Fast sampling of parameterised Gaussian random fields

Jonas Latz[a], Marvin Eisenberger[b], Elisabeth Ullmann[a]

[a] *Chair of Numerical Analysis, Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching b.M., Germany*
[b] *Chair for Computer Vision and Artificial Intelligence, Department of Computer Science, Technical University of Munich, Boltzmannstr. 3, 85747 Garching b.M., Germany*

Abstract

Gaussian random fields are popular models for spatially varying uncertainties, arising for instance in geotechnical engineering, hydrology or image processing. A Gaussian random field is fully characterised by its mean function and covariance operator. In more complex models these can also be partially unknown. In this case we need to handle a family of Gaussian random fields indexed with hyperparameters. Sampling for a fixed configuration of hyperparameters is already very expensive due to the nonlocal nature of many classical covariance operators. Sampling from multiple configurations increases the total computational cost severely. In this report we employ parameterised Karhunen-Loève expansions for sampling. To reduce the cost we construct a reduced basis surrogate built from snapshots of Karhunen-Loève eigenvectors. In particular, we consider Matérn-type covariance operators with unknown correlation length and standard deviation. We suggest a linearisation of the covariance function and describe the associated online-offline decomposition. In numerical experiments we investigate the approximation error of the reduced eigenpairs. As an application we consider forward uncertainty propagation and Bayesian inversion with an elliptic partial differential equation where the logarithm of the diffusion coefficient is a parameterised Gaussian random field. In the Bayesian inverse problem we employ Markov chain Monte Carlo on the reduced space to generate samples from the posterior measure. All numerical experiments are conducted in 2D physical space, with non-separable covariance operators, and finite element grids with $\sim 10^4$ degrees of freedom.

Keywords: Uncertainty Quantification, Bayesian Inverse Problem, Reduced Basis Methods, Spatial Statistics, Karhunen-Loève expansion
2000 MSC: 65C05, 65N21, 65N25, 60G60, 62M40, 62F15

1. Introduction

Many mathematical models in science and engineering require input parameters which are often not fully known or are perturbed by observational noise. In recent

Email addresses: jonas.latz@ma.tum.de (Jonas Latz), marvin.eisenberger@in.tum.de (Marvin Eisenberger), elisabeth.ullmann@ma.tum.de (Elisabeth Ullmann)

years it has become standard to incorporate the noise or lack of knowledge in a model by using uncertain inputs. In this work we are interested in models based on partial differential equations (PDEs) where the inputs are spatially varying random functions. These so called random fields are characterised by a probability measure on certain function spaces.

We consider two typical tasks in uncertainty quantification (UQ): (*i*) the forward propagation of uncertainty (forward problem) [22, 46], and (*ii*) the (Bayesian) inverse problem [14, 48]. In (*i*) we study the impact of uncertain model inputs on the model outputs and quantities of interest. The mathematical task is to derive the pushforward measure of the model input under the PDE solution operator. In (*ii*) we update a prior distribution of the random inputs using observations; this gives the posterior measure. Mathematically, this amounts to deriving the conditional measure of the inputs given the observations using a suitable version of Bayes' formula. Unfortunately, in most practical cases there are no analytical expressions for either the pushforward or the posterior measure. We focus on sampling based measure approximations, specifically Monte Carlo (MC) for the forward problem, and Markov chain Monte Carlo (MCMC) for the Bayesian inverse problem. Importantly, MC and MCMC require efficient sampling procedures for the random inputs.

In this work we consider Gaussian random fields which are popular models e.g. in hydrology. We recall the following sampling approaches for Gaussian random fields. Factorisation methods construct either a spectral or Cholesky decomposition of the covariance matrix. The major computational bottleneck is the fact that the covariance operator is often nonlocal, and a discretisation will almost always result in a dense covariance matrix which is expensive to handle. Circulant embedding [8, 16, 23] relies on structured spatial grids and stationary covariance operators. In this case, the covariance matrix can be factorised using the Fast Fourier Transform. Alternatively, we can approximate the covariance matrix by hierarchical matrices and low rank techniques, see e.g. [2, 10, 15, 21, 32]. A pivoted Cholesky decomposition is studied in [26]. More recently, the so called SPDE-based sampling has been developed in the works [5, 6, 36, 37, 43]. The major idea is to generate samples of Gaussian fields with Matérn covariance operators by solving a certain discretised fractional PDE with white noise forcing. The Karhunen-Loève (KL) expansion [31, 34] is another popular approach for sampling Gaussian random fields, however, it also suffers from the nonlocal nature of many covariance operators. See e.g. the works [4, 12, 32, 38, 44, 45, 53] for the efficient computation of the KL expansion.

Gaussian random fields are completely characterised by the mean function and covariance operator, and are thus simple models of spatially varying functions. They are also flexible; depending on the regularity of the covariance operator it is possible to generate realisations with different degrees of smoothness. However, in some practical situations the full information on the covariance operator might not be available, e.g. the correlation length, smoothness, and point-wise variance of the random field are not known. Of course, these model parameters can be fixed a priori. However, the posterior measure of a Bayesian inverse problem is often very sensitive to prior information. We illustrate this in the following simple example.

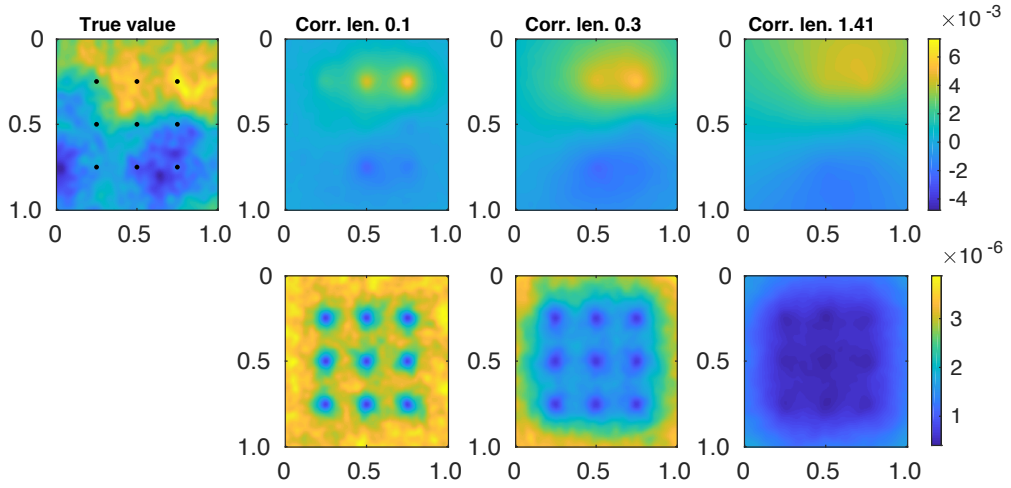

Figure 1.1: Estimation of a Gaussian random field. The top-left figure shows a realisation of the true random field. The task is to estimate this field given 9 noisy point evaluations (black dots). The three top-right (bottom-right) figures show the posterior mean (pointwise posterior variance) for mean-zero Gaussian prior random fields with exponential covariance operator, standard deviation $\sigma = 1$, and correlation lengths $\ell = 0.1, 0.3, 1.41$.

Example 1.1 *We consider a Gaussian random field with exponential covariance operator on the unit square $D = (0,1)^2$. The goal is to estimate the statistics of this field given 9 noisy point observations using Bayesian inversion (cf. §2.4). In Figure 1.1 we plot a realisation of the true field together with the posterior mean and variance for three prior fields with a different (fixed) correlation length each.*

We clearly see in Figure 1.1 that the posterior measure depends crucially on the underlying prior measure and associated correlation length. If the assumed correlation length is too small compared to the truth, then the posterior mean estimate is only accurate close to the observation locations. If, on the other hand, the assumed correlation length is too large, we obtain an overconfident posterior mean estimate. Inaccurate, fixed prior random field parameters can substantially deteriorate the estimation result in Bayesian inverse problems. We treat this problem by modelling unknown *hyperparameters* as random variables.

In statistics, a model with multiple levels of randomness is called a *hierarchical model*. In Bayesian statistics, this means that we work with parameterised prior measures (*hyperpriors*). Hierarchical models in forward and inverse UQ are discussed in [52]. We also refer to [40, Ch.10] for general hierarchical Bayesian analyses. Finally, hierarchical Bayesian inverse problems with spatially varying random fields are discussed in [5, 18, 30, 42, 47, 49]. In our work we consider *parameterised* Gaussian random fields where the covariance operator depends on random variables. Notably, the resulting random field is not necessarily Gaussian, and we can thus model a larger class of spatial variations. However, the greater flexibility of parameterised Gaussian fields brings new computational challenges as we explain next.

Assume that we discretise a Gaussian random field by a KL expansion. The basis functions in this expansion are the eigenfunctions of the covariance operator. For

fixed, deterministic hyperparameters it is sufficient to compute the KL eigenpairs only once since the covariance operator is fixed. However, changing the hyperparameters changes the covariance and often requires to re-compute the KL eigenpairs. The associated computational cost and memory requirement scales at least quadratically in the number of spatial unknowns. Hence it is often practically impossible to use uncertain hyperparameters in a (Gaussian) random field model in 2D or 3D physical space. To overcome this limitation we suggest and study a reduced basis (RB) surrogate for the efficient computation of parameter dependent KL expansions. In [47] the authors use a polynomial chaos surrogate for this task. In contrast, our reduced basis surrogate approximates the KL eigenpairs by a linear combination of snapshot eigenpairs associated with certain fixed values of the hyperparameters. Since this requires the solution of eigenproblems in a small number of unknowns (which is equal to the number of reduced basis functions) the computational cost can be reduced significantly.

Reduced basis methods were introduced in [35], and are typically used to solve PDEs for a large number of parameter configurations, see e.g. [27, 39]. In contrast, parameterised eigenproblems have not been treated extensively with reduced basis ideas. We refer to [28, 29] for applications and reviews of reduced basis surrogates to parameterised eigenproblems with PDEs. For non-parametric KL eigenproblems we mention that reduced basis methods have been combined with domain decomposition ideas. In this situation we need to solve several low-dimensional eigenproblems on the subdomains in the physical space. It is then possible to construct an efficient reduced basis by combining the subdomain solutions, see [12] for details.

We point out that many differential operators are local, and hence the associated discretised operator is often sparse. Linear equations or eigenvalue problems with sparse coefficient matrices can often be solved with a cost that scales linearly in the number of unknowns. In contrast, the reduced linear system matrix is often dense and the solution cost is at least quadratic in the number of unknowns. Hence, for reduced basis methods to be cheaper compared to solves with the full discretised PDE operator it is necessary that the size of the reduced basis is not too large compared to the number of unknowns of the full system. Notably, in most cases the discretised KL eigenproblem results in a dense matrix since the covariance integral operator is nonlocal. Hence we expect a significant reduction of the total computational cost even if the size of the reduced basis is only slightly smaller than the number of unknowns in the unreduced eigenspace.

The remainder of this report is organised as follows. In §2 we establish the mathematical theory for working with parameterised Gaussian random fields. In particular, we discuss the well-posedness of a hierarchical forward and inverse problem, respectively, with an elliptic PDE where the coefficient function is a random field with uncertain hyperparameters. In §3 we describe sampling algorithms to approximate the solution of hierarchical forward and inverse problems. In §4 we propose and study a reduced basis surrogate for the parametric KL eigenpairs which arise from the sampling methods in §3. Specifically, we consider Matérn-type covariance operators, and suggest a linearisation to enable an efficient offline-online decomposition within the reduced basis solver. In §5 we introduce an efficient *reduced basis*

sampling, and analyse its computational cost and memory requirements. Finally, we present results of numerical experiments in §6. We study the approximation accuracy and speed-up of the reduced basis surrogate, and illustrate the use of the reduced basis sampling for hierarchical forward and Bayesian inverse problems.

2. Mathematical framework

To begin, we introduce the notation and some key elements, in particular, Gaussian and parameterised Gaussian measures. Next, we describe a general setting for forward uncertainty propagation and an associated Bayesian inverse problem. Importantly, we investigate the well-posedness of these problems if the uncertain elements are modeled by *parameterised* Gaussian random fields. This is a necessary extension of the by now well-established solution theory for Gaussian random inputs.

2.1. Gaussian measures

Let $(\Omega, \mathcal{A}, \mathbb{P})$ denote a probability space. We recall the notion of a real-valued Gaussian random variable.

Definition 2.1 *The random variable $\xi : \Omega \to \mathbb{R}$ follows a non-degenerate Gaussian measure, if*

$$\mathbb{P}(\xi \leq x) := \mathrm{N}(a, b^2)((-\infty, x]) := \int_{-\infty}^{x} \frac{1}{(2\pi)^{1/2} b} \exp\left(-\frac{(x-a)^2}{b^2}\right) \mathrm{d}x, \quad x \in \mathbb{R},$$

for some $a \in \mathbb{R}$ and $b > 0$. The Gaussian measure is degenerate, if $b = 0$. In this case, we define $\mathrm{N}(a, 0) := \delta_a$, the Dirac measure concentrated in a.

Let X denote a separable $\mathbb{R}$-Hilbert space with Borel-σ-algebra $\mathcal{B}X$. We now introduce Gaussian measures on X.

Definition 2.2 *The X-valued random variable $\theta : \Omega \to X$ has a Gaussian measure, if $T\theta$ follows a Gaussian measure for any $T \in X^*$ in the dual space of X. We write $\theta \sim \mathrm{N}(m, \mathcal{C})$, where*

$$\begin{aligned} m = \mathbb{E}[\theta] &:= \int \theta \mathrm{d}\mathbb{P}, \\ \mathcal{C} = \mathrm{Cov}(\theta) &:= \mathbb{E}[(\theta - m) \otimes (\theta - m)]. \end{aligned}$$

In Definition 2.2 we distinguish two cases. If X is finite-dimensional, then we call θ a *(multivariate) Gaussian random variable* with mean vector m and covariance matrix $\mathcal{C}$. If X is infinite-dimensional, then θ is called *Gaussian random field* with mean function m and covariance operator $\mathcal{C}$.

We now discuss two options to construct a Gaussian measure with given mean function $m \in X$ and covariance operator $\mathcal{C} : X \to X$. While any $m \in X$ can be used as a mean function, we assume that $\mathcal{C}$ is a linear, continuous, trace-class, positive semidefinite and self-adjoint operator. This ensures that $\mathcal{C}$ is a valid covariance

operator, see [1, 33]. We denote the set of valid covariance operators on X by $\mathrm{CO}(X)$.

If $\dim X < \infty$, we can identify a Gaussian measure on X in terms of a probability density function w.r.t. the Lebesgue measure.

Proposition 2.3 *Let $X := \mathbb{R}^N$, $m \in X$ and $\mathcal{C} \in \mathrm{CO}(X)$ with full rank. Then*

$$\mathrm{N}(m, \mathcal{C})(B) = \int_B \mathrm{n}(\theta; m, \mathcal{C}) \mathrm{d}\theta, \quad B \in \mathcal{B}X \tag{2.1}$$

where

$$\mathrm{n}(\theta; m, \mathcal{C})(B) = \det(2\pi\mathcal{C})^{-1/2} \exp\left(-(1/2)\langle \theta - m, \mathcal{C}^{-1}(\theta - m)\rangle\right) \tag{2.2}$$

is its probability density function.

If $\dim X = \infty$, we can construct a Gaussian measure on X using the so-called *Karhunen-Loève (KL) expansion.*

Definition 2.4 *Let $\dim X = \infty$ and let $(\lambda_i, \psi_i)_{i=1}^{\infty}$ denote the eigenpairs of $\mathcal{C}$, where $(\psi_i)_{i=1}^{\infty}$ form an orthonormal basis of X. Let $\xi : \Omega \to \mathbb{R}^{\infty}$ be a measurable function. Furthermore, let the components of ξ form a sequence $(\xi_i)_{i=1}^{\infty}$ of independent and identically distributed (i.i.d.) random variables, where $\xi_1 \sim \mathrm{N}(0, 1)$. Then, the expansion*

$$\theta_{\mathrm{KL}} := m + \sum_{i=1}^{\infty} \sqrt{\lambda_i} \xi_i \psi_i$$

is called KL expansion.

One can easily verify the following proposition, see [31, 34].

Proposition 2.5 *θ_{KL} is distributed according to $\mathrm{N}(m, \mathcal{C})$.*

In the remainder of this paper we assume that all eigenpairs are ordered descendantly with respect to the absolute value of the associated eigenvalues. For illustration purposes we give a example for a Gaussian measure on an infinite-dimensional, separable Hilbert space.

Example 2.6 *Let $X := \mathcal{L}^2(D; \mathbb{R})$, where $D \subseteq \mathbb{R}^d$ ($d = 1, 2, 3$) is open, bounded and connected. We define the exponential covariance operator with correlation length $\ell > 0$ and standard deviation $\sigma > 0$ as follows,*

$$\begin{aligned} \mathcal{C}_{\mathrm{exp}}^{(\ell,\sigma)} &: X \to X, \ \varphi \mapsto \int_D c_{\mathrm{exp}}^{(\ell,\sigma)}(x, \cdot)\varphi(x)\mathrm{d}x, \\ c_{\mathrm{exp}}^{(\ell,\sigma)} &: D \times D \to \mathbb{R}, \ (x, y) \mapsto \sigma^2 \exp\left(-\ell^{-1}\mathrm{dist}(x, y)\right), \end{aligned} \tag{2.3}$$

where dist *is the metric induced by the 2-norm.*

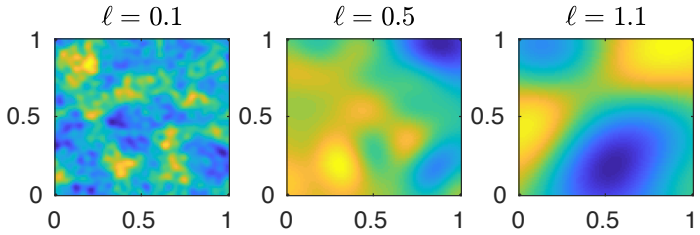

Figure 2.1: Samples of mean-zero Gaussian random fields with exponential covariance, $\sigma = 1$ and $\ell \in \{0.1, 0.5, 1.1\}$.

