the proof of Proposition 31 is nonconstructive, and does not provide concrete choices for the sets Λ_N of "active" gpc coefficients which realize (57) and, second, even if Λ_N were available, the "coefficients" $P_\nu \in V$ can not be obtained exactly, in general, but must be approximated for example by a Finite Element discretization in D.

As $P \in L^2(U, \rho; V)$, we consider the variational form "in the mean" of (4) as

$$\int_U \int_D K(x,u)\nabla P(x,u) \cdot \nabla Q(x,u)dxd\rho(u) = \int_U \int_D f(x)Q(x,u)dxd\rho(u), \tag{59}$$

for all $Q \in L^2(U, \rho; V)$. For each set $\Lambda_N \subset \mathcal{F}$ of cardinality not exceeding N that satisfies Proposition 31, and each vector $\mathcal{L} = (l_\nu)_{\nu\in\Lambda_N}$ of nonnegative integers, we define finite dimensional approximation spaces as

$$X_{N,\mathcal{L}} = \{P_{N,\mathcal{L}} = \sum_{\nu\in\Lambda_N} P_{\nu,\mathcal{L}}(x)L_\nu(u); \; P_{\nu,\mathcal{L}} \in V^{l_\nu}\}. \tag{60}$$

Evidently, $X_{N,\mathcal{L}} \subset L^2(U, \rho; V)$ is a finite-dimensional (hence closed) subspace for any N and any selection $\mathcal{L}$ of the discretization levels.

The total number of degrees of freedom, $N_{dof} = \dim(X_{N,\mathcal{L}})$, necessary for the sparse representation of the parametric forward map is given by

$$N_{\text{dof}} = O\left(\sum_{\nu \in \Lambda_N} 2^{dl_\nu}\right) \quad \text{as} \quad N, \, l_\nu \to \infty \, . \tag{61}$$

The *stochastic, sparse tensor Galerkin approximation* of the parametric forward problem (4), based on the index sets $\Lambda_N \subset \mathcal{F}$, and $\mathcal{L} = \{l_\nu : \nu \in \Lambda_N\}$, reads: find $P_{N,\mathcal{L}} \in X_{N,\mathcal{L}}$ such that for all $Q_{N,\mathcal{L}} \in X_{N,\mathcal{L}}$ holds

$$\begin{aligned} b(P_{N,\mathcal{L}}, Q_{N,\mathcal{L}}) &:= \int_U \int_D K(x, u) \nabla P_{N,\mathcal{L}} \cdot \nabla Q_{N,\mathcal{L}} dx d\rho(u) \\ &= \int_U \int_D f(x) Q_{N,\mathcal{L}}(x, u) dx d\rho(u) \, . \end{aligned} \tag{62}$$

The coercivity of the bilinear form $b(\cdot, \cdot)$ ensures the existence and uniqueness of $P_{N,\mathcal{L}}$ as well as their quasioptimality in $L^2(U, \rho; V)$: by Cea's lemma, for a constant $C > 0$ which is independent of Λ and of $\mathcal{L}$,

$$\|P - P_{N,\mathcal{L}}\|_{L^2(U,\rho;V)} \le C \inf_{Q_{\nu,\mathcal{L}} \in V^{l_\nu}} \|P - \sum_{\nu \in \Lambda} Q_{\nu,\mathcal{L}} L_\nu\|_{L^2(U,\rho;V)} \, .$$

We obtain the following error bound which consists of the error in the best N-term truncation for the gpc expansion and of the Finite Element approximation error for the "active" gpc coefficients.

$$\|P - P_{N,\mathcal{L}}\|^2_{L^2(U,\rho;V)} \le C(N^{-2\sigma} + \sum_{\nu \in \Lambda_N} \inf_{Q_{\nu,\mathcal{L}} \in V^{l_\nu}} \|P_\nu - Q_{\nu,\mathcal{L}}\|^2_V) \, . \tag{63}$$

Let us indicate sufficient conditions that ensure Assumptions 30, 9. The first condition is quantitative decay rate of the coefficient functions ψ_j in the parametric representation (5) of the random input.

To obtain convergence rates for the FE-discretization in the domain D, i.e. of the term $\|P_\nu - Q_{\nu,\mathcal{L}}\|_V$ in (63), it is also necessary to ensure spatial regularity of the solution $P(x, u)$ of the parametric problem (4). To this end we employ Assumptions 1(iii). We remark that Assumptions 1(iii) and (iv) imply Assumption 1(ii) with $q = s - 1 > 0$. Under these assumptions, the following proposition holds.

Proposition 32 *Under Assumptions 1(i),(iii) and (iv) and if, moreover, the domain D is convex and $f \in L^2(D)$, the solution $P(\cdot, u)$ of the parametric, deterministic problem (4) belongs to the space $L^2(U, \rho; W)$.*

From estimate (63), we get with Proposition 32 and standard approximation properties of continuous, piecewise linear FEM the error bound

$$\|P - P_{N,\mathcal{L}}\|^2_{L^2(U,\rho;V)} \le C(N^{-2\sigma} + \sum_{\nu \in \Lambda_N} 2^{-2l_\nu} \|P_\nu\|^2_{H^2(D)}) \, . \tag{64}$$

In order to obtain an error bound in terms of N_{dof} defined in (61) which is uniform in terms of N, we select, for $\nu \in \Lambda_N$ the discretization levels l_ν of the active gpc coefficient P_ν so that both terms in the upper bound (64) are of equal order of magnitude. This

constrained optimization problem was solved, for example, in [5], under the assumption that $(\|P_\nu\|_{H^2(D)})_\nu \in \ell^p(\mathcal{F})$.

