Bayesian nonparametric inference in PDE models: asymptotic theory and implementation

Matteo Giordano[1]

[1]University of Turin, Corso Unione Sovietica 218 bis, Turin, Italy

Abstract
Parameter identification problems in partial differential equations (PDEs) consist in determining one or more unknown functional parameters in a PDE. Here, the Bayesian nonparametric approach to such problems is considered. Focusing on the representative example of inferring the diffusivity function in an elliptic PDE from noisy observations of the PDE solution, the performance of Bayesian procedures based on Gaussian process priors is investigated. Recent asymptotic theoretical guarantees establishing posterior consistency and convergence rates are reviewed and expanded upon. An implementation of the associated posterior-based inference is provided, and illustrated via a numerical simulation study where two different discretisation strategies are devised. The reproducible code is available at: https://github.com/MattGiord.

Keywords. Inverse problems; Gaussian priors; frequentist consistency; posterior mean; Markov Chain Monte Carlo

1 Introduction

Partial differential equations (PDEs) are fundamental mathematical tools to model the behaviour of complex real-world phenomena, with ubiquitous applications across engineering and the sciences. The formulation of a PDE typically involves a number of *functional parameters*, which are often unknown in applications and not directly accessible to measurements. Employing a PDE model in practice therefore necessitates that the parameters in the equation be determined beforehand from the available data, giving rise to an *inverse problem* of *parameter identification*. Such problems have been extensively studied in applied mathematics [13, 26, 24] and, more recently, in statistics [25, 6, 23]. See the monographs [4, 3, 29] and references therein.

In the present paper, we focus on the following representative example. Many of the ideas presented hereafter have a natural extension to other classes of PDEs; see [29] for an in-depth treatment of the subject. Consider a physical quantity undergoing *diffusion* in an inhomogeneous multidimensional medium $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$. At equilibrium, the density $u(x)$ of the diffusing substance at any location $x \in \mathcal{O}$ is governed by the second-order elliptic PDE

$$\begin{aligned} \nabla \cdot (f \nabla u) &= s, \text{ on } \mathcal{O} \\ u &= b, \text{ on } \partial\mathcal{O}, \end{aligned} \tag{1}$$

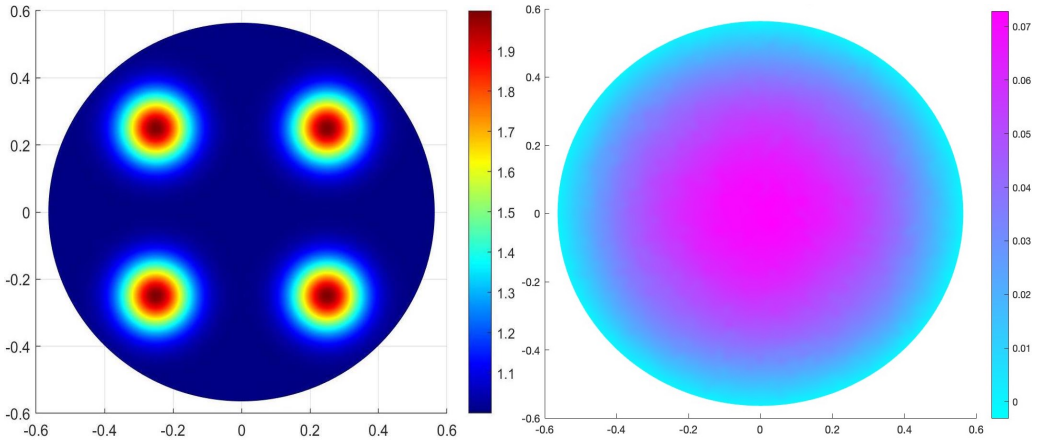

Figure 1: Left: an example of diffusivity function f with four circular regions of higher conductivity. Right: $n = 1000$ noisy observations from the corresponding PDE solution $G(f)$ at noise level $\sigma = .001$ (with constant source $s \equiv 1$ and homogeneous Dirichlet boundary conditions $b \equiv 0$).

where $s : \mathcal{O} \to \mathbb{R}$ describes the spatial distribution of local sources or sinks, $g : \partial\mathcal{O} \to \mathbb{R}$ prescribes the density values at the boundary, and the *diffusivity function* $f : \mathcal{O} \to (0, \infty)$ models spatially-varying conductivity throughout the inhomogeneous domain. Under mild regularity conditions on s, b, f and $\partial\mathcal{O}$, standard elliptic theory implies existence of a unique classical classical twice continuously-differentiable solution $G(f) \equiv u_f \in C^2(\mathcal{O})$ to (1) (e.g. [14, Chapter 6]). Assuming that s and b are known, we are then interested in the problem of estimating f from n noisy point evaluations of $G(f)$ over a grid of (possibly random) design points $X_1, \ldots, X_n$ in $\mathcal{O}$,

$$Y_i = G(f)(X_i) + \sigma W_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $W_1, \ldots, W_n$ are statistical measurement errors and $\sigma > 0$ is the noise level. In view of the central limit theorem, the Gaussian assumption $W_1, \ldots, W_n \overset{\text{iid}}{\sim} N(0, 1)$ can often be realistically maintained. The inverse problem of recovering the diffusivity in the elliptic PDE (1) from observations of its solution is an important building block in oil reservoir modelling [36], and has been studied in a large number of articles in applied mathematics, e.g. [31, 27, 7], and statistics, e.g. [6, 11, 9]. An illustration of the problem on the unit-area disk is given in Figure 1.

While the PDE (1) is linear, the parameter-to-solution map $f \mapsto G(f)$ is not, which poses several methodological and theoretical challenges. In particular, least squares functionals involving $G(f)$ (here corresponding to the negative log-likelihood, cf. eq. (4)) are then generally non-convex, so that commonly used optimisation-based methods (such as Tikhonov regularisation, maximum likelihood or maximum-a-posteriori estimation) cannot reliably be implemented by standard convex optimisation techniques. In this context, the Bayesian approach to inverse problems, popularised by influential work by Stuart [32, 12], offers an attractive alternative. In the Bayesian framework, the unknown parameter f is regarded as a random variable (with values in a function space) and assigned a *prior probability distribution* $\Pi(\cdot)$ that models the available information about f before collecting the observations. The prior is then combined, through *Bayes' formula*, with

the data $\{(Y_i, X_i)\}_{i=1}^n$ to form the *posterior distribution* $\Pi(\cdot|\{(Y_i, X_i)\}_{i=1}^n)$, which represents the updated belief about f and is used to draw the inferential conclusions. As the posterior formally involves only evaluations of the prior and the likelihood (cf. eq. (7)), approximate computation of $\Pi_n(\cdot|\{(Y_i, X_i)\}_{i=1}^n)$ and its associated *posterior mean estimator* $\bar{f}_n := E^{\Pi}[f|\{(Y_i, X_i)\}_{i=1}^n]$ via sampling methods is feasible as long as the forward map $G(\cdot)$ can be numerically evaluated. For the elliptic PDE (1), this can be done using efficient PDE solvers based on finite element methods, sidestepping altogether the need for a (possibly non-existent) inversion formula for $G(\cdot)$, as well as the use of optimisation approaches. In particular, for the class of *Gaussian process priors*, efficient ad-hoc Markov chain Monte Carlo (MCMC) algorithms, suited to the present infinite-dimensional setting, have been developed [9, 10, 5]. Finally, a further decisive advantage of the Bayesian methodology is that, alongside point estimates, it also automatically delivers *uncertainty quantification* for the recovery via the spread of the posterior, used in applications to provide interval-type estimators and to construct hypothesis tests.