We now briefly discuss the implications of ℓ and σ on a (mean-zero) Gaussian measure with exponential covariance operator $\mathrm{N}(0, \mathcal{C}_{\mathrm{exp}}^{(\ell,\sigma)})$ as in (2.3). The correlation length ℓ models the strength of the correlation between the random variables $\theta(x)$ and $\theta(y)$ in two points $x, y \in D$ in the random field $\theta \sim \mathrm{N}(0, \mathcal{C}_{\mathrm{exp}}^{(\ell,\sigma)})$. If ℓ is small, the marginals $\theta(x)$ and $\theta(y)$ are only strongly correlated if $\mathrm{dist}(x, y)$ is small. Otherwise, if ℓ is large, then $\theta(x)$ and $\theta(y)$ are also strongly correlated, if $\mathrm{dist}(x, y)$ is large. In Figure 2.1, we plot samples of mean-zero Gaussian measures with exponential covariance and different correlation lengths. The pointwise standard deviation σ is constant for all $x \in D$. One can show that the realisations of θ are a.s. continuous, independently of ℓ and σ.

2.2. Parameterised Gaussian measures

We now construct parameterised measures. We denote the space of parameters by $R \subseteq \mathbb{R}^{\mathrm{N}_R}$ and assume that R is non-empty and finite-dimensional. Moreover, we assume that $(R, \mathcal{R})$ is a measurable space. In the following we denote an element of R by τ_0 and a random variable taking values in R by $\tau : \Omega \to R$. We recall the notion of a *Markov kernel*; this is a measure-theoretically consistent object for the representation a parameterised probability measure.

Definition 2.7 *A Markov kernel from $(R, \mathcal{R})$ to $(X, \mathcal{B}X)$ is a function $K : \mathcal{B}X \times R \to [0, 1]$, where*

1. *$K(B|\cdot) : R \to [0, 1]$ is measurable for all $B \in \mathcal{B}X$,*
2. *$K(\cdot|\tau_0) : \mathcal{B}X \to [0, 1]$ is a probability measure for all $\tau_0 \in R$.*

We consider Markov kernels where $K(\cdot|\tau_0)$ is a Gaussian measure for all $\tau_0 \in R$. Hence, we define a parameterised Gaussian measure in terms of a Markov kernel from $(R, \mathcal{R})$ to $(X, \mathcal{B}X)$. Particularly, we define

$$R \ni \tau_0 \mapsto K(\cdot|\tau_0) := \mathrm{N}(m(\tau_0), \mathcal{C}(\tau_0)), \tag{2.4}$$

where

$$m : R \to X, \quad \mathcal{C} : R \to \mathrm{CO}(X)$$

are measurable functions. The family of Gaussian measures in (2.4) does not define a valid parameterised measure yet, it remains to identify a measure on the parameter

space R. To this end let $\tau : \Omega \to R$ be an R-valued random variable. We assume that τ is distributed according to some probability measure μ'. Now, let $\theta : \Omega \to X$ be an X-valued random variable. We assume that $\theta \sim K(\cdot|\tau)$. Then, the joint measure of $(\tau, \theta) : \Omega \to R \times X$ is given by

$$\mu := \mathcal{R} \otimes \mathcal{B}X \ni B \mapsto \iint_B K(\mathrm{d}\theta|\tau)\mu'(\mathrm{d}\tau) \in [0,1]. \tag{2.5}$$

The measure μ in (2.5) models a two-step sampling mechanism. Indeed, to generate a sample $(\tau, \theta) \in R \times X$ with distribution μ we proceed as follows:

1. Sample $\tau \sim \mu'$,
2. Sample $\theta \sim K(\cdot|\tau)$.

Finally, let μ'' be the marginal of μ w.r.t. to θ, i.e.,

$$\mu''(B_2) := \mu(R \times B_2), \qquad B_2 \in \mathcal{B}X. \tag{2.6}$$

Alternatively, μ'' can be defined in terms of the composition of μ' and K, $\mu'' := \mu' K$. Note that μ'' is a measure on X, and that the Markov kernel K is the parameterised Gaussian measure which we wanted to construct.

Remark 2.8 *We point out that even if $K(\cdot|\tau_0)$ is a Gaussian measure for any $\tau_0 \in R$, the marginal μ'' is not necessarily a Gaussian measure. We give two examples.*

(a) Let $R := X := \mathbb{R}$, let μ' be a Gaussian measure and $K(\cdot|\tau_0) := \mathrm{N}(\tau_0, \sigma^2)$. Then μ'' is a Gaussian measure. Indeed, this construction models a family of Gaussian random variables where the mean value is Gaussian.

(b) Let R be a finite set. Then μ'' is called Gaussian mixture and is typically not Gaussian. See §1.1 and §2.1 in [20].

Now, we return to Example 2.6 and construct an associated Markov kernel.

Example 2.9 *We consider again the exponential covariance operator that is given in (2.3). Let $\underline{\ell} > 0$ and $\overline{\sigma} > \underline{\sigma} > 0$. For any $\ell \in [\underline{\ell}, \mathrm{diam}(D)]$ and $\sigma \in [\underline{\sigma}, \overline{\sigma}]$, one can show that $\mathcal{C}_{\exp}^{(\ell,\sigma)} \in \mathrm{CO}(X)$ is a valid covariance operator. The parameters $\tau = (\ell, \sigma)$ are random variables on a non-empty set $R := [\underline{\ell}, \mathrm{diam}(D)] \times [\underline{\sigma}, \overline{\sigma}]$. The associated probability measure μ' is given by*

$$\mu' := \mu'_\ell \otimes \mu'_\sigma.$$

Here, μ'_ℓ is given such that $\ell^{-1} \sim \mathrm{Unif}[\mathrm{diam}(D)^{-1}, \underline{\ell}^{-1}]$ and μ'_σ is a Gaussian measure that is truncated outside of $[\underline{\sigma}, \overline{\sigma}]$. σ models the standard deviation in every single point $\theta(x)$, for any $x \in D$. The measure μ' and the Markov kernel $K(\cdot|\ell, \sigma) = \mathrm{N}(0, \mathcal{C}_{\exp}^{(\ell,\sigma)})$ induce a joint measure μ. This can now be understood as follows:

1. *Sample τ from μ':*
   (a) Sample the correlation length $\ell \sim \mu'_\ell$,

(b) Sample the standard deviation $\sigma \sim \mu'_\sigma$.

2. *Sample the random field $\theta \sim K(\cdot|\ell, \sigma)$ with exponential covariance operator, standard deviation σ and correlation length ℓ.*

Hence, we modelled a Gaussian random field with exponential covariance, where the correlation length and standard deviation are unknown.

In the following sections we study parameterised Gaussian measures in the setting of forward uncertainty propagation and Bayesian inversion.

2.3. Forward uncertainty propagation

We consider a classical model problem given by the elliptic partial differential equation (PDE)

$$-\nabla \cdot (\exp(\theta(x))\nabla p(x)) = f(x) \qquad \forall x \in D, \tag{2.7}$$

plus suitable boundary conditions. The PDE in (2.7) models the stationary, single-phase, incompressible flow of fluid in a permeable domain $D \subset \mathbb{R}^d$, combining Darcy's law and a mass balance equation. p is the fluid pressure, f is a source term, and θ is the spatially varying log-permeability of the fluid reservoir. We assume that D is connected, open and bounded, and that θ is an element of a separable Hilbert space X. We define the PDE solution operator

$$\mathcal{S} : X' \to H_0^1(D), \qquad \theta \mapsto p = \mathcal{S}(\theta),$$

where $X' \subseteq X$. If $X = \mathcal{L}^2(D; \mathbb{R})$, a typical choice for X' is the separable Banach space of continuous functions $C^0(D) := \{f | f \colon \overline{D} \to \mathbb{R} \text{ continuous}\}$.

We now consider θ to be uncertain. We model the spatially varying uncertainty in θ by assuming that θ is a (parameterised) Gaussian random field with continuous realisations almost surely (a.s.). Forward propagation of uncertainty means in our setting that we want to solve the elliptic PDE (2.7) with random coefficient function θ. For a well-posed problem the solution of (2.7) is a probability measure on $H_0^1(D)$.

We often consider a scalar-valued quantity of interest $Q : H_0^1(D) \to \mathbb{R}$ instead of the full space $H_0^1(D)$. For convenience we define $\mathcal{Q} = Q \circ \mathcal{S}$ which maps the uncertain parameter directly to the quantity of interest. The forward propagation of uncertainty is modeled by the push-forward measure

$$\mathcal{Q}^\# \mu'' := \mu''(\mathcal{Q} \in \cdot) := \mathbb{P}(\mathcal{Q}(\theta) \in \cdot),$$

where $\theta \sim \mu''$. If μ'' is a Gaussian measure, the well-posedness and properties of $\mathcal{Q}^\# \mu''$ have been studied in e.g. in [9, 19]. In the following, we extend this theory and discuss the existence of $\mathcal{Q}^\# \mu''$ with respect to the possibly non-Gaussian measure μ''. Moreover, we also study the existence of moments of $\mathcal{Q}^\# \mu''$. To this end we make the following assumptions.

Assumptions 2.10 *(a) $\mathcal{Q}^\# K(\cdot|\tau_0) := K(\mathcal{Q} \in \cdot|\tau_0) := \mathbb{P}(\mathcal{Q}(\theta) \in \cdot|\tau = \tau_0)$ is well-defined for μ'-almost every $\tau_0 \in R$.*

(b) For some $k \in \mathbb{N}$ it holds

$$m_k(\tau_0) := \int \mathcal{Q}(\theta)^k K(\mathrm{d}\theta|\tau_0) < \infty$$

for μ'-almost every $\tau_0 \in R$ and $\int m_k(\tau)\mu'(\mathrm{d}\tau) < \infty$.

Theorem 2.11 *Let Assumptions 2.10 hold. Then, the measure $\mathcal{Q}^{\#}\mu''$ is well-defined. Moreover, $\int \mathcal{Q}(\theta)^k \mu''(\mathrm{d}\theta) < \infty$ where $k \in \mathbb{N}$ is as in Assumption 2.10(b).*

Proof. By assumption, $\mathcal{Q}^{\#}K(\cdot|\tau_0)$ is well-defined and a probability measure for μ'-almost every $\tau_0 \in R$. Hence,

$$\mathcal{Q}^{\#}\mu'' = \int_R K(\cdot|\tau)\mu'(\mathrm{d}\tau)$$

is well-defined and a probability measure. The finiteness of the moments can be shown analogously. □

Typically, $\mathcal{Q}^{\#}\mu''$ cannot be computed analytically. We discuss the approximation of this measure by standard Monte Carlo in §3.3.

2.4. Bayesian inverse problem

Again, we consider the PDE (2.7) with a random coefficient $\theta \sim \mu_0$. First, we assume that μ_0 is a Gaussian measure and discuss the standard ('Gaussian') case of a Bayesian inverse problem.

Assume that we have collected noisy observational data. We want to use these observations to learn the uncertain parameter θ. We proceed by the Bayesian approach to inverse problems and compute the conditional probability measure of θ given the observations. This measure is called *posterior measure* - it reflects the knowledge about θ after observing the data. In contrast, the knowledge about θ without observations is modelled by μ_0, the *prior measure*.

Let $\mathcal{O} : H_0^1 \to \mathbb{R}^{N_{\mathrm{obs}}}$ be a linear operator. It is called *observation operator* and maps the PDE solution to the observations. $\mathcal{G} := \mathcal{O} \circ \mathcal{S}$ is called *forward response operator* and maps the uncertain parameter θ to the observations. Let $y \in \mathbb{R}^{N_{\mathrm{obs}}}$ denote the observations. We assume that the observations y are perturbed by additive Gaussian noise $\eta \sim \mathrm{N}(0, \Gamma)$, where $\Gamma \in \mathbb{R}^{N_{\mathrm{obs}} \times N_{\mathrm{obs}}}$ is a symmetric and positive definite matrix. η and θ are statistically independent. The event that occurs while we observe y is $\{\mathcal{G}(\theta) + \eta = y\} \in \mathcal{B}X$. The posterior measure is given by $\mathbb{P}(\theta \in \cdot|\mathcal{G}(\theta) + \eta = y)$. It can be derived using Bayes' formula which we introduce later in this section.

If Assumptions A.2 (that are identical to Assumption 2.6 from [48]) hold, and if the prior measure μ_0 is Gaussian, then one can show that a posterior measure exists, that it is uniquely defined and that it is locally Lipschitz-continuous w.r.t the data $y \in \mathbb{R}^{N_{\mathrm{obs}}}$. These statements refer to the space of probability measures on $(X, \mathcal{B}X)$

that are absolutely continuous w.r.t. μ_0. This space, equipped with the Hellinger distance, is a metric space. We denote it by $(\mathbb{P}(X, \mathcal{B}X, \mu_0), \mathrm{d}_{\mathrm{Hel}})$, where

$$\mathrm{d}_{\mathrm{Hel}}(\nu_1, \nu_2) := \sqrt{\int \left(\sqrt{\frac{\mathrm{d}\nu_1}{\mathrm{d}\mu_0}(\theta)} - \sqrt{\frac{\mathrm{d}\nu_2}{\mathrm{d}\mu_0}(\theta)} \right)^2 \mu_0(\mathrm{d}\theta)}.$$

Hence, the problem of finding a posterior measure on $\mathbb{P}(X, \mathcal{B}X, \mu_0)$ is well-posed in the sense of Hadamard [25]. We give more details in Appendix A.

The choice of the prior measure μ_0 has a considerable impact on the posterior. We discussed this in §1, see also Figure 1.1. However, it is often not possible to determine the prior sufficiently. Hence it is sensible to assume that the prior measure is (partially) unknown. μ_0 is then called *hyperprior* and can be modelled by a Markov kernel (see §2.2). Here, we consider $\mu_0(\cdot|\tau) := K(\cdot|\tau)$ where τ is a random variable and μ' is the prior measure of τ. The posterior measure in this so-called *hierarchical Bayesian inverse problem* is then

$$\mu^y := \mathbb{P}((\tau, \theta) \in \cdot | \mathcal{G}(\theta) + \eta = y). \tag{2.8}$$

It can be determined using Bayes' formula

$$\mu^y(B) = Z_y^{-1} \iint_B \exp(-\Phi(\theta)) \mu_0(\mathrm{d}\theta|\tau) \mu'(\mathrm{d}\tau), \tag{2.9}$$

where

$$\begin{aligned} B &\in \mathcal{R} \otimes \mathcal{B}X, \\ Z_y &:= \iint_{R \times X} \exp(-\Phi(\theta)) \mu(\mathrm{d}\theta|\tau) \mu'(\mathrm{d}\tau), \\ \Phi(\theta) &:= \frac{1}{2} \|\Gamma^{1/2}(\mathcal{G}(\theta) - y)\|_2^2. \end{aligned}$$

We will now show that the posterior measure based on a hyperprior is well-defined and that the Bayesian inverse problem in this setting is well-posed.

Theorem 2.12 *Let μ^{y,τ_0} be the posterior measure of the Bayesian inverse problem*

$$\mathcal{G}(\theta) + \eta = y,$$

using $\mu_0(\cdot|\tau_0)$ as a prior measure, where $\tau_0 \in R$ is fixed. In particular, let

$$\begin{aligned} \mu^{y,\tau_0}(B) &:= Z_{y,\tau_0}^{-1} \int_B \exp(-\Phi(\theta)) \mu_0(\mathrm{d}\theta|\tau_0), \quad B \in \mathcal{B}X, \\ Z_{y,\tau_0} &:= \int_X \exp(-\Phi(\theta)) \mu_0(\mathrm{d}\theta|\tau_0). \end{aligned}$$

We assume that the computation of μ^{y,τ_0} is well-posed for μ'-almost every $\tau_0 \in R$ and that a constant $L(r, \tau_0) \in (0, \infty)$ exists for μ'-almost all $\tau_0 \in R$, such that

$$\mathrm{d}_{\mathrm{Hel}}(\mu^{y_1,\tau_0}, \mu^{y_2,\tau_0}) \leq L(r, \tau_0) \|y_1 - y_2\|_2, \tag{2.10}$$

for any two datasets y_1, y_2, where $\max\{\|y_1\|_2, \|y_2\|_2\} < r$. Moreover, we assume that $L(r,\cdot) \in \mathcal{L}^2(R, \mathcal{R}; \mathbb{R})$ for all $r > 0$, and that $Z_{y,(\cdot)} \in \mathcal{L}^1(R, \mathcal{R}; \mathbb{R})$. Then, the computation of μ^y is well-posed.

Proof. By Bayes' Theorem, the posterior measure $\mu^y = \iint_{(\cdot)} \mu^{y,\tau}(\mathrm{d}\theta)\mu'(\mathrm{d}\tau)$ is well-defined and unique if the normalising constant Z_y is positive and finite. By assumption it holds $Z_{y,(\cdot)} \in (0, \infty)$. Hence,

$$Z_y = \int \underbrace{Z_{y,\tau}}_{>0} \mu'(\mathrm{d}\tau) > 0.$$

Furthermore, also by assumption we have

$$Z_y = \int Z_{y,\tau}\mu'(\mathrm{d}\tau) = \|Z_{y,(\cdot)}\|_1 < \infty.$$

Let $y_1, y_2 \in \mathbb{R}^{N_{\mathrm{obs}}}$ be a two datasets and let $\tau_0 \in R$. By assumption, a constant $L(\tau) \in (0, \infty)$ exists, such that

$$\mathrm{d}_{\mathrm{Hel}}(\mu^{y_1,\tau_0}, \mu^{y_2,\tau_0}) \leq L(\tau_0)\|y_1 - y_2\|_2. \tag{2.11}$$

Now, we have

$$\mathrm{d}_{\mathrm{Hel}}(\mu^{y_1}, \mu^{y_2})^2 = \iint \left(\sqrt{\frac{\mathrm{d}\mu^{y_1}}{\mathrm{d}\mu}(\theta,\tau)} - \sqrt{\frac{\mathrm{d}\mu^{y_2}}{\mathrm{d}\mu}(\theta,\tau)}\right)^2 \mu(\mathrm{d}\tau, \mathrm{d}\theta),$$

where

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu}(\theta,\tau) := Z_y^{-1}\exp(-\Phi(\theta)), \quad y \in \{y_1, y_2\}.$$

This density is constant in τ_0. Let $y \in \{y_1, y_2\}$. The density $\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu}(\theta, \tau_0)$ is identical to the density in the Bayesian inverse problem using $\mu_0(\cdot|\tau_0)$ as a prior with a fixed $\tau_0 \in R$. The latter is given by $\frac{\mathrm{d}\mu^{y,\tau_0}}{\mathrm{d}\mu_0(\cdot|\tau_0)}(\theta)$. Hence,

$$\begin{aligned}
\mathrm{d}_{\mathrm{Hel}}(\mu^{y_1}, \mu^{y_2})^2 &= \iint \left(\sqrt{\frac{\mathrm{d}\mu^{y_1,\tau}}{\mathrm{d}\mu_0(\cdot|\tau)}(\theta)} - \sqrt{\frac{\mathrm{d}\mu^{y_2,\tau}}{\mathrm{d}\mu_0(\cdot|\tau)}(\theta)}\right)^2 \mu(\mathrm{d}\tau, \mathrm{d}\theta) \\
&\leq \int \mathrm{d}_{\mathrm{Hel}}(\mu^{y_1,\tau}, \mu^{y_2,\tau})^2 \mu'(\mathrm{d}\tau) \\
&\leq \int L(\tau)^2 \|y_1 - y_2\|_2^2 \mu'(\mathrm{d}\tau) \\
&= \int L(\tau)^2 \mu'(\mathrm{d}\tau)\|y_1 - y_2\|_2^2 < \infty.
\end{aligned}$$

The right-hand side is finite, since L is square integrable. This implies that the computation of μ^y is well-posed. □

3. Computational framework

In this section we discuss the sampling from parameterised Gaussian measures and the respective Markov kernels. We begin with non-parameterised Gaussian measures though. We briefly discuss the spatial discretisation of random fields using finite elements. Once the spatial discretisation is fixed we only need to sample from a (high-dimensional) multivariate Gaussian measure. We review how this can be done using the Cholesky decomposition, (truncated) KL expansion and circulant embedding. Then, we analyse the computational costs when using the truncated KL expansion to sample from a parameterised Gaussian random field. We conclude this section by revisiting the forward and inverse problem, respectively, and describe sampling based methods to approximate their solutions. Both these methods require repeated sampling from a family of Gaussian random fields with different means and covariance operators each.