In recent years, several algorithms have appeared or are under current development which satisfy Assumption 9 with various exponents $\alpha \geq 1$ and $\beta \geq 0$. See for example the references [3, 27, 11, 2, 4]

Acknowledgments

References

[1] Andrea Barth, Christoph Schwab, and N. Zollinger. Multi-Level Monte Carlo Finite Element Method for elliptic PDEs with stochastic coefficients. *Numerische Mathematik*, 119(1):123–161, 2011.

[2] Marcel Bieri. A sparse composite collocation finite element method for elliptic SPDEs. *SIAM J. Numer. Anal.*, 49(6):2277–2301, 2011.

[3] Marcel Bieri, Roman Andreev, and Christoph Schwab. Sparse tensor discretization of elliptic sPDEs. *SIAM J. Sci. Computing*, 31(6):4281–4304, 2009.

[4] Abdellah Chkifa, Albert Cohen, Ron DeVore, and Christoph Schwab. Sparse adaptive Taylor approximation algorithms for parametric and stochastic elliptic pdes. Technical Report 2011-44, Seminar for Applied Mathematics, ETH Zurich, 2011. (to appear in: Math. Mod. Meth. Appl. Sciences, 2013).

[5] Albert Cohen, Ronald DeVore, and Christoph Schwab. Convergence rates of best n-term Galerkin approximations for a class of elliptic spdes. *Journ. Found. Comp. Math.*, 10(6):615–646, 2010.

[6] Albert Cohen, Ronald Devore, and Christoph Schwab. Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's. *Anal. Appl. (Singap.)*, 9(1):11–47, 2011.

[7] S. Cotter, G. Roberts, A. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science, to appear; arXiv:1202.0709*, 2013.

[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] Michael B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[10] C.J. Gittelson, J. Könnö, Ch. Schwab, and R. Stenberg. The multi-level monte carlo finite element method for a stochastic brinkman problem. Technical Report 2011-31, Seminar for Applied Mathematics, ETH Zürich, 2011. (To appear on Numer. Math. 2013).

[11] Claude Jeffrey Gittelson. *Adaptive Galerkin Methods for Parametric and Stochastic Operator Equations*. PhD thesis, ETH Zürich, 2011. ETH Dissertation No. 19533.

[12] Claude Jeffrey Gittelson. Adaptive wavelet methods for elliptic partial differential equations with

random operators. Technical Report 2011-37, Seminar for Applied Mathematics, ETH Zürich, 2011.

[13] D. Gruhlke. Convergence of multilevel MCMC methods on path spaces. Technical report, PhD Thesis, University of Bonn.

[14] M. Hairer, A. Stuart, and S. Vollmer. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *Ann. Appl. Prob., to appear; arXiv:1112.1392*, 2011.

[15] Viet Ha Hoang and Christoph Schwab. Sparse Tensor Galerkin discretizations for parametric and random parabolic pdes I: Analytic regularity and gpc approximation. *Report 2010-10, Seminar for Applied Mathematics, ETH Zürich*, ((in review)), 2010.

[16] Viet Ha Hoang and Christoph Schwab. Regularity and generalized polynomial chaos approximation of parametric and random 2nd order hyperbolic partial differential equation. *Analysis and Applications*, 2012.

[17] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer, 2004.

[18] C.W. Ketelsen, R. Scheichl, and A.L. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *arXiv preprint*, 2013.

[19] Y. Marzouk and D. Xiu. A stochastic collocation approach to Bayesian inference in inverse problems. 2009.

[20] Y.M. Marzouk and H.N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6):1862–1902, 2009.

[21] Y.M. Marzouk, H.N. Najm, and L.A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2):560–586, 2007.

[22] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[23] Siddhartha Mishra and Christoph Schwab. Sparse tensor multi-level monte carlo finite volume methods for hyperbolic conservation laws with random intitial data. *Report 2010-24, Seminar for Applied Mathematics, ETH Zürich*, ((to appear in Mathematics of Computation (2012))), 2010.

[24] Hoang Nguyen and Rob Stevenson. Finite-Element wavelets on manifolds. *IMA J. Numer. Anal.*, 23(1):149–173, 2003.

[25] Victor Nistor and Christoph Schwab. High order Galerkin approximations for parametric, second order elliptic partial differential equations. Technical Report 2012-21, Seminar for Applied Mathematics, ETH Zürich, 2012. (to appear in: Mathematical Methods and Models in Applied Sciences, 2013).

[26] Claudia Schillings and Christoph Schwab. Sparse, adaptive smolyak algorithms for bayesian inverse problems. Technical Report 2012-37, Seminar for Applied Mathematics, ETH Zürich, 2012. (to appear in: Inverse Problems, 2013).

[27] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[28] Christoph Schwab and A. M. Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4), 2012.

[29] Andrew M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 2010.

[30] L. Tierney. A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Prob.*, 8:1–9, 1998.

[31] S.J. Vollmer. Dimension-independent MCMC sampling for elliptic inverse problems with non-Gaussian priors. *arXiv preprint arXiv:1302.2213*, 2013.