The success and popularity in applications has led to recent interest in the literature for the derivation of theoretical performance guarantees for nonparametric Bayesian procedures in PDE models [34, 1, 19, 18, 28, 17, 20, 2]. Indeed, the performance of Bayesian methods depends on a suitable choice of the prior, which in infinite-dimensional statistical models primarily serves as a regularisation tool, and whose specification is a delicate task in its own merit (cf. Section 1.2 in [15]). Thus, the question arises as to whether Bayesian procedures may provide valid and prior-independent inference, at least in the presence of informative data. The established paradigm under which such investigation is carried out is the *frequentist analysis* of Bayesian procedures (cf. [15], and also Section 7.3 in [16]), assuming that the observations are generated by a fixed *ground truth* f_0 and studying the concentration of the posterior towards f_0 in the large sample size limit.

The present paper is concerned with the performance of Bayesian nonparametric methods based on Gaussian priors in the elliptic inverse problem (2). In Section , we review and expand upon recent results of Giordano and Nickl [19], which established posterior consistency and convergence rates for the conditional mean estimator for a large class of Gaussian process prior. They showed that such procedures provide statistically valid estimation of the diffusivity f, with explicit estimation error bounds that decay algebraically in the number n of observations. The class of Gaussian priors considered in [19] includes those associated to popular covariance kernels (such as the Matérn or squared-exponential kernels), as well as truncated Gaussian wavelet series expansions. In the present paper, we extend the general results of [19] to a formulation that allows to naturally deal with high-dimensional Gaussian sieve priors. We then use the general theorem to obtain statistical convergence rates for truncated Gaussian series priors defined on the eigenbasis of the Dirichlet-Laplacian, a commonly-used basis of practical interest that offers a convenient and generally applicable framework for implementation, cf. Section 4.2.

In Section 3.2, we complement the theoretical results with a discussion on implementation, devising two different discretisation strategies. In the first, we discretise the parameter space by a high-dimensional linear space of piece-wise linear functions on the elements of a deterministic triangular mesh. This approach is particularly suited to implementing procedures with Gaussian priors defined via a covariance kernel. In the numerical experiments of Section 4.1, we employ the popular Matérn kernel. The second proposed discretisation

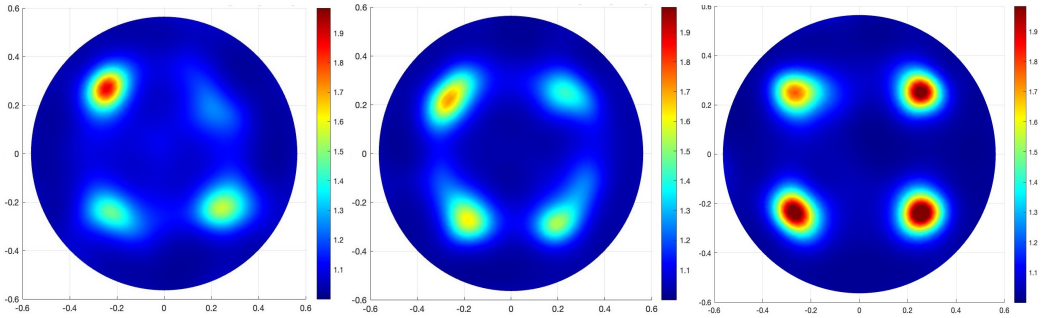

Figure 2: Right: The posterior mean estimate $\bar{f}_n$ of the diffusivity function f based on $n = 1000$ point evaluations of $G(f)$ at noise level $\sigma = .001$, computed via MCMC. The L^2-estimation error is $\|f_0 - \bar{f}_n\|_2 = .06394$. Left: estimate based on $n = 200$ observations (with estimation error .1430). Centre: estimate based on $n = 300$ observations (with estimation error .1348).

strategy is tailored to applications where a particular set of basis functions is of interest. In Section 4.2, we employ the eigenbasis of the Dirichlet-Laplacian as a broadly applicable choice, discretising the parameter space by high-dimensional truncated series expansions. A numerical simulation study is provided to illustrate the performance of the inferential procedures under the two discretisation strategies. In our numerical experiments, both approaches yielded satisfactory results, comparable in terms of reconstruction quality and running time. The posterior mean estimate (relative to a Matérn process prior), computed via a Metropolis-Hastings type MCMC algorithm, is shown in Figure 2 for increasing sample sizes, to be compared to the true diffusion coefficient pictured in Figure 1. The MATLAB code used for the study is available at: `https://github.com/MattGiord`.

2 Likelihood, prior and posterior

Throughout, $\mathcal{O} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is a given nonempty, open, convex and bounded set with smooth boundary $\partial\mathcal{O}$. For the observation model (2), with $G(f)$ the solution to the PDE (1), and for fixed constants $\alpha > 1 + d/2$ and $f_{\min} > 0$, we take as parameter space

$$\mathcal{F}_{\alpha, f_{\min}} := \left\{ f \in H^\alpha(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) \geq f_{\min},\, f_{|\partial\mathcal{O}} \equiv 1,\, \frac{\partial^j f}{\partial \nu^j} \equiv 0 \text{ for } 1 \leq j \leq \alpha - 1 \right\}, \tag{3}$$

with $H^\alpha(\mathcal{O})$ the usual Sobolev space (of regularity α) and $\nu(x)$, $x \in \partial\mathcal{O}$, the unit normal vector. Assuming that the source term s in (1) is fixed and smooth, and taking (without loss of generality) homogeneous Dirichlet boundary conditions $b \equiv 0$, the Schauder theory for elliptic PDEs (e.g. Theorem 6.14 in [14]) implies that for each $f \in \mathcal{F}_{\alpha,f_{\min}}$ there exists a unique classical solution $G(f) \in C(\overline{\mathcal{O}}) \cap C^{1+\alpha}(\mathcal{O})$ to the elliptic PDE (1). We then assume data $\{(Y_i, X_i)\}_{i=1}^n$ arising as in eq. (2) for some unknown $f \in \mathcal{F}_{\alpha,f_{\min}}$, with independent and identically distributed (i.i.d.) random design variables $X_1, \ldots, X_n$ following the uniform distribution on $\mathcal{O}$. Throughout, we regard the noise level $\sigma > 0$ in (2) as fixed and known; in practical applications it may often be replaced by an estimate (e.g. the sample variance). In view of the i.i.d. standard normal assumption on the noise variables $W_1, \ldots, W_n$ in (2), the random vectors $\{(Y_i, X_i)\}_{i=1}^n \sim P_f^{(n)}$ have joint probability density function in product

form,

$$p_f^{(n)}(\{(x_i,y_i)\}_{i=1}^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^n [y_i - G(f)(x_i)]^2/(2\sigma^2)}, \quad y_i \in \mathbb{R},\ x_i \in \mathcal{O}.$$

Accordingly, the log-likelihood is seen to be equal to, up to an additive constant, the negative least-square functional

$$l_n(f) := -\frac{1}{2\sigma^2}\sum_{i=1}^n [Y_i - G(f)(X_i)]^2, \qquad f \in \mathcal{F}_{\alpha, f_{\min}}. \tag{4}$$

In a recent paper by Giordano and Nickl [19], posterior consistency and convergence rates for the conditional mean estimator have been established for nonparametric Bayesian procedures based on a large class of Gaussian process priors. To incorporate the shape constraints in the parameter space $\mathcal{F}_{\alpha,f_{\min}}$, [19] employed the parametrisation

$$f \equiv \Phi \circ F, \tag{5}$$

where $F \in H_0^\alpha(\mathcal{O})$ (the completion of $C_c^\infty(\mathcal{O})$ with respect to $\|\cdot\|_{H^\alpha}$) and $\Phi : \mathbb{R} \to [f_{\min}, \infty)$ is a *regular link function*, that is a smooth, strictly increasing and bijective function with bounded derivatives and such that $\Phi(0) = 1$ (cf. Condition 1 in [19]). An example of regular link function is provided in Example 24 of [19]. In practice, the exponential link $\Phi(\cdot) = \exp(\cdot)$ is often used. In the following, we occasionally switch between the notation f and F, implicitly making use of the relation (5).