3.1. Sampling of Gaussian random fields

Consider a Gaussian random field $\mathrm{N}(m, \mathcal{C})$ on $(X, \mathcal{B}X)$. For practical computations the infinite-dimensional parameter space X must be discretised. Here, we use finite elements. Let $\mathrm{B}_h := (\varphi_i : i = 1, ..., N) \in X^N$ denote an N-dimensional basis of a finite element space. We approximate X by $X_h := \mathrm{span}(\mathrm{B}_h)$. Note that we can identify $X_h \cong \mathbb{R}^N$.

Let $\langle \cdot, \cdot \rangle$ denote the Euclidean inner product on $\mathbb{R}^N$. Note that $\mathbb{R}^N$ is a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle_M = \langle \cdot, M \cdot \rangle$, where $M = \mathrm{B}_h^* \mathrm{B}_h$ is the Gramian matrix associated with the finite element basis B_h. The Gaussian measure $\mathrm{N}(m, \mathcal{C})$ on $(X, \mathcal{B}X)$ can then be represented on $\mathbb{R}^N$ by the measure $\mathrm{N}(\mathrm{B}_h^* m, \mathrm{B}_h^* \mathcal{C} \mathrm{B}_h)$ with mean vector $\mathrm{B}_h^* m$ and covariance matrix $\mathrm{B}_h^* \mathcal{C} \mathrm{B}_h$.

From now on assume that $X := \mathbb{R}^N$ is a finite dimensional (and thus separable) Hilbert space with inner product $\langle \cdot, \cdot \rangle_X$. Moreover, we assume that the Gaussian measure $\mathrm{N}(m, \mathcal{C})$ is mean-zero, that is $m \equiv 0$. A simple sampling strategy uses the Cholesky decomposition LL^* of $\mathcal{C}$. Let $\xi \sim \mathrm{N}(0, \mathrm{Id}_N)$. Then, it is easy to see that $L\xi \sim \mathrm{N}(0, \mathcal{C})$. The computational cost of a Cholesky decomposition is $O(N^3; N \to \infty)$.

Alternatively, we can use the KL expansion (see Definition 2.4)

$$\sum_{i=1}^{N} \sqrt{\lambda_i} \xi_i \psi_i \sim \mathrm{N}(0, \mathcal{C}),$$

where $(\lambda_i, \psi_i)_{i=1}^N$ are the eigenpairs of $\mathcal{C}$. Recall that the eigenvectors form an orthonormal basis of X and that $\xi \sim \mathrm{N}(0, \mathrm{Id}_N)$. Computing the spectral decomposition of a symmetric matrix is typically more expensive than computing the Cholesky decomposition. However, there are cases when the spectral decomposition can be computed cheaply. Here, we briefly describe *circulant embedding*, see [16]. Assume that $\mathcal{C}$ is Toeplitz (in 1D space) or block Toeplitz with Toeplitz blocks (in 2D space). Then, $\mathcal{C}$ can be embedded into a positive definite circulant matrix (in 1D) or into a positive definite block circulant with circulant blocks matrix (in 2D)

$\mathcal{C}_{\text{circ}} \in \mathbb{R}^{N_{\text{circ}} \times N_{\text{circ}}}$. The eigenvectors of this matrix are given by the discrete Fourier transform. Hence, eigenvectors and eigenvalues can be computed by the FFT in $O(N_{\text{circ}} \log N_{\text{circ}})$.

The KL expansion offers a natural way to reduce the stochastic dimension of an X-valued Gaussian random variable by simply truncating the expansion. This can be interpreted as dimension reduction from a high-dimensional to a low-dimensional stochastic space. A reduction from an infinite-dimensional to a finite-dimensional stochastic space is also possible. Let $N_{\text{sto}} \in \mathbb{N}, N_{\text{sto}} < N$ and consider the truncated KL expansion

$$\theta_{\text{KL}}^{N_{\text{sto}}} := \sum_{i=1}^{N_{\text{sto}}} \sqrt{\lambda_i} \xi_i \psi_i.$$

The remaining eigenvalues of the truncated KL expansion give an error bound in $\mathcal{L}^2$,

$$\mathbb{E}\left[\|\theta_{\text{KL}} - \theta_{\text{KL}}^{N_{\text{sto}}}\|_X^2\right] = \sum_{i=N_{\text{sto}}+1}^{N} \lambda_i.$$

Furthermore, the truncated KL expansion solves the minimisation problem

$$\min_{\widehat{\theta} \in \mathcal{L}^2(\mathbb{R}^{N_{\text{sto}}}; X)} \mathbb{E}\left[\|\theta_{\text{KL}} - \widehat{\theta}\|_X^2\right],$$

for any given $N_{\text{sto}} \in \mathbb{N}$. Hence, the truncated KL expansion $\theta_{\text{KL}}^{N_{\text{sto}}}$ is the optimal N_{sto}-dimensional function which approximates θ_{KL}. In the statistics literature this method is known by the name *principal component analysis*.

Observe that $\theta_{\text{KL}}^{N_{\text{sto}}}$ is a Gaussian random field on X with covariance operator

$$\mathcal{C}^{N_{\text{sto}}} := \sum_{i=1}^{N_{\text{sto}}} \lambda_i (\psi_i \otimes \psi_i). \tag{3.1}$$

This covariance operator has rank $\leq N_{\text{sto}} < N$. Sampling with the truncated KL expansion requires only the leading N_{sto} eigenpairs. We assume that this reduces the computational cost asymptotically to $O(N^2 N_{\text{sto}}; N \to \infty)$. We discuss this in more detail in §3.2. In the remainder of this report, we generate samples with the truncated KL expansion.

3.2. Sampling of parameterised Gaussian random fields

We will use sample-based techniques to approximate the pushforward and posterior measure in the forward and Bayesian inverse problem, respectively. To this end we require samples $(\tau_1, \theta_1), ..., (\tau_{N_{\text{smp}}}, \theta_{N_{\text{smp}}}) \sim \mu$. We assume that sampling from μ' is accessible and inexpensive. However, for any sample $\tau_n \sim \mu'$ we also need to sample $\theta_n \sim \mu_0(\cdot|\tau_n)$. This requires the assembly of the (dense) covariance matrix $\mathcal{C}(\tau_n)$, and its leading N_{sto} eigenpairs. We abbreviate this process by the function $\text{eigs}(\mathcal{C}(\tau_n), N_{\text{sto}})$ which returns $\Psi := (\lambda_i^{1/2}(\tau_n)\psi_i(\tau_n))_{i=1}^{N_{\text{sto}}}$. We give the full procedure in Algorithm 1.

The cost for the assembly of the covariance matrix is of order $O(N^2; N \to \infty)$. We assume that the cost of a single function call $\text{eigs}(\cdot, N_{\text{sto}})$ is of order $O(N^2 \cdot N_{\text{sto}}; N \to \infty)$. This corresponds to an *Implictly Restarted Lanczos Method*, where $p = O(N_{\text{sto}})$. See [7, 24] for details. Thus, the total computational cost of Algorithm 1 is $O(N_{\text{smp}} \cdot (N^2 \cdot (N_{\text{sto}} + 1)); N \to \infty)$. The largest contribution to the computational cost is the repeated computation of the leading eigenpairs of $\mathcal{C}(\tau_n)$. We can avoid this cost in certain special cases. See e.g. the work by Dunlop et al. [18] where a family of covariance operators with identical eigenfunctions is considered. The associated eigenvalues can be computed analytically. However, in this paper we focus on parameterised covariance operators where the full eigenproblems have to be solved numerically for all parameter values.

Algorithm 1: Sampling from parameterised Gaussian measure μ

for $n \in \{1, ..., N_{\text{smp}}\}$ **do**
  Sample $\tau_n \sim \mu'$
  $\Psi_n \leftarrow \text{eigs}(\mathcal{C}(\tau_n), N_{\text{sto}})$
  Sample $\xi \sim \text{N}(0, \text{Id}_{N_{\text{sto}}})$
  $\theta_n \leftarrow m_0(\tau_n) + \Psi_n \xi$
end

3.3. Monte Carlo for forward uncertainty propagation

First, we return to the forward uncertainty propagation problem (see §2.3). A straightforward way to approximate the pushforward measure $\mathcal{Q}^{\#}\mu''$ is a Monte Carlo method. In particular, we draw $N_{\text{smp}} \in \mathbb{N}$ independent samples from the parameterised Gaussian measure μ,

$$(\tau_1, \theta_1), \ldots, (\tau_{N_{\text{smp}}}, \theta_{N_{\text{smp}}}) \sim \mu.$$

Then, we evaluate the quantity of interest $\mathcal{Q}(\theta_1), ..., \mathcal{Q}(\theta_{N_{\text{smp}}})$. Finally, we construct the discrete measure

$$\widehat{\mathcal{Q}^{\#}\mu''} = \frac{1}{N_{\text{smp}}} \sum_{n=1}^{N_{\text{smp}}} \delta_{\mathcal{Q}(\theta_n)}.$$

The Glivenko-Cantelli Theorem (see [51, §5]) implies that $\widehat{\mathcal{Q}^{\#}\mu''} \to \mathcal{Q}^{\#}\mu''$ weakly, as $N_{\text{smp}} \to \infty$.

3.4. Markov Chain Monte Carlo for Bayesian inverse problems

Since sampling independently from the posterior measure is in general not possible we use Markov Chain Monte Carlo (MCMC). Meaning that we construct an ergodic Markov chain which has the posterior measure as a stationary measure. The samples

$$(\tau_1, \theta_1), ..., (\tau_{N_{\text{smp}}}, \theta_{N_{\text{smp}}}) \sim \mu^y$$

can then be used to approximate integrals with respect to the posterior measure.

In our setting the prior measure is a parameterised Gaussian measure on a (discretised) function space. Following [18] we suggest to use a Metropolis-within-Gibbs sampler. This allows us to sample τ and θ in an alternating way using two different proposal kernels, one for θ and one for τ.

In most cases, the hyperparameter space R will be low dimensional. Hence, various Metropolis-Hastings proposals can be used to efficiently propose samples of τ. We denote the conditional density of this Metropolis-Hastings proposal by $q_R : R \times R \to \mathbb{R}$. On the other hand, θ is a X-valued random variable where X is high-dimensional or possibly infinite-dimensional. Therefore, we suggest a preconditioned Crank-Nicholson (pCN) proposal. Combining the Metropolis-Hastings and the pCN proposal we arrive at the MCMC method in Algorithm 2.

Some comments are in order. In Algorithm 2 we require $\dim X < \infty$. See §3.1 for the discretisation of the space X. In this case, the Gaussian measure $\mathrm{N}(m(\tau_0), \mathcal{C}(\tau_0))$ has a probability density function (see Proposition 2.3), and we use this probability density function to compute the acceptance probability α_R for the Gibbs move of τ. It is however possible to compute α_R without access to the probability density function. In this case one can also define an algorithm for an infinite dimensional space X. We refer to [18] for a rigorous introduction of the infinite-dimensional version.

Algorithm 2: Metropolis-within-Gibbs to sample from the posterior measure of (τ, θ)

Let (τ_0, θ_0) be the initial sample of the Markov chain.
for $n \in \{1, ..., N_{\mathrm{smp}}\}$ **do**
  Sample $\tau^* \sim q_R(\cdot|\tau_{n-1})$
  $\alpha_R(\tau_{n-1}; \tau^*) \leftarrow \min\left\{1, \frac{q_R(\tau_{n-1}|\tau^*)}{q_R(\tau^*|\tau_{n-1})} \frac{\mathrm{n}(\theta_{n-1}; m(\tau^*), \mathcal{C}(\tau^*))}{\mathrm{n}(\theta_{n-1}; m(\tau_{n-1}), \mathcal{C}(\tau_{n-1}))}\right\}$
  Sample $U_R \sim \mathrm{Unif}[0, 1]$
  if $U_R < \alpha_R$ **then**
    $\tau_n \leftarrow \tau^*$
  else
    $\tau_n \leftarrow \tau_{n-1}$
  end
  Sample $\theta^* \sim \mathrm{N}(\sqrt{1-\beta^2}\theta_{n-1}, \beta\mathcal{C}(\tau_n))$
  $\alpha_X(\theta_{n-1}; \theta^*) \leftarrow \min\{1, \exp(-\Phi(\theta^*) + \Phi(\theta_{n-1}))\}$
  Sample $U_X \sim \mathrm{Unif}[0, 1]$
  if $U_X < \alpha_X$ **then**
    $\theta_n \leftarrow \theta^*$
  else
    $\theta_n \leftarrow \theta_{n-1}$
  end
end

It can be proved that Algorithm 2 is valid to sample from μ^y. We summarise this in the following Proposition and refer to §3 of [18] for a similar statement and proof.

Proposition 3.1 *Algorithm 2 defines a Markov kernel* MK *from* $(X \times R, \mathcal{B}X \otimes \mathcal{R})$ *to* $(X \times R, \mathcal{B}X \otimes \mathcal{R})$ *that has* μ^y *as a stationary measure. In particular,*

$$\mu^y \mathrm{MK} = \mu^y :\Leftrightarrow \int \mathrm{MK}(B|\tau,\theta)\mu^y(\mathrm{d}\tau,\mathrm{d}\theta) = \mu^y(B) \ (B \in \mathcal{B}X \otimes \mathcal{R}).$$

4. Reduced basis approach to parameterised eigenproblems

In §3.2 we discuss the sampling from parameterised Gaussian random fields. The largest contribution to the computational cost in this sampling procedure is the repeated computation of eigenpairs of the associated parameterised covariance matrix $\mathcal{C}(\tau_0)$ for multiple parameter values $\tau_0 \in R$. Reduced basis (RB) methods construct a low-dimensional trial space for a family of parameterised eigenproblems. This is the cornerstone in our fast sampling algorithm. To begin we explain the basic idea behind reduced basis (RB) approaches for eigenproblems. There are multiple ways to construct a reduced basis, such as the proper orthogonal decomposition (POD), as well as single- and multi-choice greedy approaches. POD and greedy approaches for parameterised eigenproblems are discussed and compared in [28]. In this paper we focus on the POD.

RB algorithms consist of two parts, an offline or preprocessing phase and an online phase. The construction of the reduced basis is the offline phase. In the online phase the reduced basis is used to solve a large number of low-dimensional eigenproblems. Finally, in §4.3, we discuss the offline-online decomposition for Matérn-type covariance operators. To be able to implement the RB approach efficiently we approximate the full covariance operators by linearly separable operators. We explain how this can be done, and finally analyse the proposed class of approximate operators.

4.1. Basic idea

Let $\mathcal{C} : R \to \mathrm{CO}(X)$ be a measurable map, where $(X, \langle\cdot,\cdot\rangle_X) := (\mathbb{R}^N, \langle\cdot,\cdot\rangle_M)$ is a finite-dimensional space (see §3.1). In Algorithm 1 we need to solve the generalised eigenproblem associated with $\mathcal{C}(\tau_0)$ for multiple parameter values $\tau_0 \in R$. That is, we want to find $(\lambda_i(\tau_0), \psi_i(\tau_0))_{i=1}^{N_{\mathrm{sto}}} \in (\mathbb{R} \times X)^{N_{\mathrm{sto}}}$, such that

$$\mathcal{C}(\tau_0)\psi_i(\tau_0) = \lambda_i(\tau_0) M \psi_i(\tau_0). \tag{4.1}$$

X is in general high-dimensional, which results in a large computational cost for solving the eigenproblems. However, it is often not necessary to consider the full space X. If we assume that the eigenpairs corresponding to different parameter values are closely related, then the space

$$\mathrm{span}\{\psi_i(\tau_0) : i = 1, ..., N_{\mathrm{sto}}, \tau_0 \in R\} \subseteq X$$

can be approximated by a low dimensional subspace X_{RB}, where $N_{\mathrm{RB}} := \dim X_{\mathrm{RB}} \ll N$. We point out that the truncated KL expansion requires N_{sto} eigenpairs by assumption. However, the reduced operators are $N_{\mathrm{RB}} \times N_{\mathrm{RB}}$ matrices with N_{RB} eigenpairs. Hence, $N_{\mathrm{RB}} \geq N_{\mathrm{sto}}$ is required.