Under the parametrisation (5), placing a prior probability distribution $\Pi(\cdot)$ on F induces a (push-forward) prior on f, which, in slight abuse of notation, we still denote by $\Pi(\cdot)$. Following [19], we consider *scaled* Gaussian priors, constructed starting from a base (possibly n-dependent) centred Gaussian Borel probability measure Π_{W_n}, which we assume to be supported on a measurable linear subspace of the Hölder space $C^\beta(\mathcal{O})$, for some $\beta \geq 1$, and to have reproducing Kernel Hilbert space (RKHS) $\mathcal{H}_{W_n}$ continuously embedded into $H_0^\alpha(\mathcal{O})$. See Chapter 2 in [16] for background and terminology on Gaussian processes and measures. Given Π_{W_n}, we then construct the scaled prior $\Pi_n(\cdot)$ as the law of the random function

$$V(x) := \frac{W(x)}{n^{d/(4\alpha+4+2d)}}, \qquad x \in \mathcal{O}, \qquad W \sim \Pi_{W_n}. \tag{6}$$

By linearity, $\Pi_n(\cdot)$ also defines a centred Gaussian Borel probability measure with the same support as Π_{W_n}; however, as discusses in [19], the scaling enforces additional regularisation used in the theoretical analysis to deal with the non-linearity of the inverse problem.

Given prior $\Pi_n(\cdot)$ as above, the joint measurability of the map $(F,(y,x)) \mapsto p^{(1)}_{\phi\circ F}(y,x)$ implies that, by Bayes' formula (p.7, [15]), the posterior distribution $\Pi_n(\cdot|\{(Y_i,X_i)\}_{i=1}^n)$ of $F|\{(Y_i,X_i)\}_{i=1}^n$ arising from data in model (2) equals

$$\Pi_n(B|\{(Y_i,X_i)\}_{i=1}^n) = \frac{\int_B e^{l_n(\Phi\circ F)} d\Pi_n(F)}{\int_{C(\mathcal{O})} e^{l_n(\Phi\circ F')} d\Pi_n(F')}, \qquad B \subseteq C(\mathcal{O}) \text{ measurable}, \tag{7}$$

with $l_n(\cdot)$ the log-likelihood in (4).

3 Statistical convergence rates

3.1 A general posterior consistency result

We study the asymptotic concentration of the posterior distribution $\Pi_n(\cdot|\{(Y_i,\, X_i)\}_{i=1}^n)$ in (7) around the ground truth diffusivity function $f_0 = \Phi \circ F_0$, assuming that the data $\{(Y_i, X_i)\}_{i=1}^n \sim P_{f_0}^{(n)}$ have been generated according to the observation model (2) with $f = f_0$. The following theorem is a slightly extended reformulation of the main results in [19]. The expanded statement allows to conveniently include in the analysis general sieve-type Gaussian priors, cf. Example 3. The proof follows through largely unchanged from [19], in fact arguing similar to Section 3.2 in [19], and is therefore omitted for conciseness.

Theorem 1. *For fixed integer $\alpha > \beta + d/2$, with $\beta \geq 1$, consider the scaled prior Π_n in (6), where Π_{W_n} is a centred Gaussian Borel probability measure supported on a measurable linear subspace of $C^\beta(\mathcal{O})$, with RKHS $\mathcal{H}_{W_n} \subset H_0^\alpha(\mathcal{O})$ satisfying, for some constant $c > 0$ (independent of n),*

$$\|F\|_{H^\alpha} \leq c\|F\|_{\mathcal{H}_{W_n}}, \qquad \forall F \in \mathcal{H}_{W_n}.$$

Further assume that

$$\sup_{n\in\mathbb{N}} E^{\Pi_{W_n}} \|F\|_{C^1} < \infty.$$

For fixed $F_0 \in H_0^\alpha(\mathcal{O})$, suppose that there exists a sequence $F_{0,n} \in \mathcal{H}_{W_n}$ such that, as $n \to \infty$,

$$\|F_0 - F_{0,n}\|_{(H^1)^*} = O(n^{-\frac{\alpha+1}{2\alpha+2+d}}); \qquad \max\{\|F_{0,n}\|_{C^1}, \|F_{0,n}\|_{\mathcal{H}_{W_n}}\} = O(1).$$

Then, there exists $L > 0$ large enough such that, as $n \to \infty$,

$$\Pi_n\Big(f : \|G(f) - G(f_0)\|_{L^2} > Ln^{-\frac{\alpha+1}{2\alpha+2+d}} \Big| \{(Y_i, X_i)\}_{i=1}^n\Big) \to 0, \tag{8}$$

in $P_{f_0}^{(\infty)}$-probability as $n \to \infty$. If in addition $\beta \geq 2$ and $\inf_{x\in\mathcal{O}} s(x) > 0$, then there exists $L > 0$ large enough and a constant $\lambda > 0$ such that

$$\Pi_n(f : \|f - f_0\|_{L^2} > Ln^{-\lambda}|\{(Y_i, X_i)\}_{i=1}^n) \to 0, \tag{9}$$

in $P_{f_0}^{(\infty)}$-probability as $n \to \infty$, and moreover the estimator $\bar{f}_n = \Phi \circ \bar{F}_n$, with $\bar{F}_n = E^{\Pi_n}[F|\{(Y_i, X_i)\}_{i=1}^n]$, satisfies as $n \to \infty$,

$$P_{f_0}^{(n)}\big(\|\bar{f}_n - f_0\|_{L^2} > n^{-\lambda}\big) \to 0. \tag{10}$$

The first statement (eq. (8)) of Theorem 1 establishes posterior consistency in *prediction risk*: the induced posterior on the PDE solution $G(f), f \sim \Pi_n(\cdot|\{(Y_i,\, X_i)\}_{i=1}^n)$, concentrates around the true PDE solution $G(f_0)$ in L^2-distance at rate $n^{-(\alpha+1)/(2\alpha+2+d)}$. Since such rate is known to be minimax optimal [30, Theorem 10], the procedure is seen to optimally solve the PDE-constrained regression problem of recovering $G(f_0)$ in prediction risk from data $\{(Y_i, X_i)\}_{i=1}^n$.