Now, let $W \in X_{\mathrm{RB}}^{N_{\mathrm{RB}}}$ be an orthonormal basis of X_{RB} with respect to the inner product $\langle \cdot, \cdot \rangle_{X_{\mathrm{RB}}} := \langle \cdot, \cdot \rangle_X := \langle \cdot, \cdot \rangle_M$. W is called *reduced basis* and X_{RB} is called *reduced space*. We can represent any function $\psi \in X_{\mathrm{RB}}$ by a coefficient vector $w \in \mathbb{R}^{N_{\mathrm{RB}}}$, such that $\psi = Ww$. The *reduced eigenproblem* is obtained by a Galerkin projection of the full eigenproblem in (4.1), and is again a generalised eigenproblem. The task is to find $(\lambda_i^{\mathrm{RB}}(\tau_0), w_i^{\mathrm{RB}}(\tau_0))_{i=1}^{N_{\mathrm{sto}}} \in (\mathbb{R} \times \mathbb{R}^{N_{\mathrm{RB}}})^{N_{\mathrm{sto}}}$, such that

$$\mathcal{C}^{\mathrm{RB}}(\tau_0) w_i(\tau_0) = \lambda_i^{\mathrm{RB}}(\tau_0) M^{\mathrm{RB}} w_i(\tau_0), \quad i = 1, ..., N_{\mathrm{sto}}. \tag{4.2}$$

In (4.2) we have the *reduced operator* $\mathcal{C}^{\mathrm{RB}}(\tau_0) := W^* \mathcal{C}(\tau_0) W$, and the *reduced Gramian matrix* $M^{\mathrm{RB}} := W^* M W$, that are both $N_{\mathrm{RB}} \times N_{\mathrm{RB}}$ matrices. The eigenvectors in X_{RB} can then be recovered by $\psi_i^{\mathrm{RB}}(\tau_0) = W w_i(\tau_0)$, $i = 1, ..., N_{\mathrm{sto}}$.

4.2. Offline-online decomposition

A reduced basis method typically has two phases. In the offline phase the reduced basis W is constructed. In the online phase the reduced operator $\mathcal{C}^{\mathrm{RB}}(\tau_0)$ is assembled, and the reduced eigenproblem (4.2) is solved for selected parameter values $\tau_0 \in R$. To be able to shift a large part of the computational cost from the online to the offline phase we assume that the following offline-online decomposition is available for the family of parameterised covariance operators.

Assumptions 4.1 *Let $N_{\mathrm{lin}} \in \mathbb{N}$. We assume that there are functions $F_k : R \to \mathbb{R}$ and linear operators $\mathcal{C}_k$, $k = 1, ..., N_{\mathrm{lin}}$, such that*

$$\mathcal{C}(\tau_0) = \sum_{k=1}^{N_{\mathrm{lin}}} F_k(\tau_0) \mathcal{C}_k, \quad \tau_0 \in R.$$

In this case, $\mathcal{C}(\tau_0)$ is called a linearly separable operator.

4.2.1. Offline phase

We use snapshots of the full eigenvectors to construct the reduced basis. Meaning that we choose a vector $\tau^{\mathrm{snap}} \in R^{N_{\mathrm{snap}}}$ and solve the full eigenproblem (4.1) for all elements of τ^{snap}. We then have

$$W_{\mathrm{snap}} = (\psi_i(\tau_s^{\mathrm{snap}}) : s = 1, ..., N_{\mathrm{snap}}, i = 1, ..., N_{\mathrm{sto}}),$$

and define the reduced space $X_{\mathrm{RB}} := \mathrm{span}(W_{\mathrm{snap}})$. Next, we construct an orthonormal basis for this vector space. One option to do this is the *proper orthogonal decomposition*. Note that the POD is in fact equivalent to a principal component analysis of the 'covariance operator' $W_{\mathrm{snap}}^{(2)} := W_{\mathrm{snap}} W_{\mathrm{snap}}^*$. As result of the POD we obtain a spectral decomposition of $W_{\mathrm{snap}}^{(2)}$ of the form

$$W_{\mathrm{snap}}^{(2)} = Q \Lambda Q^*,$$

where $\Lambda := \mathrm{diag}(\lambda_1^{\mathrm{snap}}, ..., \lambda_N^{\mathrm{snap}})$ is a diagonal matrix containing the eigenvalues of $W_{\mathrm{snap}}^{(2)}$ and each column of Q contains the associated orthonormal eigenvectors. We use the eigenvectors with non-zero eigenvalues as basis vectors of X_{RB}, that is,

$$W := (Q_{\cdot,i} : \lambda_i^{\mathrm{snap}} > 0, i = 1, ..., N).$$

The magnitude of the eigenvalues of $W_{\text{snap}}^{(2)}$ is an indicator for the error when the corresponding eigenvectors are not included in W. See the discussion in §3.1. Neglecting eigenvectors however is beneficial due to the smaller dimension of the reduced basis. Depending on the pay-off of the dimension reduction compared to the approximation accuracy of the reducd basis one can choose a threshold $\underline{\lambda} > 0$ and work with the basis

$$W := (Q_{\cdot,i} : \lambda_i^{\text{snap}} > \underline{\lambda}, i = 1, ..., N).$$

Remark 4.2 *In this paper we compute the singular value decomposition (SVD) of W_{snap} instead of the spectral decomposition of $W_{\text{snap}}^{(2)}$. This is possible since the squares of the singular values of W_{snap} are identical to the eigenvalues of $W_{\text{snap}}^{(2)}$.*

There are many ways to choose the snapshot parameter values τ^{snap}. In our applications τ is a random variable with probability measure μ'. Hence, a straightforward method is to sample independently $\tau_s^{\text{snap}} \sim \mu'$ $(s = 1, ..., N_{\text{snap}})$. Alternatively, one can select deterministic points in R, e.g. quadrature nodes. We will come back to this question in §6 where we discuss the numerical experiments.

4.2.2. Online phase

In the online phase we iterate over various hyperparameter values $\tau_0 \in R$. In every step, we assemble the operator $\mathcal{C}^{\text{RB}}(\tau_0)$, and then we solve the associated eigenproblem (4.2). By Assumptions 4.1, it holds

$$\mathcal{C}(\tau_0) = \sum_{n=1}^{N_{\text{lin}}} F_k(\tau_0)\mathcal{C}_k.$$

Hence, the reduced operator can be assembled efficiently as follows,

$$\mathcal{C}^{\text{RB}}(\tau) = W^* \sum_{k=1}^{N_{\text{lin}}} F_m(\tau)\mathcal{C}_k W = \sum_{k=1}^{N_{\text{lin}}} F_k(\tau) W^* \mathcal{C}_k W = \sum_{k=1}^{N_{\text{lin}}} F_k(\tau)\mathcal{C}_k^{\text{RB}}.$$

The reduced operators $\mathcal{C}_k^{\text{RB}}$ $(k = 1, ..., N_{\text{lin}})$ can be computed in the offline phase and stored in the memory. In the online phase, we then only need to compute a certain linear combination of $(\mathcal{C}_k^{\text{RB}})_{k=1}^{N_{\text{lin}}}$. This reduces the computational cost of the assembly of the reduced operator significantly. After the assembly step we solve the reduced eigenproblem (4.2) and obtain the eigenvectors $\psi_i^{\text{RB}}(\tau_0) := W w_i(\tau_0) \in X$ and eigenvalues $\lambda_i^{\text{RB}}(\tau), i = 1, ..., N_{\text{sto}}$.

4.3. Matérn-type covariance operators

Matérn-type covariance operators are widely used in spatial statistics and uncertainty quantification. They are particularly popular for modelling spatially variable uncertainties in porous media. We are interested in solving the KL eigenproblem with Matérn covariance operators with hyperparameters e.g. the correlation length. Unfortunately, the Matérn-type covariance operators are not linearly separable with respect to the hyperparameters of interest (see Assumptions 4.1). For this reason we introduce and analyse a class of linearly separable covariance operators which can approximate Matérn-type covariance operators with arbitrary accuracy.

Definition 4.3 *Let $D \subseteq \mathbb{R}^d, d = 1, 2, 3$ be an open, bounded and connected domain, and let $X := \mathcal{L}^2(D; \mathbb{R})$. Furthermore, let $\ell \in (0, \mathrm{diam}(D)), \nu \in (0, \infty], \sigma \in (0, \infty)$. Define the covariance kernel $c(\nu, \ell, \sigma) : [0, \infty) \to [0, \infty)$ as*

$$z \mapsto c(\nu, \ell, \sigma)(z) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{z}{\ell}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\frac{z}{\ell}\right),$$

where K_ν is the modified Bessel function of the second kind. Then, the Matérn-type covariance operator with smoothness ν, standard deviation σ and correlation length ℓ is given by

$$\mathcal{C}(\nu, \ell, \sigma) : X \to X, \varphi \mapsto \int_D \varphi(x) c(\nu, \ell, \sigma)(\mathrm{dist}(x, \cdot))\mathrm{d}x,$$

where $\mathrm{dist} : D \times D \to [0, \infty)$ *is the Euclidean distance in D.*

Remark 4.4 *The exponential covariance operator which we discuss in Examples 2.6 and 2.9 is a Matérn-type covariance operator. Indeed, $\mathcal{C}(1/2, \ell, \sigma) \equiv \mathcal{C}_{\exp}^{(\ell,\sigma)}$.*

In Example 2.9 we discussed the possibility of using the standard deviation σ and the correlation length ℓ as hyperparameters in a Matérn-type covariance operator. What are the computational implications for the KL expansion? Changing σ only rescales the KL eigenvalues, and does not require a re-computation of the KL. However, changing the correlation length clearly changes the KL eigenfunctions. We can see this in Figure 4.1. However, the good news is that the KL eigenfunctions for different correlation lengths are very similar, for example, the number and location of extrema is preserved. This suggests that we might be able to construct a useful reduced basis from selected snapshots of KL eigenpairs corresponding to different correlation lengths. To be able to construct and use the reduced basis efficiently requires the linear separability of the covariance operator, see Assumptions 4.1. The Matérn operator is linearly separable with respect to σ. On the other hand it is not linearly separable with respect to ℓ, since the covariance function $c(\nu, \ell, \sigma)$ is not linearly separable. However, it is possible to approximate $\mathcal{C}(\nu, \ell, \sigma)$ with any precision by a linearly separable operator. We show this in the remainder of this section.

Assumptions 4.5 *The correlation length ℓ satisfies $0 < \underline{\ell} \leq \ell$ with fixed $\underline{\ell}$.*

Lemma 4.6 *Let $\nu \in (0, \infty]$, and let Assumption 4.5 hold. Then, there is a linearly separable operator*

$$\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}}) : X \to X$$

consisting of $N_{\mathrm{lin}} \in 2\mathbb{N}$ terms, such that asymptotically

$$\|\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}}) - \mathcal{C}(\nu, \ell, \sigma)\|_X \leq O(1/(N_{\mathrm{lin}}!); N_{\mathrm{lin}} \to \infty). \tag{4.3}$$

Proof. See Appendix B. □

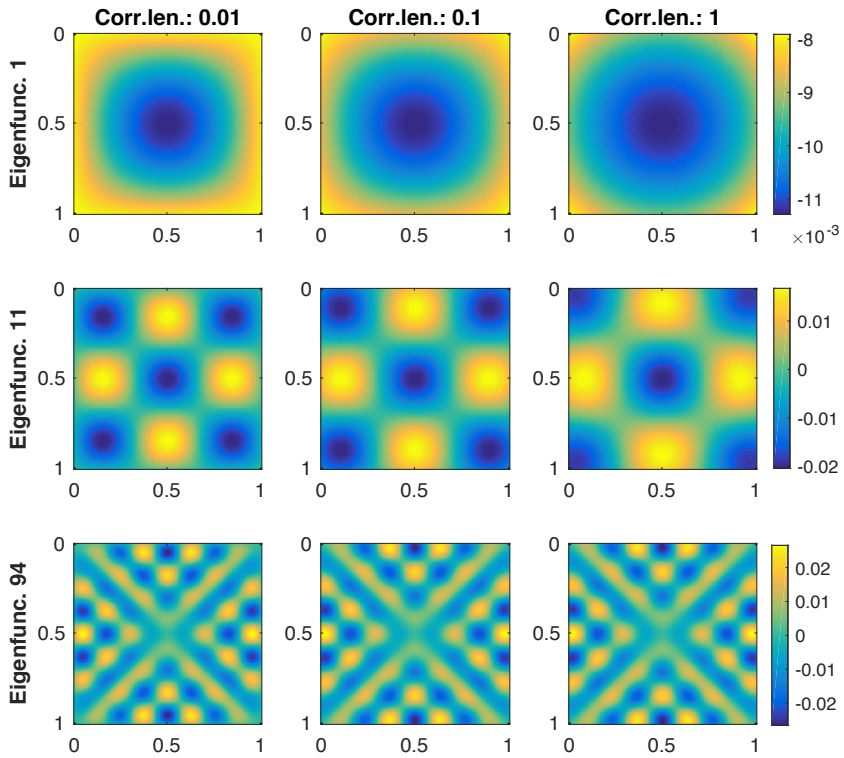

Figure 4.1: Eigenfunctions 1, 11 and 94 of the Matérn-type covariance operator with correlation lengths $\ell = 0.01, 0.1, 1$ and $\nu = 1/2$.

The covariance operator approximation brings new issues. The Matérn-type covariance operators $\mathcal{C}(\nu, \ell, \sigma)$ are valid covariance operators in $\mathrm{CO}(X)$. However, this is not necessarily the case for $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$. One can easily verify the following.

Lemma 4.7 *The operator $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$ is self-adjoint, trace-class and continuous.*

Proof. See Appendix B. □

However, $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$ is not necessarily positive definite. Under weak assumptions we can cure this by replacing $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$ by an operator $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\mathrm{lin}})$ which has the exact same eigenfunctions and positive eigenvalues as $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$, however, all negative eigenvalues are set to zero. Formally, we define $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\mathrm{lin}})$ by

$$\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\mathrm{lin}}) = \sum_{i=1; \widetilde{\lambda}_i > 0}^{\infty} \widetilde{\lambda}_i (\widetilde{\psi}_i \otimes \widetilde{\psi}_i), \tag{4.4}$$

where $(\widetilde{\lambda}_i, \widetilde{\psi}_i)_{i=1}^{\infty}$ are eigenpairs of $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$ and the eigenfunctions are orthonormal. Note that the same technique has been applied in [11] to remove the degeneracy of multilevel sample covariance estimators. Fortunately, we can show that the approximation error of $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\mathrm{lin}})$ is of the same order as the error of $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$.

Lemma 4.8 *The Matérn-type covariance operator $\mathcal{C}(\nu, \ell, \sigma)$ and the approximate operator $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}})$ in (4.4) satisfy*

$$\|\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}) - \mathcal{C}(\nu, \ell, \sigma)\|_X \leq 2\|\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\text{lin}}) - \mathcal{C}(\nu, \ell, \sigma)\|_X.$$

Proof. See Appendix B. □

We summarise the results in Lemma 4.6–4.8 as follows.

Proposition 4.9 *Let $\nu \in (0, \infty)$. Under Assumption 4.5 there is a linearly separable, valid covariance operator $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}) \in \text{CO}(X)$ consisting of $N_{\text{lin}} \in 2\mathbb{N}$ terms, such that asymptotically*

$$\|\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}) - \mathcal{C}(\nu, \ell, \sigma)\|_X \leq O(1/(N_{\text{lin}}!); N_{\text{lin}} \to \infty).$$

The expansion in (4.4) has infinitely many terms. We truncate this expansion and retain only the leading N_{sto} terms, denoting the resulting covariance operator by $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}, N_{\text{sto}})$.

Finally, we discuss the sample path continuity. Samples of Gaussian random fields with Matérn-type covariance operators are almost surely continuous functions. In the following proposition we show that this also holds for the realisations of $\text{N}(0, \widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}, N_{\text{sto}}))$.

Proposition 4.10 *Let $\theta \sim \text{N}\left(0, \widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}, N_{\text{sto}})\right)$, where $\nu \in (0, \infty)$. Then it holds $\theta \in C^0(D)$.*

Proof. We consider the random field in terms of its (finite) KL expansion,

$$\theta := \theta_{\text{KL}}^{N_{\text{sto}}} := \sum_{i=1}^{N_{\text{sto}}} \sqrt{\widetilde{\lambda}_i}\widetilde{\psi}_i\xi_i,$$

where $(\widetilde{\lambda}_i, \widetilde{\psi}_i)_{i=1}^{N_{\text{sto}}}$ are the eigenpairs of $\widetilde{\mathcal{C}}_0(\nu, \ell, \sigma, N_{\text{lin}}, N_{\text{sto}})$ with positive eigenvalues. Then, θ is a continuous function, if $(\widetilde{\psi}_i)_{i=1}^{N_{\text{sto}}}$ is a family of continuous functions. This however is clear by the definition of $(\psi_i)_{i=1}^{N_{\text{sto}}}$. Indeed, let $i = 1, ..., N_{\text{sto}}$. Then it holds

$$\widetilde{\psi}_i(y) = \frac{1}{\lambda_i} \int_D \widetilde{\psi}_i(x)\widetilde{c}^{(\nu,\ell,\sigma,N_{\text{lin}})}(\text{dist}(x, y))\text{d}x, \quad y \in D.$$

By definition, $\widetilde{c}^{(\nu,\ell,\sigma,N_{\text{lin}})}(\text{dist}(\cdot, \cdot))$ is a continuous function. Hence, one can easily verify that the right-hand side is continuous. □

5. Reduced basis sampling

In §3.2 we discuss sampling from a Markov kernel, or, equivalently, a parameterised Gaussian measure

$$K(\cdot|\tau) := \text{N}(m(\tau), \mathcal{C}(\tau)),$$

where $\tau \sim \mu'$. Now, to reduce the computational cost, we combine Algorithm 1 and reduced bases in §5.1. We discuss the computational cost of the offline and the online phase of the suggested reduced basis sampling in §5.2. Finally, in §5.3 we explain how the reduced basis induces an alternative expansion for the parameterised Gaussian random field.