The second statement shows that the posterior contracts around f_0 also in the standard L^2-risk, thereby solving the inverse problem of estimating the diffusivity. It follows combining

(8) with the regularisation properties implied by the rescaling in the prior construction (6) and a suitable stability estimate for G^{-1}. The latter was proved in [30, Lemma 24], and requires the slightly stronger assumption on β and the strict positivity of the source s. The exponent $\lambda > 0$ is explicitly computed in [19] and equals $\lambda = (\alpha+1)(\beta-1)/(2\alpha+2+d)(\beta+1)$. Note that $\lambda < (\alpha+1)/(2\alpha+2+d)$. While minimax optimal rates for estimating the diffusivity f in model (2) are currently unknown, inspection of the proofs in [19] shows that when $f_0 \in C^\infty(\mathcal{O})$, then the prior can be tuned so to attain a rate as closed as desired to the parametric rate $n^{-1/2}$.

The last statement of Theorem 1 entails that the 'posterior mean' estimator $\bar{f}_n$ converges towards f_0 in L^2-risk at the same rate $n^{-\lambda}$ attained by the whole posterior distribution. It is indeed a corollary of (9), following from uniform integrability arguments for Gaussian measures and the Lipschitzianity of the composition with the link function Φ, cf. Section 3 in [19].

3.2 Examples of Gaussian priors

We now provide two concrete instances of Gaussian priors to which Theorem 1 applies. For both examples, an implementation of the associated posterior-based inference is provided in Section 3.2 below. We will maintain the assumption that $f_0 = \Phi \circ F_0$ for some $F_0 \in H^\alpha(\mathcal{O})$ supported inside a given compact subset $K \subset \mathcal{O}$. This correspond to the common assumption that f_0 is known near the boundary $\partial\mathcal{O}$ (specifically $f_0 \equiv 1$ on $\mathcal{O}\backslash K$).

Example 2 (Example 25 in [19], Matérn covariance kernel)**.** *Consider a Matérn process $W = \{W(x),\ x \in \mathcal{O}\}$ on $\mathcal{O}$ with regularity $\alpha - d/2$, that is a centred stationary Gaussian process with covariance kernel*

$$C(x,y) = \frac{2^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{|x-y|\sqrt{2\alpha}}{\ell} \right)^\alpha B_\alpha \left(\frac{|x-y|\sqrt{2\alpha}}{\ell} \right), \qquad \ell > 0, \tag{11}$$

where Γ denotes the gamma function and B_α is the modified Bessel function of the second kind. Fix any smooth cut-off function $\chi \in C_c^\infty(\mathcal{O})$ such that $\chi = 1$ on K. It can then be shown (cf. Example 25 in [19]) that the law Π_W of χW defines a centred Gaussian Borel probability measure supported on the separable linear subspace $C^{\beta'}(\mathcal{O})$ of $C^\beta(\mathcal{O})$ for any $\beta < \beta' < \alpha - d/2$. Furthermore, its RKHS is given by $\mathcal{H}_W = \{\chi F,\ F \in H^\alpha(\mathcal{O})\} \subset H_0^\alpha(\mathcal{O})$, with RKHS norm satisfying

$$\|\chi F\|_{H^\alpha} \lesssim \|\chi F\|_{\mathcal{H}_W}, \qquad \forall F \in H^\alpha(\mathcal{O}).$$

For ground truths $F_0 \in H^\alpha(\mathcal{O})$ compactly supported inside K, we have $\chi F = F$, so that $F \in \mathcal{H}_W$. We conclude that Theorem 1 applies for a base Matérn process prior $\Pi_{W_n} := \Pi_W$, choosing the trivial approximating sequence $F_{0,n} := F_0$ for all $n \in \mathbb{N}$.

Next, we consider high-dimensional Gaussian sieve priors obtained via truncating Karhunen-Loève-type random series expansions, a frequently used approach in computation, e.g. [12, 21]. Section 2.2.2 in [19] studied truncated Gaussian wavelet series priors, and we here provide an extension, via Theorem 1, to expansions on the Dirichlet-Laplacian eigenbasis. Such constructions corresponds to commonly used Gaussian process priors with covariance kernel given by an inverse power of the Laplacian, e.g. [32, Section 2.4]. The eigenbasis

can be numerically computed via efficient finite element methods for elliptic eigenvalue problems, offering a broadly applicable framework for implementation on general domains $\mathcal{O}$. Details on computation and a numerical simulation study are provided in Section 4.2.

Example 3 (Dirichlet-Laplacian eigenbasis)**.** *Let* $\{e_j,\ j \geq 0\} \subset H_0^1(\mathcal{O}) \cap C^\infty(\overline{\mathcal{O}})$ *be the orthonormal basis of* $L^2(\mathcal{O})$ *formed by the eigenfunctions of the (negative) Dirichlet-Laplacian:*

$$\begin{aligned} -\Delta e_j - \lambda_j e_j &= 0, \text{ on } \mathcal{O} \\ e_j &= 0, \text{ on } \partial\mathcal{O}, \end{aligned} \qquad j \geq 0, \tag{12}$$

with associated eigenvalues $0 < \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots$, *satisfying* $\lambda_j \to \infty$ *as* $j \to \infty$ *according to Weyl's asymptotics:* $\lambda_j = O(j^{2/d})$ *as* $j \to \infty$. *See Example 6.3 and Section 7.4 in [22] for details. Define the Hilbert scale*

$$\mathbb{H}^s := \left\{ w \in L^2(\mathcal{O}) \ :\ \|w\|_{\mathbb{H}^s}^2 := \sum_{j=0}^{\infty} \lambda_j^s |\langle w, e_j\rangle_2|^2 < \infty \right\}, \qquad s \geq 0.$$

We then have $\mathbb{H}^0 = L^2(\mathcal{O})$ *(with equality of norms),* $\mathbb{H}^1 = H_0^1(\mathcal{O})$ *(with equivalence of norms),* $\mathbb{H}^2 = H_0^1(\mathcal{O}) \cap H^2(\mathcal{O})$, *and the continuous embedding* $\mathbb{H}^s \subset H^s(\mathcal{O})$ *for all* $s \geq 0$ *(holding generally with strict inclusion), cf. p.472f. in [33]. In fact, it holds that* $\|w\|_{\mathbb{H}^s} \simeq \|w\|_{H^s}$ *for all* $w \in \mathbb{H}^s$ *and* $s \geq 0$ *(proved initially for* $s = 2$, *then extended by induction to all larger integers, and finally by interpolation to all positive reals), and if* $F \in H^s(\mathcal{O})$ *is compactly supported within* $\mathcal{O}$, *then* $F \in \mathbb{H}^s$. *Finally, defining* $\mathbb{H}^{-1} := (\mathbb{H}^1)^* = (H_0^1(\mathcal{O}))^*$, *we have the equivalence (cf. eq. (A15) in [33])*

$$\|w\|_{(H_0^1)^*}^2 \simeq \|w\|_{\mathbb{H}^{-1}}^2 \equiv \sum_{j=0}^{\infty} \lambda_j^{-1} |\langle w, e_j\rangle_2|^2. \tag{13}$$

Now for fixed $J \in \mathbb{N}$, *the Gaussian random sum*

$$\overline{W}_J := \sum_{j \leq J} \lambda_j^{-\alpha/2} W_j e_j, \qquad W_j \overset{\text{iid}}{\sim} N(0,1),$$

defines a centred Gaussian Borel probability measure supported on (and with RKHS equal to) the finite-dimensional space $\mathcal{H}_{\overline{W}_J} := \text{span}\{e_j,\ j \leq J\}$, *with RKHS norm*

$$\|\bar{w}\|_{\mathcal{H}_{\overline{W}_J}}^2 = \sum_{j \leq J} \lambda_j^\alpha w_j^2 = \|\bar{w}\|_{\mathbb{H}^\alpha}^2 \simeq \|\bar{w}\|_{H^\alpha}^2, \qquad \forall w \in \overline{W}_J.$$