5.1. Algorithm

First we describe the offline phase of the reduced basis sampling. We use a POD approach to compute a reduced basis W for $\mathcal{C}(\cdot)$. Here, it is important that the dimension of $X_{\mathrm{RB}} = \mathrm{span}(W)$ is larger than N_{sto}. Furthermore, we assume that $\mathcal{C}(\cdot)$ fulfils Assumptions 4.1, i.e., it has the linearly separable form

$$\mathcal{C}(\cdot) := \sum_{k=1}^{N_{\mathrm{lin}}} F_k(\cdot)\mathcal{C}_k.$$

Having constructed the reduced basis W, we compute $\mathcal{C}_k^{\mathrm{RB}} = W^*\mathcal{C}_k W, k = 1, ..., N_{\mathrm{lin}}$. Then, we proceed with the online phase (see Algorithm 3). We iterate over $n =$

Algorithm 3: Reduced basis sampling from the parameterised Gaussian measure

for $n \in \{1, ..., N_{\mathrm{smp}}\}$ **do**
- Sample $\tau_n \sim \mu'$
- $\mathcal{C}^{\mathrm{RB}}(\tau_n) \leftarrow \sum_{k=1}^{N_{\mathrm{lin}}} F_k(\tau_n)\mathcal{C}_k^{\mathrm{RB}}$
- $\Psi_n^{\mathrm{RB}}(\tau_n) \leftarrow \mathrm{eigs}(\mathcal{C}^{\mathrm{RB}}(\tau_n), N_{\mathrm{sto}})$
- $\Psi_n(\tau_n) \leftarrow W\Psi_n^{\mathrm{RB}}(\tau_n)$
- Sample $\xi \sim \mathrm{N}(0, \mathrm{Id}_{N_{\mathrm{sto}}})$
- $\theta_n \leftarrow m(\tau_n) + \Psi_n(\tau_n)\xi$

end

$1, ..., N_{\mathrm{smp}}$. In each step we first sample $\tau_n \sim \mu'$. Then, we evaluate the reduced covariance operator $\mathcal{C}^{\mathrm{RB}}(\tau_n)$, compute the eigenpairs $(\lambda_i^{\mathrm{RB}}(\tau_n), w_i^{\mathrm{RB}}(\tau_n))_{i=1}^{N_{\mathrm{sto}}}$ of $\mathcal{C}^{\mathrm{RB}}(\tau_n)$, and return $\Psi_n^{\mathrm{RB}}(\tau_n) := \left(\sqrt{\lambda_i^{\mathrm{RB}}(\tau_n)}w_i^{\mathrm{RB}}(\tau_n) : i = 1, ..., N_{\mathrm{sto}}\right)$. Next, we compute the representation of $\Psi_n^{\mathrm{RB}}(\tau_n)$ on the full space X, that is, $\Psi_n(\tau_n) := W\Psi_n^{\mathrm{RB}}(\tau_n)$. Finally, we proceed as in Algorithm 1: We sample a multivariate standard Gaussian random variable with N_{sto} components and return $m(\tau_n) + \Psi_n(\tau_n)\xi$.

5.2. Computational cost

We assume again that $X = \mathbb{R}^N$. The number of solves of the full eigenproblem in the offline phase is N_{snap}. We assume that N_{snap} is much smaller than the number of samples N_{smp} we want to generate. Furthermore, we assume that the dimension of the reduced basis $\dim(X_{\mathrm{RB}}) = N_{\mathrm{RB}}$ is smaller than the dimension of the full problem N, and that $\mathcal{C}(\tau)$ and $\mathcal{C}^{\mathrm{RB}}(\tau)$ are both dense matrices.

The computational cost of the tasks in the offline phase is given in Table 1. The total offline cost is $O(N_{\mathrm{snap}}N^2 + N_{\mathrm{snap}}N^2N_{\mathrm{sto}} + N_{\mathrm{snap}}N^2N_{\mathrm{sto}}; N \to \infty)$. Since

Task	Computational Cost
Construct the full operator	$O(N_{\text{snap}} N^2)$
Solve the full eigenproblem	$O(N_{\text{snap}} N^2 N_{\text{sto}})$
POD	$O(N_{\text{snap}} N^2 N_{\text{sto}})$

Table 1: Computational cost of the offline phase.

$N_{\text{snap}} \ll N_{\text{smp}}$ the offline cost is asymptotically much cheaper than the cost of Algorithm 1 where we solve the full eigenproblem for each sample.

The computational cost of the tasks in the online phase is given in Table 2. The total online cost is $O(N_{\text{smp}} N_{\text{RB}}^2 N_{\text{lin}} + N_{\text{smp}} N_{\text{RB}}^2 N_{\text{sto}} + N_{\text{smp}} N_{\text{RB}} N; N \to \infty)$. In

Task	Computational Cost
Construct the reduced operator	$O(N_{\text{smp}} N_{\text{RB}}^2 N_{\text{lin}})$
Solve the reduced eigenproblem	$O(N_{\text{smp}} N_{\text{RB}}^2 N_{\text{sto}})$
Map the reduced solution to the full space	$O(N_{\text{smp}} N_{\text{RB}} N)$

Table 2: Computational cost of the online phase.

the online phase we solve the covariance eigenproblems in the reduced space. The high-dimensional full space X is only required when we map the reduced sample to X. The cost of these steps is linear in the dimension N of X, and quadratic in the dimension of the reduced basis N_{RB} for every sample. In constrast, the cost of Algorithm 1 is at least quadratic in N, for every sample. Hence, for every sample, we need to solve an $O(N; N \to \infty)$ problem using RB, but an $O(N^2; N \to \infty)$ problem in the full space. This clearly demonstrates the advantages of RB sampling.

5.3. Reduced basis random field expansion

In Algorithm 3 we describe sampling from the full random field θ. However, this can be inefficient for two reasons:

1. In some applications, e.g. the Bayesian inverse problem, the random field samples need to be stored and kept in memory. This is often impossible due to memory limitations.
2. In some methods, e.g. the MCMC method in Algorithm 2, we need to assess random fields with respect to some covariance operators. Hence, not only the random fields have to be kept in memory but also at least one covariance operator. This requires one order of magnitude more memory than a single random field, that is, $O(N^2; N \to \infty)$ compared to $O(N; N \to \infty)$.

The reduced basis enables a natural compression of the full random field θ. Let $\theta \sim K(\cdot|\tau_0)$ for some $\tau_0 \in R$. The reduced basis implies the representation

$$\theta = m(\tau_0) + \Psi_0(\tau_0)\xi = m(\tau_0) + W\Psi_0^{\text{RB}}(\tau_0)\xi$$

for some $\xi \sim \mathrm{N}(0, \mathrm{Id}_{N_{\mathrm{sto}}})$. We can represent θ in terms of $\theta_{\mathrm{RB}} := (\Psi_0^{\mathrm{RB}}(\tau_0)\xi) \in \mathbb{R}^{N_{\mathrm{RB}}}$ which gives

$$\theta = m(\tau_0) + W\theta_{\mathrm{RB}}, \tag{5.1}$$

the *reduced basis expansion* of the random field θ. Note that $\theta_{\mathrm{RB}} \sim \mathrm{N}(0, \mathcal{C}^{\mathrm{RB}}(\tau_0))$ is a Gaussian random field on the reduced space. In contrast to a KL expansion the random field expansion in (5.1) is not necessarily optimal when we compare the number of terms N_{RB} required to achieve a certain approximation accuracy in the mean-square error. We comment on the KL expansion of the non-Gaussian random field $\theta \sim \mu''$ in the following.

Let

$$\overline{\mathcal{C}} := \int_R \mathcal{C}(\tau)\mu'(\mathrm{d}\tau)$$

denote the covariance operator of $\theta \sim \mu''$ and $\overline{m} := \int m(\tau)\mu'(\mathrm{d}\tau)$ denote its mean. We assume that $\overline{m}$ and $\overline{\mathcal{C}}$ are well-defined. Based on $\overline{m}$ and $\overline{\mathcal{C}}$ we obtain the KL expansion

$$\theta_{\mathrm{KL}} := \overline{m} + \sum_{i=1}^{\infty} \sqrt{\overline{\lambda}_i}\overline{\psi}_i\overline{\xi}_i, \tag{5.2}$$

where $(\overline{\lambda}_i, \overline{\psi}_i)_{i=1}^{\infty}$ are eigenpairs of $\overline{\mathcal{C}}$, and $(\overline{\xi}_i)_{i=1}^{\infty}$ is a family of countably many uncorrelated scalar random variables. These random variables are typically neither standard Gaussian nor stochastically independent. One can show that $\theta_{\mathrm{KL}} \sim \mu''$. The truncated KL expansion associated with (5.2) satisfies a best approximation property. Let

$$\theta_{\mathrm{KL}}^{N_{\mathrm{sto}}} := \overline{m} + \sum_{i=1}^{N_{\mathrm{sto}}} \sqrt{\overline{\lambda}_i}\overline{\psi}_i\overline{\xi}_i.$$

Then, $\theta_{\mathrm{KL}}^{N_{\mathrm{sto}}}$ solves the minimisation problem

$$\min_{\widehat{\theta} \in \mathcal{L}^2(\mathbb{R}^{N_{\mathrm{sto}}}; X)} \mathbb{E}\left[\|\theta_{\mathrm{KL}} - \widehat{\theta}\|_X^2\right].$$

There are many alternative random field expansions depending on the choice of the basis. One could for instance use eigenpairs $(\lambda_i(\tau_0), \psi_i(\tau_0))_{i=1}^{\infty}$ for a fixed $\tau_0 \in R$. Various approaches for random field expansions have been proposed and compared in [47]. We advocate the reduced basis random field expansion for two reasons:

1. The KL expansion associated with μ'' requires the construction of $\overline{\mathcal{C}}$ and the computation of its eigenpairs.
2. The offline-online decomposition has been computed for the basis W. A basis change requires the computation of new reduced operators $\mathcal{C}^{\mathrm{RB}}(\tau_0)$ and $(\mathcal{C}_k^{\mathrm{RB}})_{k=1}^{N_{\mathrm{lin}}}$, resp.

Observe that the covariance of θ for a fixed τ_0 can be fully described by $\mathcal{C}^{\mathrm{RB}}(\tau_0) \in \mathbb{R}^{N_{\mathrm{RB}} \times N_{\mathrm{RB}}}$. Hence, we can sample θ cheaply, and we can also represent the covariance of θ by a low dimensional matrix. This is particularly useful in Bayesian inverse problems as we explain next.

Consider a Bayesian inverse problem with parameterised prior random field and assume that $m(\tau_0) \in X_{\mathrm{RB}}$, $\tau_0 \in R$. If this is not the case, one can either add $m(\tau_0)$ for the required $\tau_0 \in R$ to X_{RB} or project $m(\tau_0)$ orthogonally onto X_{RB}. In Algorithm 2 we need the full random field θ as input for the potential Φ to compute the acceptance probability of the pCN-MCMC proposal. Moreover, to compute the acceptance probability of the τ-proposal, again we need to construct the full precision operator of the proposed $\tau^* \in R$. The computational cost of this construction is of order $O(N; N \to \infty)$. However, since θ_{RB} already contains the full covariance information, we can replace $\mathcal{C}(\tau^*)^{-1}$ by $\mathcal{C}^{\mathrm{RB}}(\tau^*)^{-1}$. Then, the computational cost of the Gibbs step in τ-direction is independent of N. We summarise the associated Reduced Basis MCMC method in Algorithm 4.

Algorithm 4: Reduced Basis Markov Chain Monte Carlo

Let $(\tau_0, \theta_{\mathrm{RB},0}) \in \mathbb{R}^{N_{\mathrm{RB}}+1}$ be the initial state of the Markov chain.
for $n \in \{1, ..., N_{\mathrm{smp}}\}$ **do**
  Sample $\tau^* \sim q_R(\cdot|\tau_{n-1})$
  $\alpha_R(\tau_{n-1}; \tau^*) \leftarrow \min\left\{1, \frac{q_R(\tau_{n-1}|\tau^*)}{q_R(\tau^*|\tau_{n-1})} \frac{\mathrm{n}(\theta_{\mathrm{RB},n-1}; W^* m(\tau^*), \mathcal{C}^{\mathrm{RB}}(\tau^*))}{\mathrm{n}(\theta_{\mathrm{RB},n-1}; W^* m(\tau_{n-1}), \mathcal{C}^{\mathrm{RB}}(\tau_{n-1}))}\right\}$
  Sample $U_R \sim \mathrm{Unif}[0,1]$
  if $U_R < \alpha_R$ **then**
    $\tau_n \leftarrow \tau^*$
  else
    $\tau_n \leftarrow \tau_{n-1}$
  end
  Sample $\theta^*_{\mathrm{RB}} \sim \mathrm{N}(\sqrt{1-\beta^2}\theta_{\mathrm{RB},n-1}, \beta\mathcal{C}^{\mathrm{RB}}(\tau_n))$
  $\alpha_X(\theta_{\mathrm{RB},n-1}; \theta^*_{\mathrm{RB}}) \leftarrow \min\{1, \exp(-\Phi(W\theta^*_{\mathrm{RB}}) + \Phi(W\theta_{\mathrm{RB},n-1}))\}$
  Sample $U_X \sim \mathrm{Unif}[0,1]$
  if $U_X < \alpha_X$ **then**
    $\theta_{\mathrm{RB},n} \leftarrow \theta^*_{\mathrm{RB}}$
  else
    $\theta_{\mathrm{RB},n} \leftarrow \theta_{\mathrm{RB},n-1}$
  end
end

6. Numerical experiments

In this section we illustrate and verify the reduced basis sampling for use with forward and Bayesian inverse problems. We start by measuring runtime and accuracy of the reduced basis approximation to the parametric KL eigenproblems. In Example 6.1, we consider a forward and a Bayesian inverse problem in a low-dimensional test setting. This allows us to compare the reduced basis sampling with the samples obtained by using the full, unreduced KL eigenproblems. We then move on to high-dimensional estimation problems in Examples 6.2–6.4. In Examples 6.2–6.3 we consider the elliptic PDE

$$-\nabla \cdot (\exp(\theta(x))\nabla p(x)) = f(x) \qquad (x \in D) \tag{6.1}$$

on the unit square domain $D = (0,1)^2$ together with suitable boundary conditions. The PDE (6.1) is discretised with linear, continuous finite elements on a uniform, triangular mesh. The coefficient function θ is a parameterised Gaussian random field with exponential covariance operator, and random correlation length and standard deviation, respectively (see Example 2.9). The spatial discretisation of θ is done with piecewise constant finite elements on a uniform, rectangular mesh. The evaluation of the covariance operator on this finite element space requires to evaluate an integral. We approximate this integral using a composite midpoint rule, with one quadrature node in each finite element.

We further discretise θ by a truncated KL expansion where we retain the leading N_{sto} terms. The parameter N_{sto} is selected such that the truncated KL captures at least 90% of the total variance. We list the random field parameters for Examples 6.1–6.4 in Table 3. We introduce the estimation problems in more detail in the following subsections. Note that we solve the test problems in Examples 6.1–6.4 using the reduced basis samplers presented in §5.

	Example 6.1	Example 6.2	Example 6.3	Example 6.4
$\underline{\sigma}$	1	0.1	0.5	0.1
$\overline{\sigma}$	1	1	0.5	1
m_σ	1	0.5	0.5	0.5
σ_σ^2	0	0.1	0	0.1
$\underline{\ell}$	0.3	0.3	0.3	0.1
N_{sto}	200	100	100	800

Table 3: Random field parameters in Examples 6.1–6.4.

6.1. Accuracy and speed up

First we assess the accuracy and time consumption of the reduced basis approximation. We measure the quality of the reduced basis surrogate by comparing reduced basis eigenvalues with full eigenvalues in a simplified setting. The *full matrix* is the finite element approximation of $\mathcal{C}_{\text{exp}}^{(\ell,1)}$ with 100×100 piecewise constant finite elements. The goal is to compute the leading 100 eigenpairs of $\mathcal{C}_{\text{exp}}^{(\ell,1)}$ for selected values $\ell \in [0.1, \sqrt{2}]$. We compute reference solutions for $\ell = 0.1, 0.5, 1.4$ using the full matrix. The reduced basis is constructed using 10 snapshots

$$\ell^{\text{snap}} = (2^{1/2}, (2^{-1/2}+1)^{-1}, (2^{-1/2}+2)^{-1}, \dots, (2^{-1/2}+9)^{-1}).$$

We compute the leading 100 eigenpairs for all correlation lengths in ℓ^{snap}, and assemble the associated eigenvectors in a single matrix. Then we apply the POD and retain $N_{\text{RB}} = 2^1, \dots, 2^{13}$ orthonormal basis vectors. Recall that the offline-online decomposition requires a linearisation of the covariance operator (see §4.3). Throughout this section (§6) we use $N_{\text{lin}} = 39$ linearisation terms.

We plot the relative error of the reduced compared to the exact eigenvalue in Figure 6.1 for various reduced basis dimensions N_{RB}. Note that it is not possible

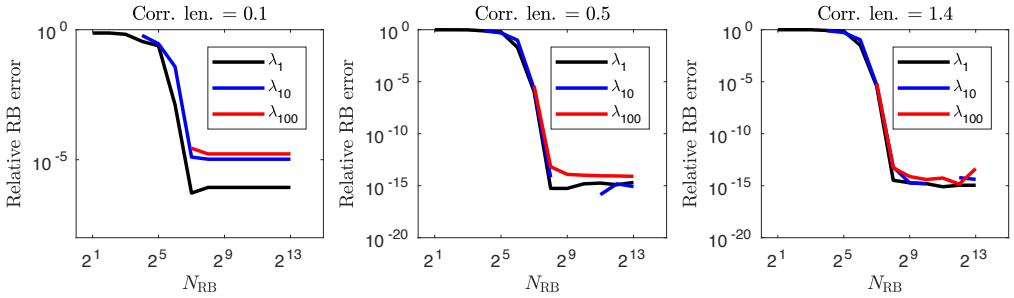

Figure 6.1: Relative reduced basis error of the eigenvalues $\lambda_1(\ell), \lambda_{10}(\ell), \lambda_{100}(\ell)$ for correlation lengths $\ell = 0.1, 0.5, 1.4$ and reduced basis dimensions $N_{\mathrm{RB}} = 2^1, ..., 2^{13}$.

to compute eigenvalues with an index larger than N_{RB}. For $\ell = 0.1$ the relative reduced basis error stagnates at a level that is not smaller than 10^{-6}. In further experiments not reported here we observed that this stagnation is caused by the relatively large linearisation error of the covariance kernel (recall that we use only $N_{\mathrm{lin}} = 39$ terms). We remark that the root mean square error of the MC and MCMC estimation results in this section is of order $O(10^{-2})$. Hence, an eigenvalue error of magnitude 10^{-6} is acceptable. We point out, however, that a full error analysis of the reduced basis samplers (including the linearisation and RB error) is beyond the scope of this study. For $\ell = 0.5$ and $\ell = 1.4$ we achieve the accuracy $O(10^{-6})$ for $N_{\mathrm{RB}} \approx 128$. For $N_{\mathrm{RB}} > 128$ the relative errors are of the size of the machine epsilon. This error unnecessarily outperforms the sampling error mentioned above and introduces a higher computational cost in the online phase. Hence, in our test problems $N_{\mathrm{RB}} = 128$ would be a sufficient choice.