Fix any smooth cut-off function $\chi \in C_c^\infty(\mathcal{O})$ *such that* $\chi = 1$ *on* K, *and consider the random function*

$$W_n := \chi W_{J_n} = \chi \sum_{j \leq J_n} \lambda_j^{-\alpha/2} W_j e_j, \qquad W_j \overset{\text{iid}}{\sim} N(0,1), \tag{14}$$

where $J_n \in \mathbb{N}$ *is such that* $J_n \simeq n^{d/(2\alpha+2+d)}$. *By linearity and boundedness of multiplication by* χ, *the law* Π_{W_n} *of* W_n *defines, according to Exercises 2.6.5 in [16], a centred*

Gaussian prior supported on (and with RKHS equal to)

$$\mathcal{H}_{W_n} = \text{span}\{\chi e_j,\ j \le J_n\} \subset C_c^\infty(\mathcal{O}) \subset \bigcap_{s=0}^{\infty} H_0^s(\mathcal{O}) \cap \bigcap_{s=0}^{\infty} \mathbb{H}^s,$$

with RKHS norm satisfying, with multiplicative constant independent of n,

$$\|\chi\bar{w}\|_{\mathcal{H}_{W_n}} \le \|\bar{w}\|_{\mathcal{H}_{\overline{W}_{J_n}}} \simeq \|\bar{w}\|_{H^\alpha}, \qquad \forall \bar{w} \in \mathcal{H}_{\overline{W}_{J_n}}.$$

Arguing as in Example 25 in [19], one further shows that for some constant $c > 0$ (independent of n),

$$\|w\|_{H^\alpha} \le c\|w\|_{\mathcal{H}_{W_n}}, \qquad \forall w \in \mathcal{H}_{W_n}.$$

Finally, by a Sobolev embedding and the above inequality,

$$E^{\Pi_{W_n}}\|F\|_{C^1}^2 \lesssim E^{\Pi_{W_n}}\|F\|_{H^\alpha}^2 \le cE^{\Pi_{W_n}}\|F\|_{\mathcal{H}_{W_n}}^2 \le E\left[\sum_{j\le J_n} \lambda_j^{-\alpha} W_j^2\right],$$

which is uniformly bounded in n recalling that $W_i \overset{\text{iid}}{\sim} N(0,1)$ and the fact that by Weyl's asymptotics $\lambda_j^{-\alpha} = O(j^{-2\alpha/d})$ with $\alpha > 1 + d/2$. This shows that the sequence of base priors Π_{W_n} satisfies the first two assumptions of Theorem 1. For ground truths $F_0 \in H^\alpha(\mathcal{O})$ compactly supported inside K, we have $F_0 \in \mathbb{H}^\alpha$. Construct the finite-dimensional approximations

$$F_{0,n} = \sum_{j\le J_n} \langle F_0, e_j\rangle_2 \chi e_j \in \mathcal{H}_{W_n}, \qquad n \in \mathbb{N}. \tag{15}$$

Then for all $n \in \mathbb{N}$,

$$\begin{aligned}\|F_{0,n}\|_{\mathcal{H}_{W_n}} &\le \left\|\sum_{j\le J_n} \langle F_0, e_j\rangle_2 e_j\right\|_{\mathcal{H}_{\overline{W}_{J_n}}} \\ &= \left\|\sum_{j\le J_n} \langle F_0, e_j\rangle_2 e_j\right\|_{\mathbb{H}^\alpha} \le \|F_0\|_{\mathbb{H}^\alpha} \simeq \|F_0\|_{H^\alpha} < \infty.\end{aligned}$$

By a Sobolev embedding, we similarly have $\|F_{0,n}\|_{C^1} \le \|F_0\|_{H^\alpha} < \infty$ for all $n \in \mathbb{N}$. Furthermore, since both F_0 and $F_{0,n}$ have compact support within $\mathcal{O}$,

$$\begin{aligned}\|F_0 - F_{0,n}\|_{(H^1)^*} &= \sup_{H\in H^1(\mathcal{O})} \int_{\mathcal{O}} (F_0(x) - F_{0,n}(x))H(x)dx \\ &= \sup_{H\in H_0^1(\mathcal{O})} \int_{\mathcal{O}} (F_0(x) - F_{0,n}(x))H(x)dx \\ &= \|F_0 - F_{0,n}\|_{(H_0^1)^*},\end{aligned}$$

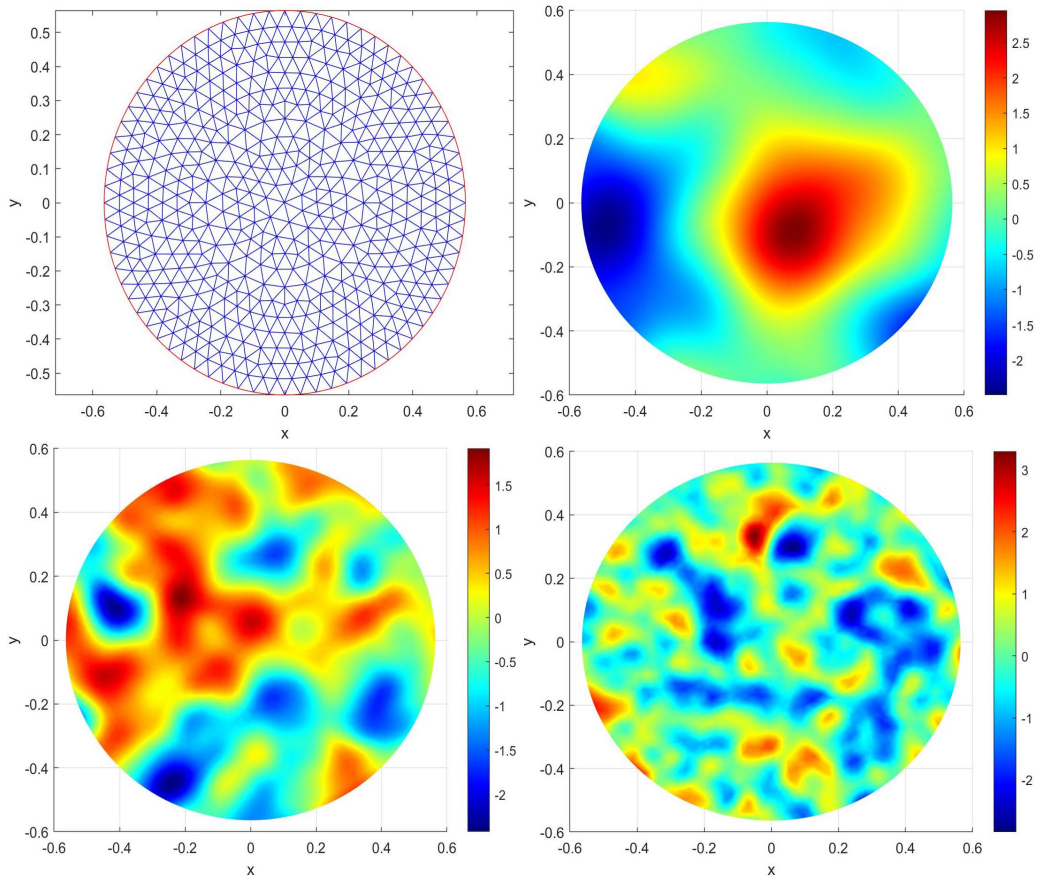

Figure 3: Left to right, top to bottom: deterministic triangular mesh with $M = 1981$ nodes; samples from a Matérn process prior with smoothness parameter $\alpha = 5$ and decreasing length parameter $\ell = .25, .1, .05$ respectively.

and recalling (13)*, Weyl's asymptotics, and the choice of* J_n*,*

$$\|F_0 - F_{0,n}\|^2_{(H^1)^*} = \sum_{j>J_n} \lambda_j^{-(1+\alpha)} \lambda_j^{\alpha} |\langle F_0, e_j \rangle_2|^2$$
$$\leq \lambda_{J_n}^{-(1+\alpha)} \|F_0\|^2_{\mathbb{H}^\alpha} \lesssim (J_n)^{-2(1+\alpha)/d} \simeq n^{-2(\alpha+1)/(2\alpha+2+d)}.$$

We conclude that Theorem 1 applies with the sequence of base Gaussian sieve priors in (14)*, choosing the approximations* $F_{0,n}$ *according to* (15)*.*

4 Implementation of the algorithms

4.1 Discretisation via piecewise linear functions

We take the unit-area disk $\mathcal{O} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1/\pi\}$ as the working domain, and assume we are given n noisy point evaluations of the PDE solution as in (2) over a grid of design points uniformly drawn at random on $\mathcal{O}$.

Following the parametrisation described in Section , we incorporate the positivity constraint on the diffusion coefficient f by modelling f as the composition $f = \Phi \circ F$, for $F : \mathcal{O} \to \mathbb{R}$

and where $\Phi : \mathbb{R} \to (0, \infty)$ is a fixed bijective smooth link function. Throughout this section, we use the link $\Phi(\cdot) = f_{\min} + \exp(\cdot)$, with $f_{\min} = 1$. We then discretise the parameter space by assuming that F is given by the finite sum

$$F = \sum_{m=1}^{M} F_m \psi_m, \qquad F_m \in \mathbb{R}, \qquad M \in \mathbb{N}, \tag{16}$$

where $\{\psi_1, \dots, \psi_M\}$ are piecewise linear functions on a deterministic triangular mesh with nodes $\{z_1, \dots, z_M\} \subset \mathcal{O}$ (displayed in Figure 3), uniquely characterised by the relation $\psi_m(z_{m'}) = 1_{\{m=m'\}}$. Accordingly, F in (16) satisfies $F(z_m) = F_m$, and for any $x \in \mathcal{O}$ the value $F(x)$ is obtained by linear interpolation over the pairs $\{(z_m, F_m),\ m = 1, \dots, M\}$.

We assign to F a centred Gaussian process prior $\Pi(\cdot)$, which can readily be implemented under the above discretisation scheme. In particular, if $C(x, y),\ x, y \in \mathcal{O}$, is the covariance kernel of $\Pi(\cdot)$, a sample $F \sim \Pi(\cdot)$ is drawn in practice by sampling the vector of coefficients $\mathbf{F} := [F_1, \dots, F_M]$ in (16) from the multivariate Gaussian distribution on $\mathbb{R}^M$,

$$\mathbf{F} \sim N(0, \mathbf{C}), \qquad \mathbf{C} = [\mathbf{C}_{mm'}]_{m,m'=1}^{M}, \qquad \mathbf{C}_{mm'} = C(z_m, z'_m).$$

Specifically, we consider the Matérn process prior, shown in Example 2 to yield asymptotically consistent posterior distributions according to Theorem 1. The Matérn covariance kernel is given in eq. (11); the hyperparameter $\alpha > 0$ controls the Sobolev regularity of the sample paths, while $\ell > 0$ determines the characteristic length-scale (cfr. Figure 3).

While the prior is Gaussian, the nonlinearity of the map $f \mapsto G(f)$ implies that the posterior distribution is generally non-Gaussian and not available in closed form. For Gaussian priors, specific MCMC algorithms have been developed to (approximately) sample from the posterior distribution, including the preconditioned Crank-Nicholson (pCN) method [9]. In the present setting, the pCN algorithm generates a Markov chain $(\vartheta_h)_{h \geq 1}$ with invariant measure equal to the posterior distribution of $F | \{(Y_i, X_i)\}_{i=1}^n$, starting from an initialisation point ϑ_0 and then, for $h \geq 0$, repeating the following steps:

1. Draw a prior sample $\xi \sim \Pi(\cdot)$ and for $\delta > 0$ define the proposal $p := \sqrt{1 - 2\delta}\vartheta_h + \sqrt{2\delta}\xi$;

2. Set
$$\vartheta_{h+1} := \begin{cases} p, & \text{with probability } 1 \wedge e^{l_n(\Phi \circ p) - l_n(\Phi \circ \vartheta_h)}, \\ \vartheta_h, & \text{otherwise,} \end{cases}$$
where l_n is the log-likelihood function in (4).

For each iteration, step 2. requires the evaluation of the log-likelihood $l_n(\Phi \circ p)$, which in turn entails the numerical evaluation of the PDE solution $G(\Phi \circ p)$ at the design points $\{X_1, \dots, X_N\}$, performed in practice using Matlab PDE Toolbox (which also provides an implementation of the triangular mesh for the discretisation of the parameter space). Prior samples $\xi \sim \Pi(\cdot)$ required in step 1. are drawn as described above.

The algorithm is terminated after H steps, returning approximate samples $\{\vartheta_h,\ h = 0, \dots, H\}$ from the posterior distribution of $F | \{(Y_i, X_i)\}_{i=1}^n$. Under certain assumptions on the forward map that are compatible with the present setting, Hairer, Stuart and Vollmer [21]

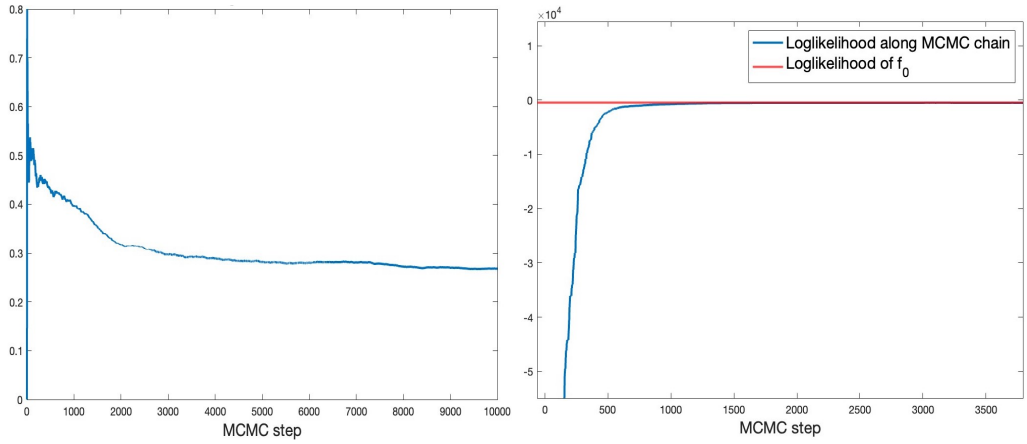

Figure 4: Left: acceptance rate over $H = 10000$ pCN samples. The rate stabilises around 27% after the initial burn-in time (first 2000 iterates). Right: in blue, the log-likelihood $l_n(\Phi \circ \vartheta_h)$ of the first 3500 iterates; in red, the log-likelihood $l_n(f_0)$ of the true diffusion coefficient f_0.

n	100	200	300	500	1000
$\|\bar{f}_n - f_0\|_2$	.1808	.1430	.1348	.08257	.06394

Table 1: L^2-estimation errors for increasing sample sizes.

derived dimension-free spectral gaps which imply rapid convergence towards the invariant measure. As a consequence, the posterior mean $\bar{F}_n = E^{\Pi}[F|\{(Y_i, X_i)\}_{i=1}^n]$ can be reliably computed numerically by the MCMC average

$$\bar{\vartheta} = \frac{1}{H+1} \sum_{h=0}^{H} \vartheta_h, \tag{17}$$

with non-asymptotic bounds for the numerical approximation error. Posterior credible sets can likewise be reliably computed by considering the empirical quantiles of the pCN samples.