To get an idea about the speed-ups that are possible with reduced basis sampling we repeat the experiment. This time we vary the dimension of the finite element space and use $N = 4^4, \ldots, 4^7$. The dimension of the reduced basis is fixed with $N_{\mathrm{RB}} = 256$. We plot the test results in Figure 6.2. The time measurements correspond to serial simulations in MATLAB with an Intel i7 (2.6 Ghz) CPU and 16 GB RAM memory. The dashed lines show the theoretical asymptotic behaviour, that is, $O(N; N \to \infty)$ for the reduced basis sampling and $O(N^2; N \to \infty)$ for the full sampling. We see that the theoretical and observed timings for the full sampling are almost identical. In contrast, the observed timings for RB sampling are smaller than predicted by the theory. This is caused by the fact that the dimension N of the finite element space is quite small. As N increases we observe a massive speed-up of the reduced basis sampling compared to the full sampling. For example, for $N = 4^7$ the reduced basis sampling requires less than 10^{-1} seconds, while the full sampling requires several minutes. In this case the speed-up in the online phase is of order $\mathcal{O}(10^3)$.

For the estimation problems in Examples 6.1–6.4 we use Monte Carlo and Markov Chain Monte Carlo with 10^4 up to $1.5 \cdot 10^5$ samples and a random field resolution with $N = 256 \times 256$ finite elements in space. In these cases, MC and MCMC estimations based on the full KL eigenproblem would take a couple of days up to a couple of years to terminate. In contrast, the (serial) run-time of the reduced

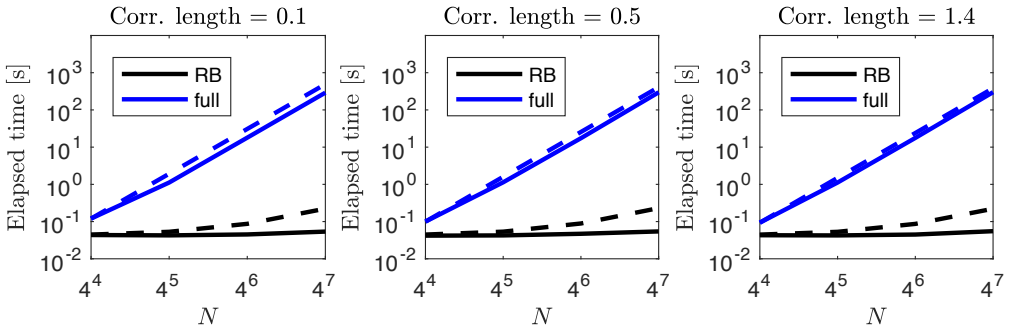

Figure 6.2: Timings for the full and reduced problem with different FE resolutions and correlation lengths. The elapsed time is shown as solid line, and the asymptotic behaviour is shown as dashed line.

basis sampling is ~ 15 minutes in Example 6.2 and ~ 18 hours in Example 6.4. Of course, standard Monte Carlo simulations are trivially parallelisable. MCMC is a serial algorithm by design, and parallelisation is not trivial, see e.g. [50] for suitable strategies. Our experiments show that RB sampling can reduce the computational cost without the need for parallelisation.

6.2. Verification of reduced basis sampling

Next, we test the accuracy of RB sampling using coarse spatial discretisations. This allows us to obtain reference solutions in a reasonable amount of time.

Example 6.1 *Let μ be the joint probability measure from Example 2.9 together with the parameter values given in Table 3. We discretise the random field $\theta \sim \mu''$ using 32^2 finite elements. The test problems are as follows.*

(a) Forward uncertainty propagation: We consider a flow cell in 2D with log-permeability θ. See Example 6.2 for the definition of the flow cell. Given the random coefficient θ we want to estimate the probability distribution of the outflow over the boundary $\mathcal{Q}(\theta)$. We discretise the PDE with $2 \cdot 16^2$ finite elements.

(b) Bayesian inverse problem: We observe a random field realisation on $D = (0,1)^2$ at the centre point $(0.5, 0.5)$. Given this (noisy) observation $y = 0.1$ we want to reconstruct the random field. The prior measure is μ as specified above. The likelihood is given by

$$\exp\left(-1/(2 \cdot 10^{-3}) \|y - \theta(0.5, 0.5)\|^2\right).$$

We want to estimate the posterior mean and variance of the correlation length ℓ given the data y. In addition, we compute the model evidence.

We solve the test problems in Example 6.1(a)–(b) with Monte Carlo. In part (b) we use importance sampling with samples from the prior as proposal. We solve (a) and

(b) with reduced basis sampling as well as standard sampling based on the full (discretised) eigenvectors of the parameterised covariance operator. The standard sampling serves as reference solution for the reduced basis sampling. The reduced basis is constructed using the snapshot correlation lengths $\ell^{\text{snap}} = (0.322, 0.433, 0.664, 1.414)$; these are simply the inverses of four equidistant points in the interval $[1/\sqrt{2}, 3.11]$ including the boundary points. This choice clusters snapshots near zero which is desirable due to the singularity of the exponential covariance at $\ell = 0$. We apply the POD to construct three reduced bases with different accuracies $\underline{\lambda} := 10^{-1}, 10^{-5}, 10^{-9}$.

For each of the settings we run 61 Monte Carlo simulations with 10^4 samples each to estimate the mean and the variance of the pushforward measure $\mu''(\mathcal{Q} \in \cdot)$ in part (a), as well as the posterior mean, posterior variance and model evidence in the Bayesian inverse setting in part (b). We compute a reference solution for all those quantities using 6.15×10^5 samples. With respect to the reference solutions we compute the relative error of the 61 estimates in each setting. In Table 4 we give the means and the associated standard deviations (StD) of the relative errors. We observe that the (mean of the) relative error is of order $O(10^{-2})$ up to $O(10^{-3})$.

$\underline{\lambda}$	10^{-1}	(StD)	10^{-5}	(StD)	10^{-9}	(StD)
Pushforward mean	0.0055	(0.0041)	0.0057	(0.0051)	0.0054	(0.0040)
Pushforward variance	0.0451	(0.0262)	0.0442	(0.0352)	0.0321	(0.0282)
Evidence	0.0169	(0.0116)	0.0156	(0.0129)	0.0265	(0.0167)
Posterior mean	0.0098	(0.0076)	0.0077	(0.0059)	0.0093	(0.0067)
Posterior variance	0.0914	(0.0706)	0.0720	(0.0554)	0.0867	(0.0623)

Table 4: Relative errors in the Monte Carlo estimation of the mean and variance of the pushforward measure, and posterior mean, posterior variance and model evidence in the Bayesian inverse problem (Example 6.1). Each error value is the mean taken over 61 simulations with 10^4 samples each. The simulations are performed with reduced basis sampling with POD accuracies $\underline{\lambda} = 10^{-1}, 10^{-5}, 10^{-9}$. The relative errors are computed with respect to a reference solution computed with 6.15×10^5 samples based on the full eigenproblem.

6.3. Forward uncertainty propagation

Next we study the forward uncertainty propagation of a hierarchical random field through an elliptic PDE.

Example 6.2 *Consider a flow cell problem on $D = (0,1)^2$ where the flow only takes place in the x_1-direction. The boundary conditions are as follows,*

$$\begin{aligned} p(x) &= 0 \quad (x \in \{1\} \times [0,1]), \\ p(x) &= 1 \quad (x \in \{0\} \times [0,1]), \\ \frac{\partial p}{\partial \vec{n}}(x) &= 0 \quad (x \in (0,1) \times \{0,1\}). \end{aligned}$$

There are no sources within the flow cell ($f \equiv 0$). The random field θ is as in Example 2.9 with the parameters given in Table 3. The PDE is discretised with

2×128^2 *finite elements, and the random field with* 256^2 *finite elements. The quantity of interest is the outflow over the (western) boundary* $\Gamma_{\text{out}} := \{0\} \times [0,1]$. *It can be approximated by*

$$\mathcal{Q}(\theta) := -\int_D \kappa(\theta)\nabla p \cdot \nabla\psi \mathrm{d}x,$$

where $\psi|_{D\setminus\Gamma_{\text{out}}} \approx 0$ *and* $\psi|_{\Gamma_{\text{out}}} \approx 1$ *(see e.g. [17]). We discretise the outflow using a piecewise linear, continuous finite element function* ψ *on the same mesh that we used for the PDE discretisation.*

The log-permeability is modelled as a parameterised Gaussian random field. We employ the reduced basis sampling with $N_{\text{RB}} = 191$. We construct the reduced basis analogously to the simple test setting in §6.2. However, now we let $N_{\text{sto}} = 100$, and remove vectors from the POD where the corresponding eigenvalue is smaller than $\underline{\lambda} = 10^{-10}$ (see also §4.2.1).

We estimate the mean and variance of the output quantity of interest. We compare 24 estimations by computing the associated coefficient of variation (CoV) for the mean and variance estimator, respectively. The CoV is defined as the ratio of the standard deviation of the estimator and the absolute value of its mean. We present the estimation results in Table 5. The small CoVs tell us that $N_{\text{smp}} =$

	Mean	(CoV)	Variance	(CoV)
MC estimate	157.286	(0.0028)	3012.2	(0.0355)

Table 5: Mean and variance estimates with 10^4 samples (Example 6.2). We compare these estimates to 23 further simulation results by computing the coefficient of variation within the 24 estimates.

10^4 samples were sufficient to estimate the pushforward measure of the quantity of interest, as well as its mean and variance. Note that with the reduced basis sampling a single Monte Carlo simulation run took about 18 minutes.

6.4. Hierarchical Bayesian inverse problem

We consider two hierarchical Bayesian inverse problems based on random fields. Note that we use again 256^2 finite elements to discretise the random fields in both problems and 2×128^2 finite elements to discretise the elliptic PDE in Example 6.3.

Example 6.3 *Consider the Bayesian estimation of a random field which is propagated through the elliptic PDE* (2.4) *together with Dirichlet boundary conditions*

$$p(x) = 0 \qquad (x \in \partial D),$$

and 9 Gaussian-type source terms

$$f(x) = \sum_{n,m=1}^{3} \mathrm{n}\left(x_1; 0.25n, 0.001\right) \cdot \mathrm{n}\left(x_2; 0.25m, 0.001\right).$$

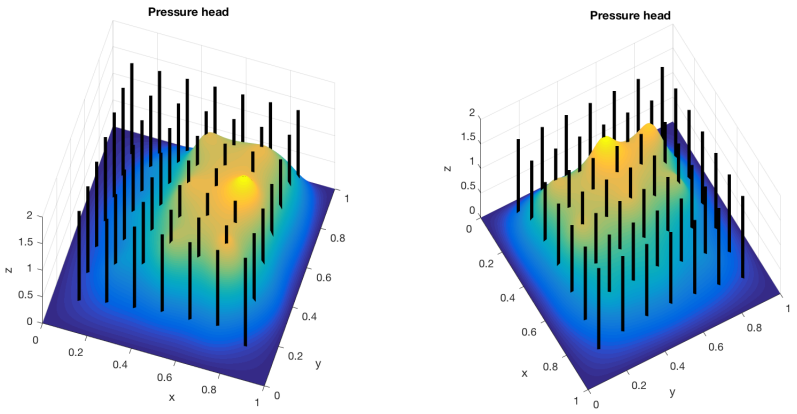

Figure 6.3: Synthetic solution p observed in the Bayesian inverse problem (Example 6.3). The black lines indicate the measurement locations. The figures show the model output for a log-permeability with correlation length $\ell = 0.5$ (left) and $\ell = 1.1$ (right), respectively.

We observe the solution p at 49 locations. In particular, the observation operator is given by

$$\mathcal{O}(p) := (p(n/8, m/8) : n, m = 1, ..., 7).$$

The (synthetic) observations are generated with log-permeability fields that are samples of a Gaussian random field with exponential covariance operator with $\ell \in \{0.5, 1.1\}$ and $\sigma = \sqrt{1/2}$. We show the corresponding PDE outputs and the measurement locations in Figure 6.3. The Gaussian random fields have been sampled with the full (unreduced) $N_{\text{sto}} = 100$ leading KL terms. Every observation is perturbed with i.i.d. Gaussian noise $\eta_1 \sim \mathrm{N}(0, 10^{-6})$. We use the measure μ in Example 2.9 with parameter values given in Table 3 as prior measure.

Example 6.4 *Consider the Bayesian estimation of a Gaussian random field together with its standard deviation and correlation length. We asssume that we can observe the field directly, however, the observations are again noisy. We perform estimations with two data sets that have been generated with fixed hyperparameters $\ell \in \{0.2, 1.1\}$ and $\sigma = 1/(\sqrt{2} \cdot 256)$. We have set $\sigma = 1/\sqrt{2}$ and rescaled the KL eigenfunctions by $1/256$. We discretise the random field using an $N_{\text{sto}} = 800$ dimensional full (unreduced) KL basis. We observe the random field at 2500 positions. Each observation is perturbed by i.i.d. Gaussian noise $\eta_1 \sim \mathrm{N}(0, 10^{-6})$. As prior we consider the measure μ in Example 2.9 with parameter values given Table 3.*

We remark that in Examples 6.3–6.4 we use the same PDE and random field discretisation for the generation of the data and the estimation problem. The reason is that we are mainly interested in the reduced basis error, and not in the reconstruction error of the inverse problem. Note further that in Examples 6.3–6.4 the standard deviation $\sigma = \sqrt{1/2}$ is fixed a priori, and is not estimated. The hierarchical Bayesian inverse problems in Example 6.3–6.4 are well-posed since the associated Bayesian inverse problem with fixed, deterministic hyperparameters is well-posed (see [13]), and since the hyperparameter set R is compact.

6.4.1. Observations from PDE output

We consider Example 6.3 and the settings in Table 3. We employ the Reduced Basis MCMC method presented in Algorithm 4 to sample from the posterior measure. We use correlation lengths $\ell \in [0.3, \sqrt{2}]$. Since this is the same range as in Example 6.2 we reuse the reduced basis computed for Example 6.2. Recall that the standard deviation $\sigma = 1/\sqrt{2}$ of the random field θ is fixed and not estimated. Moreover, we assume that the observational noise is given by $\eta \sim \mathrm{N}(0, 10^{-3}\mathrm{Id})$. This corresponds to a noise level of $\sqrt{10^{-3}}/\|y\|_Y \approx 0.6\%$.

We perform experiments for two synthetic data sets with $\ell = 0.5$ and $\ell = 1.1$, respectively. For both data sets we compute a Markov chain of length $N_{\mathrm{smp}} = 10^5$. To avoid burn-in effects we choose initial states close to the true parameter values for the Markov chains. In a setting with real world data it is often possible to obtain suitable initial states with Sequential Monte Carlo (see [3]).

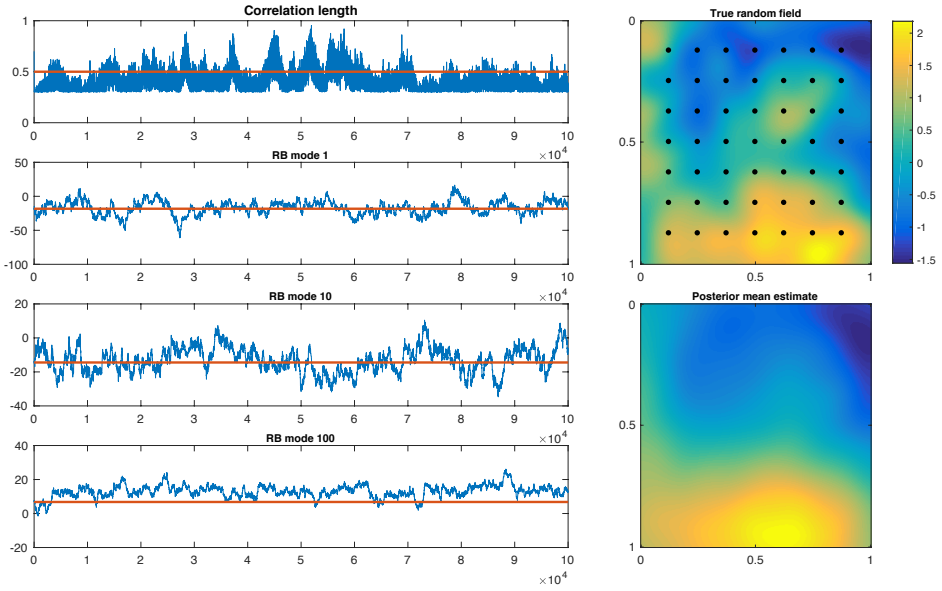

Figure 6.4: Results of the MCMC estimation (Example 6.3, $\ell = 0.5$). The top-right plot shows the synthetic truth together with the measurement locations (black dots). Below we plot the posterior mean estimate computed with MCMC. The four path plots on the left side of the figure show the Markov chains for the correlation length ℓ, and the reduced basis modes $(\theta_{\mathrm{RB}})_1$, $(\theta_{\mathrm{RB}})_{10}$, and $(\theta_{\mathrm{RB}})_{100}$, respectively. The red lines mark the truth.

We show the estimation results in Figure 6.4 and Figure 6.5. We observe in both figures that the Markov chain for ℓ mixes very fast, however, it takes some time for the Markov chains of the reduced basis modes to explore the whole space. To investigate this further we conduct a heuristic convergence analysis. To this end we consider multiple Markov chains (see §12.1.2 in [41] for a review of MCMC convergence analysis with multiple Markov chains). For each of the two test data sets we compute 4 additional Markov chains starting at different initial states. In results not reported here we observed a similar mixing and coverage of the parameter space of the additional chains. Given these mixing properties, it can reasonably be assumed that the Markov chains have reached the stationary regime.

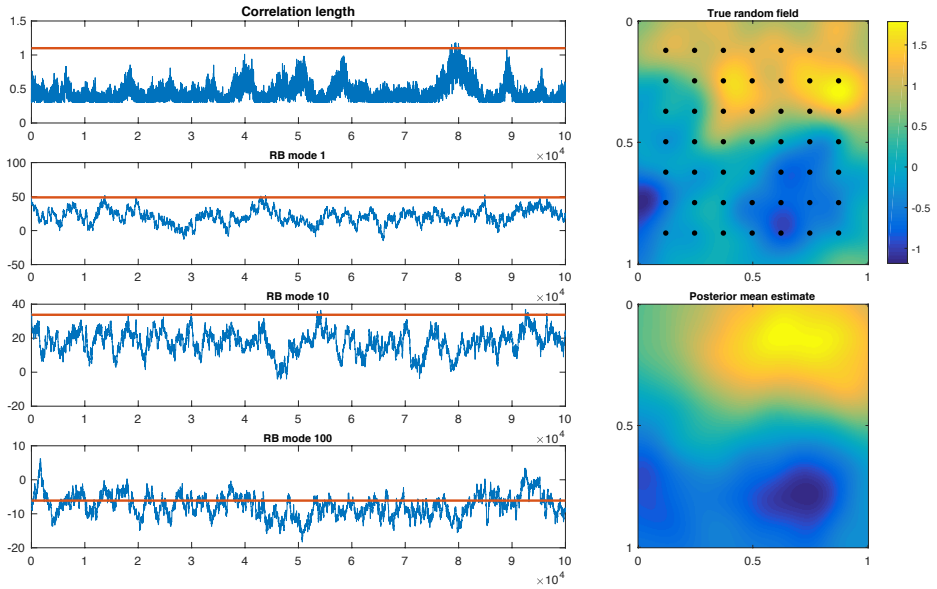

Figure 6.5: Results of the MCMC estimation (Example 6.3, $\ell = 1.1$). The top-right plot shows the synthetic truth together with the measurement locations (black dots). Below we plot the posterior mean estimate computed with MCMC. The four path plots on the left side of the figure show the Markov chains for the correlation length ℓ, and the reduced basis modes $(\theta_{\mathrm{RB}})_1$, $(\theta_{\mathrm{RB}})_{10}$, and $(\theta_{\mathrm{RB}})_{100}$, respectively. The red lines mark the truth.

Moreover, we have computed posterior mean and posterior variance of the correlation length parameter ℓ for each of the 5 Markov chains. We assess the accuracy of these estimates by computing the coefficient of variation (CoV) within these five estimates. We tabulate the posterior mean and variance estimates for ℓ of a single Markov chain as well as the associated CoVs in Table 6. The single Markov chains in this table are the ones shown in Figures 6.4-6.5. The coefficients of variation of the posterior mean and variance estimates are considerably small. This tells us that the posterior mean and variance estimates are reasonably accurate.

	Mean	(CoV)	Variance	(CoV)
MCMC ℓ given y (Truth: $\ell = 0.5$)	0.4105	(0.0040)	0.0081	(0.3235)
MCMC ℓ given y (Truth: $\ell = 1.1$)	0.4403	(0.0524)	0.0157	(0.2346)

Table 6: Estimation results of the Bayesian inverse problem with observations from PDE output (Example 6.3). We tabulate the posterior mean and variance estimates of the correlation length ℓ of one Markov chain each and the CoVs within the estimates of 5 different Markov chains.

Discussion of the estimation results. The correlation length is underestimated in both cases. In the first case, where the true parameter is given by $\ell = 0.5$, the posterior mean is close to the true parameter. The relative distance between truth and posterior mean is about 18%. In the second setting, where in truth $\ell = 1.1$, the posterior mean is far away from the true parameter. Here, the relative distance between truth and posterior mean is about 60%. In both cases, we conclude that the

data likelihood was not sufficiently informative to estimate the correlation length more accurately. However, we note that the succession of the estimates is correct: The posterior mean estimate in the problem with the larger true correlation is larger than the posterior mean estimate in the other case. Hence we observe a certain consistency with the data in the estimation.

6.4.2. Observations from a random field

Finally, we consider Example 6.4. Here, we allow for much smaller correlation lengths $\ell \in [0.1, \sqrt{2}]$. This requires more KL terms for an accurate approximation, in particular, we use the 800 leading KL terms. This also means that we cannot reuse the reduced basis computed in Example 6.2. Instead, we construct a reduced basis as follows. We solve the KL eigenproblem for 5 snapshots

$$\ell^{\text{snap}} = (0.1148, 0.1491, 0.2124, 0.3694, 1.4142)$$

of the correlation length. The rationale behind this choice is explained in §6.3. Given the collection of snapshot KL eigenvectors we apply a POD and retain only the basis vectors with $\lambda_i^{\text{snap}} \geq \underline{\lambda} = 10^{-10}$.

Recall that in this example the observational noise is given by $\eta \sim \mathrm{N}(0, 10^{-4}\mathrm{Id})$. This corresponds to a noise level of $\sqrt{10^{-4}}/\|y\|_Y \approx 6.6\%$. We employ the Reduced

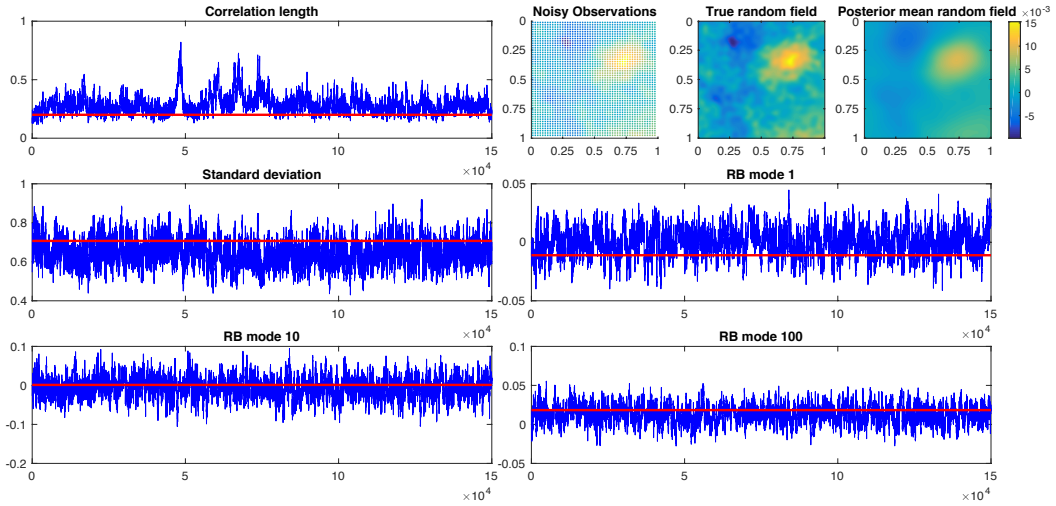

Figure 6.6: Results of the MCMC estimation given random field observations (Example 6.4, $\ell = 0.2$, $\sigma = 1/\sqrt{2}$). In the top-right corner we plot the positions and values of the noisy observations (left), the synthetic truth (middle), and the posterior mean (right). The five path plots show the Markov chains for ℓ and σ, and the reduced basis modes $(\theta_{\text{RB}})_1$, $(\theta_{\text{RB}})_{10}$, $(\theta_{\text{RB}})_{100}$, respectively. The red lines mark the truth.

Basis MCMC sampler in Algorithm 4 to generate $N_{\text{smp}} = 1.5 \times 10^5$ samples of the posterior measure. We present the Markov chains and estimation results in Figure 6.6 and in Figure 6.7. We observe a fast mixing of the Markov chains. To conduct a heuristic convergence assessment we again compute 4 additional Markov chains with $N_{\text{smp}} = 1.5 \times 10^5$ samples each and different initial states. We found that the additional Markov chains mix similarly compared to the Markov chains shown in

Figures 6.6–6.7. They also cover the same area of the parameter space. Hence, it can reasonably be concluded that the Markov chains reached a stationary regime.

	Mean	(CoV)	Variance	(CoV)
MCMC ℓ given y (Truth: $\ell = 0.2$)	0.2847	(0.0465)	0.0064	(0.1520)
MCMC σ given y (Truth: $\sigma = 1/\sqrt{2}$)	0.6438	(0.0077)	0.0042	(0.0207)
MCMC ℓ given y (Truth: $\ell = 1.1$)	0.7248	(0.0161)	0.0575	(0.1308)
MCMC σ given y (Truth: $\sigma = 1/\sqrt{2}$)	0.5484	(0.0096)	0.0052	(0.0453)

Table 7: Estimation results of the Bayesian inverse problem with observations from a random field (Example 6.4). We tabulate the posterior mean and variance of the correlation length ℓ and standard deviation σ.

In addition we present in Table 7 the posterior mean and posterior variance estimates of ℓ and σ associated with the Markov chains given in Figures 6.6–6.7. To assess the accuracy of these estimates we compare them with the posterior mean and variance estimates of the 4 other Markov chains by computing the coefficients of variations of the estimators. Again, the coefficients of variation are reasonably small.

Discussion of the estimation results. While the likelihood was rather uninformative in the PDE-based Bayesian inverse problem, we see overall more consistent estimates in Example 6.4. For the short correlation length $\ell = 0.2$ the relative distance between posterior mean and truth is 42%. The long correlation length $\ell = 1.1$ is again underestimated. The relative distance between truth and posterior mean is 34% in this case. This result could be explained by the uncorrelated noise that has an influence on the observation of the correlation structure. In particular, we actually observe a random field $\theta' := \theta + \eta'$, where $\theta \sim \mathrm{N}(0, \mathcal{C}_{\mathrm{exp}}^{(\ell,\sigma)})$ and $\eta' \sim \mathrm{N}(0, \sigma_*^2 \cdot \mathrm{Id}_X)$, for some $\sigma_*^2 > 0$. In this situation, the random field η' can be understood as a random field with correlation length 0. This might explain the underestimation of the correlation lengths. The standard deviations are slightly underestimated and some of the reduced basis modes are overestimated – this is a consistent result. The posterior mean random fields appear to be smoother than the true random fields. This might be due to the high noise level.

7. Conclusions

We developed a mathematical and computational framework for working with parameterised Gaussian random fields arising from hierarchical forward and inverse problems in uncertainty quantification. Under weak assumptions we proved the well-posedness of the associated hierarchical problems. We discretised the family of parameterised Gaussian random fields by (parametric) KL expansions. We showed how the overall discretisation cost can be reduced substantially by a reduced basis surrogate for the parametric KL eigenpairs. Moreover, we developed a reduced basis sampler for use with Monte Carlo and Markov chain Monte Carlo. For Matérn-type

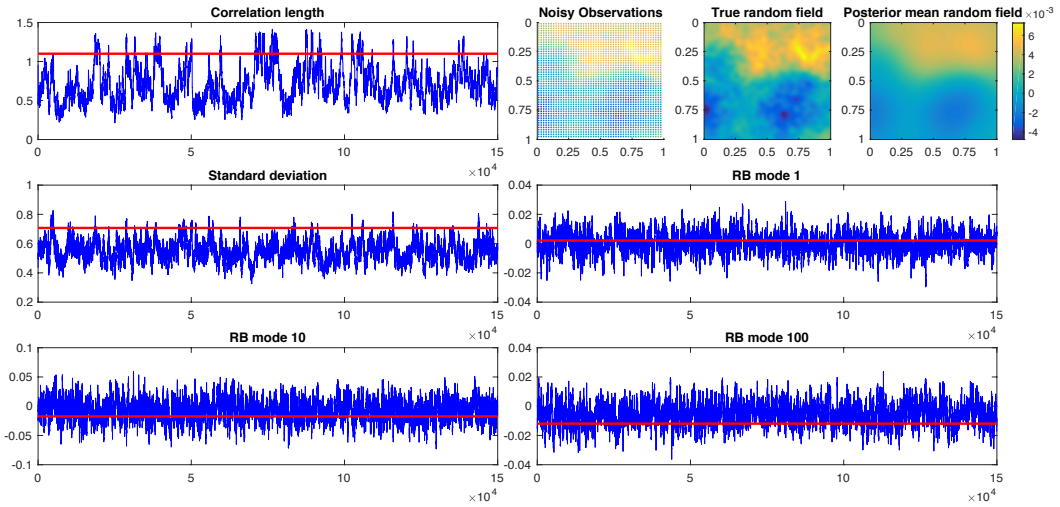

Figure 6.7: Results of the MCMC estimation given random field observations (Example 6.4, $\ell = 1.1$, $\sigma = 1/\sqrt{2}$). In the top-right corner we plot the positions and values of the noisy observations (left), the synthetic truth (middle), and the posterior mean (right). The five path plots show the Markov chains for ℓ and σ, and the reduced basis modes $(\theta_{\mathrm{RB}})_1$, $(\theta_{\mathrm{RB}})_{10}$, $(\theta_{\mathrm{RB}})_{100}$, respectively. The red lines mark the truth.

covariance operators with uncertain correlation length we suggested and analysed a linearisation technique to enable an efficient offline-online decomposition for the reduced basis solver. We illustrated the accuracy and speed-up of the reduced basis surrogate and RB sampling in simple low-dimensional test problems. Finally, we applied the reduced basis sampling to more realistic, high-dimensional, forward and Bayesian inverse test problems in 2D physical space. The test results illustrate that the parametric KL eigenproblem can be approximated with acceptable accuracy by a reduced basis surrogate. Moreover, the RB sampling gives acceptable accuracies compared to the full, unreduced sampling. This enables an efficient hierarchical uncertainty quantification with parameterised random fields. Of course, the size of the acceptable error level is heavily problem dependent, and our numerical experiments are a proof-of-concept. A rigorous error analysis of the reduced basis samplers including the RB error and the linearisation error is beyond the scope of this study.

A. Well-posed Bayesian inverse problem

In §2.4 we discuss the well-posedness of the Bayesian inverse problem for hyperpriors. We recall the well-posedness definition by Hadamard [25].

Definition A.1 *A problem is well-posed in the sense of Hadamard, if the following holds;*

1. *the problem has a solution,*
2. *the solution is unique,*
3. *the solution depends continuously on the inputs.*

If the problem is a parameter identification or inverse problem, then the third assumption in Definition A.1 postulates that the solution depends continuously on the observations. Under certain assumptions on the likelihood and prior it can be proved that Bayesian inverse problems are well-posed. We state and explain these assumptions, following and summarising the results in [48].

Let X denote a separable Hilbert space and let $X' \subseteq X$ denote a separable Banach space. The operator $\mathcal{G} : X' \to Y$ is called *forward-response operator*. It maps a feasible parameter θ to the observation space $Y := \mathbb{R}^{N_{\text{obs}}}$. The observations $y \in Y$ are given by $y = \mathcal{G}(\theta) + \eta$, where $\eta \sim \mathrm{N}(0, \Gamma)$ is non-degenerate Gaussian noise.

Let μ_0 be a Gaussian measure on $(X, \mathcal{B}X)$ that is concentrated on X', i.e. $\mu_0(X') = 1$. We assume that the parameter θ is distributed according to the prior measure μ_0. The solution of the Bayesian inverse problem is the *posterior measure*

$$\mu^y := \mathbb{P}(\theta \in \cdot | \mathcal{G}(\theta) + \eta = y).$$

It can be written in terms of a Radon-Nikodym derivative w.r.t. the prior measure μ_0 as follows:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0} = Z_y^{-1} \exp(-\Phi(\cdot; y)),$$

where $\Phi(\theta; y) := \frac{1}{2}\|\Gamma^{-1/2}(\mathcal{G}(\theta) - y)\|_X^2$, and $Z_y := \int \exp(-\Phi(\theta; y))\mathrm{d}\mu_0$. The operator Φ is a so-called *potential*. If the potential satisfies certain conditions, then the associated Bayesian inverse problem is well-posed.

Assumptions A.2 (Potential) *Consider a potential $\Phi^\dagger : X \times Y \to \mathbb{R}$ which satisfies the following conditions:*

1. *For every $\varepsilon, r > 0$ there is a constant $M(\varepsilon, r) \in \mathbb{R}$ such that*

$$\Phi^\dagger(\theta; y) \geq M(\varepsilon, r) - \varepsilon\|\theta\|_X^2. \qquad (\theta \in X, y \in Y, \text{ where } \|y\|_Y < r)$$

2. *For every $r > 0$ there is a constant $K(r) > 0$ such that*

$$\Phi^\dagger(\theta; y) \leq K(r). \qquad (\theta \in X, y \in Y, \text{ where } \max\{\|\theta\|_X, \|y\|_Y\} < r)$$

3. *For every $r > 0$ there is a constant $L(r) > 0$ such that*

$$|\Phi^\dagger(\theta_1; y) - \Phi^\dagger(\theta_2; y)| < L(r)\|\theta_1 - \theta_2\|_X. \qquad (\theta_1, \theta_2 \in X, y \in Y, \text{ where } \max\{\|\theta_1\|_X, \|\theta_2\|_X, \|y\|_Y\} < r)$$

4. *For every $\varepsilon, r > 0$ there is a constant $C(\varepsilon, r) \in \mathbb{R}$ such that*

$$|\Phi^\dagger(\theta; y_1) - \Phi^\dagger(\theta; y_2)| \leq \exp(\varepsilon\|\theta\|_X^2 + C(\varepsilon, r))\|y_1 - y_2\|_Y. \quad (\theta \in X, y_1, y_2 \in Y, \text{ where } \max\{\|y_1\|_Y, \|y_2\|_Y\} < r)$$

Theorem A.3 *Assume that the potential Φ satisfies Assumptions A.2. Then μ^y is a well-defined probability measure on $(X, \mathcal{B}X)$ for any $y \in Y$. Moreover, the function*

$$(Y, \|\cdot\|_2) \ni y \mapsto \mu^y \in (\mathbb{P}(X, \mathcal{B}X, \mu_0), \mathrm{d_{Hel}})$$

is locally Lipschitz-continuous. Hence, the Bayesian inverse problem is well-posed in this setting.