Figure 1 (right) shows $n = 1000$ noisy discrete observations of the PDE solution corresponding to the true diffusion coefficient

$$\begin{aligned} f_0(x,y) = e^{-(10x-2.5)^2-(10y-2.5)^2} + e^{-(10x-2.5)^2-(10y-2.5)^2} \\ + e^{-(10x+2.5)^2-(10y+2.5)^2} + e^{-(10x+2.5)^2-(10y-2.5)^2} \end{aligned} \tag{18}$$

(shown above in Figure 1, left). The source function was taken to be constant $s(x, y) = 1$, $x, y \in \mathcal{O}$. The noise standard deviation was set to $\sigma = .001$. The posterior mean estimate $\bar{f}_n = \Phi \circ \bar{F}_n$, computed via the pCN algorithm (with $H = 10000$ iterations), is shown in Figure 2 (right), to be compared to the true diffusion coefficient pictured in Figure 1, left. The parameter space was discretised using a mesh with $M = 1981$ nodes, and the hyperparameter for the Matérn process prior were set to $\alpha = 10$ and $\ell = .125$. The step-size δ for the pCN algorithm was set to $\delta = .00125$, tuned so that after the initial burn-in time (here seen to roughly correspond to the first 2000 iterates), the acceptance rate of new

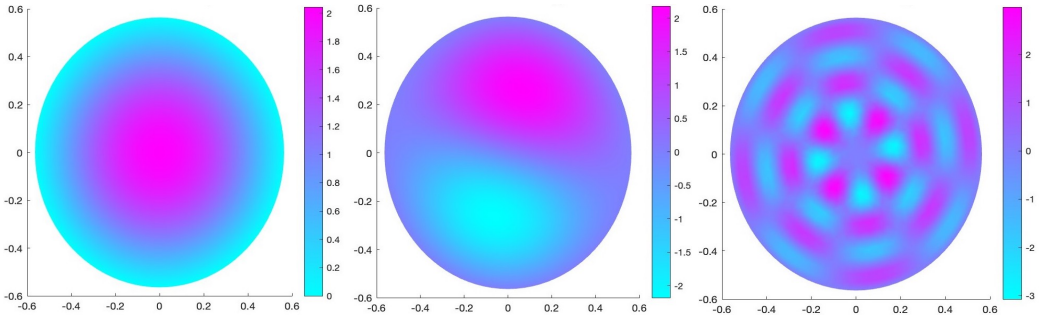

Figure 5: Left to right: the first two eigenfunctions $e_0,\ e_1$, and the eigenfunction e_J corresponding to the largest eigenvalue λ_J in the prescribed range $[0, \lambda_{\text{max}}]$.

proposals stabilises just below 30%; see Figure 4 (left). A non-informative initialisation point $\vartheta_0 \equiv 0$ for the pCN algorithm was chosen. The computation time was 1185 seconds (approximately 20 minutes) on a 2020 M1 MacBook Pro 13 running MATLAB 2023a. Figure 4 (right) shows the log-likelihood $l_n(\Phi \circ \vartheta_h)$ along the first 3500 pCN iterates, seen to rapidly increase towards, and then stabilise around, the log-likelihood $l_n(f_0)$ attained by the true diffusion coefficient f_0.

Figure 2 (left), shows the posterior mean estimate obtained with $n = 200$ observations, while Figure 2 (centre) shows the estimate with $n = 300$. Table 1 shows the decaying L^2-estimation error for increasing sample sizes $n = 100, 200, 300, 500, 1000$. For each of the experiments, the posterior mean was computed through 10000 pCN samples, discarding the first 1000 as burnin. All the pCN chains were started at the non-informative initialisation $\vartheta_0 \equiv 0$. The same discretisation of F based on a triangular mesh with $M = 1981$ nodes was used, and in each case the hyperparameters for the Matérn process prior were set to $\alpha = 10$ and $\ell = .125$. The only specific tuning regarded the pCN step size δ, set respectively to $\delta = .005, .0045, .0035, .002, .00125$, in order to have comparable final acceptance ratios (between 30% and 40%) after the initial burnin.

4.2 Discretisation via truncated series expansions

Taking again the unit-area disk $\mathcal{O}$ as the working domain, assuming n observations from (2) with design points drawn uniformly at random, and recalling the parametrisation $f = \Phi \circ F$ employed in Section 4.1, we discretise the parameter space by modelling F as the finite basis expansion

$$F = \sum_{j=1}^{J} F_j \psi_j, \qquad F_j \in \mathbb{R}, \qquad J \in \mathbb{N}, \tag{19}$$

where $\{e_j,\ j \geq 1\}$ is the collection of eigenfunctions of the Dirichlet-Laplacian on $\mathcal{O}$, with associated eigenvalues $\{\lambda_j,\ j \geq 1\}$; see Example 3 for details. For the special case of $\mathcal{O}$ being the unit-area disk, the eigenpairs are known explicitly [8], and are implemented in MATLAB within the Chebfun package [35]. On the other hand, we remark that the present discretisation approach is readily applicable to general domains $\mathcal{O}$, by computing numerically the Dirichlet-Laplacian eigenpairs via standard elliptic PDE solvers. For example, MATLAB PDE toolbox allows to specify a maximal range $[0, \lambda_{\text{max}}]$ for the eigenvalues, and then solves the eigenvalue problem (12) via finite element methods, returning (numerical

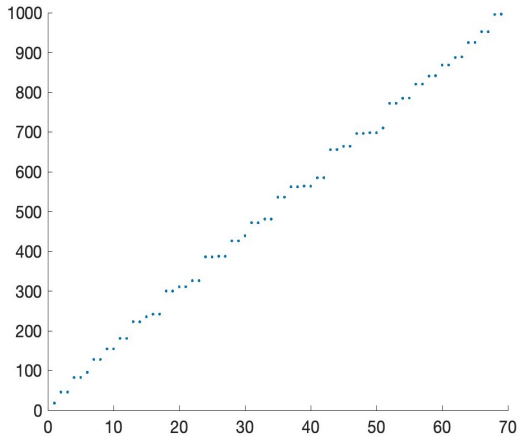

Figure 6: Dirichlet-Laplacian eigenvalues λ_j in the range $[0, \lambda_{\text{max}}]$.

approximations of) the eigenvalues in the prescribed range and the corresponding eigenfunctions. Figure 5 shows the first, second and J^{th} eigenfunction returned by the numerical solver. The range of search was set to $[0, \lambda_{\text{max}}] = [0, 1000]$, for which $J = 69$ eigenvalues were found. The numerical eigenvalues are plotted in Figure 6, showing a linear growth in accordance to Weyl's asymptotics.