Proof. See Theorems 4.1 and 4.2 in [48]. □

B. Proofs of auxiliary results

Lemma B.1 *Let $\nu \in (0, \infty]$, and let Assumption 4.5 hold. Then, there is a linearly separable operator*

$$\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}}) : X \to X$$

consisting of $N_{\mathrm{lin}} \in 2\mathbb{N}$ terms, such that asymptotically

$$\|\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}}) - \mathcal{C}(\nu, \ell, \sigma)\|_X \leq O(1/(N_{\mathrm{lin}}!); N_{\mathrm{lin}} \to \infty). \tag{B.1}$$

Proof. Let $\nu \in (0, \infty)\backslash\mathbb{N}$. Consider the function

$$f_{(\nu)} : [0, \infty) \to [0, \infty), \quad \zeta \mapsto \zeta^\nu \cdot K_\nu(\zeta).$$

It holds $f_{(\nu)}(\sqrt{2\nu}z/\ell) = \mathrm{const}(\nu, \sigma)c(\nu, \ell, \sigma)(z)$, where $\mathrm{const}(\nu, \sigma) > 0$ is a constant that does not depend on the correlation length ℓ. Moreover, we assume that we work in a bounded computational domain D, and that ℓ is bounded from below by a fixed positive constant $\underline{\ell} > 0$ (see Assumption 4.5). Now, for $\nu \in (0, \infty)\backslash\mathbb{N}$ the function $f_{(\nu)}$ can be written in terms of a series

$$f_{(\nu)}(\zeta) = \frac{\pi \csc(\pi\nu)}{2} \sum_{k=0}^{\infty} \left(\frac{1}{2^{2k-\nu}\Gamma(k-\nu+1)k!} - \frac{\zeta^{2\nu}}{2^{2k+\nu}\Gamma(k+\nu+1)k!} \right) \zeta^{2k}.$$

If we truncate the series after the first $1 + N_{\mathrm{lin}}/2$ terms we obtain the following function:

$$f_{(\nu, N_{\mathrm{lin}})}(\zeta) = \frac{\pi \csc(\pi\nu)}{2} \sum_{k=0}^{M} \left(\frac{1}{2^{2k-\nu}\Gamma(k-\nu+1)k!} - \frac{\zeta^{2\nu}}{2^{2k+\nu}\Gamma(k+\nu+1)k!} \right) \zeta^{2k}$$

This (truncated) series expansion is associated with the integral operator $\widetilde{\mathcal{C}}(\nu, \ell, \sigma, N_{\mathrm{lin}})$ that is given by the kernel

$$c(\nu, \ell, \sigma, N_{\mathrm{lin}})(z) := \frac{f_{(\nu, N_{\mathrm{lin}})}(\sqrt{2\nu}z/\ell)}{\mathrm{const}(\nu, \sigma)}. \tag{B.2}$$

Now, we bound the asymptotic truncation error, and assume w.l.o.g. that $N_{\text{lin}} > \nu$. Note that then $\Gamma(k-\nu+1) > 1$. Moreover, $\zeta > 0$. Then,

$$\begin{aligned}
|f_{(\nu,N_{\text{lin}})}(\zeta) - f_{(\nu)}(\zeta)| &\leq \frac{\pi|\csc(\pi\nu)|}{2}\left|\sum_{k=N_{\text{lin}}+1}^{\infty}\left(\frac{1-\zeta^{2\nu}}{2^{2k-\nu}\Gamma(k-\nu+1)k!}\right)\zeta^{2k}\right| \\
&\leq \frac{\pi|\csc(\pi\nu)|}{2^{1-\nu}}(1+\zeta^{2\nu})\sum_{k=N_{\text{lin}}+1}^{\infty}\left(\frac{1}{\Gamma(k-\nu+1)k!}\right)\left(\frac{\zeta}{2}\right)^{2k} \\
&\leq \frac{\pi|\csc(\pi\nu)|}{2^{1-\nu}}(1+\zeta^{2\nu})\sum_{k=N_{\text{lin}}+1}^{\infty}\left(\frac{1}{k!}\right)\left(\frac{\zeta}{2}\right)^{2k},
\end{aligned}$$

using the triangular inequality. The infinite sum on the right-hand side can be bounded by the remainder term of a Taylor series approximation of the exponential function with N_{lin} terms at $\zeta = 0$. Hence,

$$\begin{aligned}
|f_{(\nu,N_{\text{lin}})}(\zeta) - f_{(\nu)}(\zeta)| &\leq \frac{\pi|\csc(\pi\nu)|}{2^{1-\nu}}(1+\zeta^{2\nu})\exp\left(\frac{\zeta_{\max}^2}{4}\right)\frac{\zeta^{2N_{\text{lin}}+2}}{(N_{\text{lin}}+1)!} \\
&\leq \frac{\pi|\csc(\pi\nu)|}{2^{1-\nu}}(1+\zeta_{\max}^{2\nu})\exp\left(\frac{\zeta_{\max}^2}{4}\right)\frac{\zeta_{\max}^{2N_{\text{lin}}+2}}{(N_{\text{lin}}+1)!} =: \text{const}'(N_{\text{lin}}),
\end{aligned}$$

where $\zeta_{\max} = \frac{\text{diam}(D)}{\underline{\ell}}$. Finally, let $\varphi \in X$. Then it holds

$$\begin{aligned}
&\|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\text{lin}})\varphi - \mathcal{C}(\nu,\ell,\sigma)\varphi\|_X^2 \\
&= \int_D\left(\int_D(\widetilde{c}^{(\nu,\ell,\sigma,N_{\text{lin}})}(\text{dist}(x,y)) - c(\nu,\ell,\sigma)(\text{dist}(x,y)))\varphi(x)\text{d}x\right)^2\text{d}y \\
&\leq \int_D\left(\int_D(\widetilde{c}^{(\nu,\ell,\sigma,N_{\text{lin}})}(\text{dist}(x,y)) - c(\nu,\ell,\sigma)(\text{dist}(x,y)))^2\text{d}x\right)\cdot\left(\int_D\varphi(x)^2\text{d}x\right)\text{d}y \\
&\leq \text{Leb}(d)(D)^2\cdot\text{const}'(N_{\text{lin}})^2\cdot\|\varphi\|_X^2,
\end{aligned}$$

by the Cauchy-Schwarz inequality. Taking the square root on both sides and dividing by $\|\varphi\|_X$ proves the error bound for the case $\nu \in (0,\infty)\backslash\mathbb{N}$. We now comment on the cases $\nu \in \mathbb{N}$ and $\nu = \infty$. Let $n \in \mathbb{N}\cup\{\infty\}$. The function $f_{(n)} := \lim_{\nu\uparrow n} f_{(\nu)}$ is analytic and its truncated power series can be used to construct a linearly separable operator analogously to the proof given above. □

Lemma B.2 *The operator $\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\text{lin}})$ is self-adjoint, trace-class and continuous.*

Proof. The integral operator $\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\text{lin}})$ is self-adjoint since the associated kernel function is symmetric. The operator is trace-class since D is a bounded domain, and $\int_D \widetilde{c}(\nu,\ell,\sigma,N_{\text{lin}})(\text{dist}(x,x))\text{d}x = \widetilde{c}(\nu,\ell,\sigma,N_{\text{lin}})(0)\cdot\text{Leb}(d)(D) < \infty$. The boundedness of D also implies the continuity of the operator. □

Lemma B.3 *The Matérn-type covariance operator $\mathcal{C}(\nu,\ell,\sigma)$ and the approximate operator $\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\text{lin}})$ in (4.4) satisfy*

$$\|\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\text{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X \leq 2\|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\text{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X.$$

Proof. Let $(\widetilde{\lambda}_i)_{i=1}^{\infty}$ denote the eigenvalues of the operator $\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})$. Without loss of generality, we assume that the spectrum of $\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})$ contains a negative eigenvalue. Since $\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})$ is trace-class, it holds $|\widetilde{\lambda}_i| \to 0$ for $i \to \infty$. Hence, there is an eigenpair $(\widetilde{\lambda}_{\max},\widetilde{\psi}_{\max})$ which realises the maximum in the expression

$$\max_{i\in\mathbb{N}\colon \widetilde{\lambda}_i<0} |\widetilde{\lambda}_i|. \tag{B.3}$$

Thus,

$$\|\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})\|_X = \|\sum_{i=1;\lambda_i<0}^{\infty} \widetilde{\lambda}_i\widetilde{\psi}_i\otimes\widetilde{\psi}_i\|_X = |\widetilde{\lambda}_{\max}|.$$

Moreover, since $\mathcal{C}(\nu,\ell,\sigma)$ is positive definite, we have $\widetilde{\psi}_{\max}^*\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})\widetilde{\psi}_{\max} \geq 0 > \widetilde{\psi}_{\max}^*\mathcal{C}(\nu,\ell,\sigma)\widetilde{\psi}_{\max}$. Hence, we obtain

$$\begin{aligned}
\|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X &\geq |\widetilde{\psi}_{\max}^*(\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma))\widetilde{\psi}_{\max}| \\
&= \widetilde{\psi}_{\max}^*\mathcal{C}(\nu,\ell,\sigma)\widetilde{\psi}_{\max} - \widetilde{\psi}_{\max}^*\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})\widetilde{\psi}_{\max} \\
&= \widetilde{\psi}_{\max}^*\mathcal{C}(\nu,\ell,\sigma)\widetilde{\psi}_{\max} - \widetilde{\lambda}_{\max}\widetilde{\psi}_{\max}^*\widetilde{\psi}_{\max} \\
&\geq |\widetilde{\lambda}_{\max}|.
\end{aligned}$$

This gives the bound

$$\|\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})\|_X \leq \|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X.$$

Finally, using the triangular inequality, we arrive at

$$\begin{aligned}
&\|\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X \\
&\quad\leq \|\widetilde{\mathcal{C}}_0(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}})\|_X + \|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X \\
&\quad\leq 2\|\widetilde{\mathcal{C}}(\nu,\ell,\sigma,N_{\mathrm{lin}}) - \mathcal{C}(\nu,\ell,\sigma)\|_X
\end{aligned}$$

□

Acknowledgements

The authors would like to thank Barbara Wohlmuth for pointing them to the work [28] and for many helpful comments. JL and EU would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme *Uncertainty quantification for complex systems: theory and methodologies* when work on this paper was undertaken. This work was supported by the DFG through the International Graduate School of Science and Engineering at the Technical University of Munich within the project 10.02 BAYES, and by EPSRC Grant Number EP/K032208/1.

References

[1] R Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS, 1990.

[2] P Benner, Y Qiu, and M Stoll. Low-rank computation of posterior covariance matrices for Bayesian inverse problems. *Preprint*, arXiv:1703.05638, 2017.

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

[4] W Betz, I Papaioannou, and D Straub. Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion. *Comput. Methods Appl. Mech. Eng.*, 271:109–129, 2014.

[5] D Bolin and K Kirchner. The SPDE approach for Gaussian random fields with general smoothness. *Preprint*, arXiv:1711.04333, 2017.

[6] D Bolin, K Kirchner, and M Kovàcs. Numerical solution of fractional elliptic stochastic PDEs with spatial white noise. *Preprint*, arXiv:1705.06565, 2017.

[7] D Calvetti, L Reichel, and D Sorensen. An Implicitly Restarted Lanczos Method for Large Symmetric Eigenvalue Problems. *Electron. Trans. Numer. Anal*, 2:1–21, 1994.

[8] G Chan and A T A Wood. An algorithm for simulating stationary Gaussian random fields. *J. R. Stat. Soc., Ser. C*, 46(1):171–181, 1997.

[9] J Charrier. Strong and Weak Error Estimates for Elliptic Partial Differential Equations with Random Coefficients. *SIAM J. Numer. Anal.*, 50(1):216–246, 2012.

[10] J Chen and M L Stein. Linear-cost covariance functions for Gaussian random fields. *Preprint*, arXiv:1711.05895, 2017.

[11] A Chernov, H Hoel, K J H Law, F Nobile, and R Tempone. Multilevel Ensemble Kalman Filtering for spatially extended models. *Preprint*, arXiv:1608.08558, 2016.

[12] A A Contreras, P Mycek, O P Le Maître, F Rizzi, B Debusschere, and O M Knio. Parallel Domain Decomposition Strategies for Stochastic Elliptic Equations Part A : Local KL Representations. *Preprint*, 2017. Available from https://perso.limsi.fr/olm/archives/paper_DDA.pdf.

[13] M Dashti and A M Stuart. Uncertainty Quantification and Weak Approximation of an Elliptic Inverse Problem. *SIAM J. Numer. Anal.*, 49(6):2524–2542, 2011.

[14] M Dashti and A M Stuart. The Bayesian Approach to Inverse Problems. In R Ghanem, D Higdon, and H Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 311–428. Springer, 2017.

[15] M D'Elia and M Gunzburger. Coarse-Grid Sampling Interpolatory Methods for Approximating Gaussian Random Fields. *SIAM/ASA J. Uncertain. Quantif.*, 1(1):270–296, 2013.

[16] C R Dietrich and G N Newsam. Fast and Exact Simulation of stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, 1997.

[17] J Douglas Jr., T Dupont, and M F Wheeler. A Galerkin procedure for approximating the flux on the boundary for elliptic and parabolic boundary value problems. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 8(R2):47–59, 1974.

[18] M M Dunlop, M A Iglesias, and A M Stuart. Hierarchical Bayesian level set inversion. *Statistics and Computing*, 27(6):1555–1584, 2017.

[19] O G Ernst and B Sprungk. Stochastic Collocation for Elliptic PDEs with Random Data: The Lognormal Case. In Jochen Garcke and Dirk Pflüger, editors, *Sparse Grids and Applications - Munich 2012*, pages 29–53, Cham, 2014. Springer International Publishing.

[20] B S Everitt and D J Hand. *Finite Mixture Distributions*. Springer Netherlands, Dordrecht, 1981.

[21] M Feischl, F Y Kuo, and I H Sloan. Fast random field generation with H-matrices. *Preprint*, arXiv:1702.08637, 2017.

[22] R Ghanem, D Higdon, and H Owhadi, editors. *Handbook of Uncertainty Quantification*. Springer, 2017.

[23] I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl, and I. H. Sloan. Analysis of circulant embedding methods for sampling stationary random fields. *Preprint*, arXiv:1710.00751, 2017.

[24] M Gu, A Ruhe, R Lehoucq, D Sorensen, R Freund, G Sleijpen, H van der Vorst, Z Bai, and R Li. *Hermitian Eigenvalue Problems*, chapter 4, pages 45–107. Society for Industrial and Applied Mathematics, 2000.

[25] J Hadamard. Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[26] H Harbrecht, M Peters, and M Siebenmorgen. Efficient approximation of random fields for numerical applications. *Numer. Linear Algebra Appl.*, 22(4):596–617, 2015.

[27] J S Hesthaven, G Rozza, and B Stamm. *Certified reduced basis methods for parametrized partial differential equations*. SpringerBriefs in Mathematics. Springer, Cham; BCAM Basque Center for Applied Mathematics, Bilbao, 2016.

[28] T Horger, B Wohlmuth, and T Dickopf. Simultaneous Reduced Basis Approximation of Parameterized Elliptic Eigenvalue Problems. *ESAIM: M2AN*, 51(2):443–465, 2017.

[29] T Horger, B Wohlmuth, and L Wunderlich. Reduced Basis Isogeometric Mortar Approximations for Eigenvalue Problems in Vibroacoustics. In P Benner, M Ohlberger, A Patera, G Rozza, and K Urban, editors, *Model Reduction of Parametrized Systems*, pages 91–106. Springer International Publishing, Cham, 2017.

[30] L Jiang and N Ou. Multiscale model reduction method for Bayesian inverse problems of subsurface flow. *J. Comput. Appl. Math.*, 319:188–209, 2017.

[31] K Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 37:1–79, 1947.

[32] B N Khoromskij, A Litvinenko, and H G Matthies. Application of hierarchical matrices for computing the Karhunen-Loève expansion. *Computing*, 84(1):49–67, 2009.

[33] M A Lifshits. *Gaussian Random Functions*. Springer, 1995.

[34] M Loeve. *Probability Theory II*. Springer, New York, 1978.

[35] A K Noor and J M Peters. Reduced Basis Technique for Nonlinear Analysis of Structures. *AIAA Journal*, 18(4):455–462, 1980.

[36] S Osborn, P S Vassilevski, and U Villa. A multilevel, hierarchical sampling technique for spatially correlated random fields. *SIAM J. Sci. Comput.*, 39(5):543–562, 2017.

[37] S Osborn, P Zulian, T Benson, U Villa, R Krause, and P S Vassilevski. Scalable hierarchical PDE sampler for generating spatially correlated random fields using non-matching meshes. *Preprint*, arXiv:1712.06758, 2017.

[38] S Pranesh and D Ghosh. Faster computation of the Karhunen-Loève expansion using its domain independence property. *Comput. Methods Appl. Mech. Engrg.*, 285:125–145, 2015.

[39] A Quarteroni, A Manzoni, and F Negri. *Reduced basis methods for partial differential equations*, volume 92 of *Unitext*. Springer, Cham, 2016.

[40] C P Robert. *The Bayesian Choice*. Springer, 2nd edition, 2007.

[41] C P Robert and G Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[42] L Roininen, M Girolami, S Lasanen, and M Markkanen. Hyperpriors for Matérn fields with applications in Bayesian inversion. *Preprint*, arXiv:1612.02989, 2016.

[43] L Roininen, S Lasanen, M Orispää, and S Särkkä. Sparse approximations of Fractional Matérn fields. *Scandinavian Journal of Statistics*, 45:194–216, 2018.

[44] A K Saibaba, J Lee, and P K Kitanidis. Randomized algorithms for generalized Hermitian eigenvalue problems with application to computing Karhunen-Loève expansion. *Numer. Linear Algebra Appl.*, 23(2):314–339, 2016.

[45] C Schwab and R A Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122, 2006.

[46] R C Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Society for Industrial and Applied Mathematics, 2014.

[47] I Sraj, O P Le Maître, O M Knio, and I Hoteit. Coordinate Transformation and Polynomial Chaos for the Bayesian Inference of a Gaussian Process with Parametrized Prior Covariance Function. *Computer Methods in Applied Mechanics and Engineering*, 298:205–228, 2016.

[48] A M Stuart. Inverse problems: A Bayesian perspective. In *Acta Numerica*, volume 19, pages 451–559. Cambridge University Press, 2010.

[49] P M Tagade and H-L Choi. A generalized polynomial chaos-based method for efficient Bayesian calibration of uncertain computational models. *Inverse Probl. Sci. Eng.*, 22(4):602–624, 2014.

[50] D N VanDerwerken and S C Schmidler. Parallel Markov Chain Monte Carlo. *Preprint*, arXiv:1312.7479, 2013.

[51] V N Vapnik and A Ya. Chervonenkis. On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.

[52] C K Wikle. Hierarchical Models for Uncertainty Quantification. In R Ghanem, D Higdon, and H Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 193–218. Springer, 2017.

[53] Z Zheng and H Dai. Simulation of multi-dimensional random fields by Karhunen-Loève expansion. *Comput. Methods Appl. Mech. Engrg.*, 324:221–247, 2017.