Under the discretisation (19), the truncated Gaussian series prior $\Pi(\cdot)$ considered in Example 3 is implemented in practice sampling the vector of coefficients $\mathbf{F} := [F_1, \dots, F_J]$ in (19) from

$$\mathbf{F} \sim N(0, \mathbf{C}), \qquad \mathbf{C} := \text{diag}(\lambda_1^{-\alpha}, \dots, \lambda_J^{-\alpha}). \tag{20}$$

Posterior inference with such truncated Gaussian series prior $\Pi(\cdot)$ is then implemented via the pCN algorithm described Section 4.1: at each step, definition of the pCN proposal p in step 1. requires sampling $\xi \sim \Pi(\cdot)$, done in practice according to eq. (20). Computation of the acceptance probability in step 2. requires evaluation of the log-likelihood $l_n(\Phi \circ p)$, computed via (4) and the numerical evaluation of the PDE solution $G(\Phi \circ p)$ through a numerical elliptic PDE solver.

For the ground truth f_0 in eq. (18), Figure 7 shows the posterior mean estimate $\bar{f}_n = \Phi \circ \bar{F}_n$ based on $n = 1000$ observations from model (2) with $f = f_0$. The source function was set to $s \equiv 1$ and the noise level to $\sigma = .001$. The obtained L^2-estimation error was $\|\bar{f}_n - f_0\|_{L^2} = .0697$. For this experiment, the exponential link $\Phi(\cdot) = \exp(\cdot)$ was used. The parameter space was discretised using $J = 69$ basis functions. The prior regularity parameter in eq. (20) was set to $\alpha = .625$. The estimate $\bar{F}_n$ was numerically evaluated via an ergodic average as in eq. (17) over $H = 10000$ pCN iterates, discarding the first 1000 as burn-in. The step-size for the pCN algorithm was set to $\delta = .0001$, resulting in a stabilisation of the acceptance rate around 40% after the burnin. A non-informative pCN initialiser $\vartheta_0 \equiv 0$ was chosen. The computation time was 598.593 seconds (approximately 10 minutes) on a 2020 M1 MacBook Pro 13 running MATLAB 2023a.

Acknowledgement. This research has been partially supported by MUR, PRIN project 2022CLTYP4.

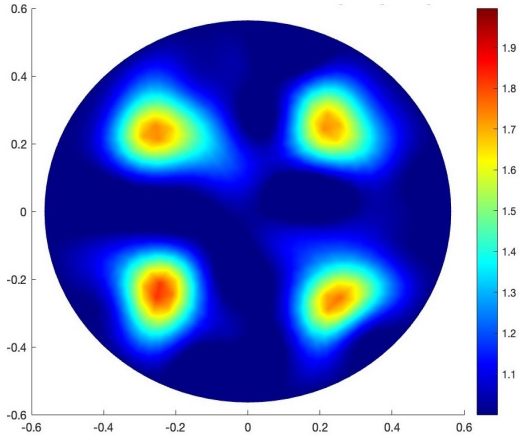

Figure 7: The posterior mean estimate $\bar{f}_n$ of the diffusivity function f based on $n = 1000$ observations of $G(f)$ at noise level $\sigma = .001$, computed via MCMC. The parameter space was discretised via the Dirichlet-Laplacian eigenbasis, cf. (19). The L^2-estimation error is $\|f_0 - \bar{f}_n\|_2 = .0697$.

References

[1] ABRAHAM, K., AND NICKL, R. On statistical Calderón problems. *Math. Stat. Learn. 2*, 2 (2019), 165–216.

[2] AGAPIOU, S., AND WANG, S. Laplace priors and spatial inhomogeneity in bayesian inverse problems. *Bernoulli* (to appear).

[3] ARRIDGE, S., MAASS, P., ÖKTEM, O., AND SCHÖNLIEB, C.-B. Solving inverse problems using data-driven models. *Acta Numer. 28* (2019), 1–174.

[4] BENNING, M., AND BURGER, M. Modern regularization methods for inverse problems. *Acta Numerica 27* (2018), 1–111.

[5] BESKOS, A., GIROLAMI, M., LAN, S., FARRELL, P., AND STUART, A. Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics 335* (2017), 327–351.

[6] BISSANTZ, N., HOHAGE, T., AND MUNK, A. Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems 20*, 6 (2004), 1773–1789.

[7] BONITO, A., COHEN, A., DEVORE, R., PETROVA, G., AND WELPER, G. Diffusion coefficients estimation for elliptic partial differential equations. *SIAM J. Math. Anal. 49*, 2 (2017), 1570–1592.

[8] CHURCHILL, R. V., AND BROWN, J. W. *Fourier series and boundary value problems*, third ed. McGraw-Hill Book Co., New York, 1978.

[9] COTTER, S., ROBERTS, G., STUART, A., AND WHITE, D. MCMC methods for functions: Modifying old algorithms to make them faster. *Statistical Science 28*, 3 (2013), 424–446.

[10] CUI, T., LAW, K. J. H., AND MARZOUK, Y. M. Dimension-independent likelihood-informed MCMC. *J. Comput. Phys. 304* (2016), 109–137.

[11] DASHTI, M., AND STUART, A. M. Uncertainty quantification and weak approximation of an elliptic inverse problem. *SIAM J. Numer. Anal. 49*, 6 (2011), 2524–2542.

[12] DASHTI, M., AND STUART, A. M. The Bayesian approach to inverse problems. In *Handbook of uncertainty quantification. Vol. 1, 2, 3*. Springer, Cham, 2017, pp. 311–428.

[13] ENGL, H. W., HANKE, M., AND NEUBAUER, A. *Regularization of inverse problems*, vol. 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[14] EVANS, L. C. *Partial differential equations*, second ed., vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010.

[15] GHOSAL, S., AND VAN DER VAART, A. W. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, New York, 2017.

[16] GINÉ, E., AND NICKL, R. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, New York, 2016.

[17] GIORDANO, M. Asymptotic theory for bayesian nonparametric inference in statistical models arising from partial differential equations. *Doctoral Thesis* (2021).

[18] GIORDANO, M., AND KEKKONEN, H. Bernstein–von Mises theorems and uncertainty quantification for linear inverse problems. *SIAM/ASA J. Uncertain. Quantif. 8*, 1 (2020), 342–373.

[19] GIORDANO, M., AND NICKL, R. Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inverse Problems 36*, 8 (2020), 085001–85036.

[20] GIORDANO, M., AND RAY, K. Nonparametric Bayesian inference for reversible multidimensional diffusions. *Ann. Statist. 50*, 5 (2022), 2872–2898.

[21] HAIRER, M., STUART, A., AND VOLLMER, S. Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. *The Annals of Applied Probability 24*, 6 (2014), 2455–2490.

[22] HAROSKE, D. D., AND TRIEBEL, H. *Distributions, Sobolev Spaces, Elliptic Equations*. EMS Press, 2007.

[23] HOHAGE, T., AND PRICOP, M. Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise. *Inverse Probl. Imaging 2*, 2 (2008), 271–290.

[24] ISAKOV, V. *Inverse problems for partial differential equations*, third ed., vol. 127 of *Applied Mathematical Sciences*. Springer, Cham, 2017.

[25] KAIPIO, J., AND SOMERSALO, E. *Statistical and Computational Inverse Problems*. No. 160 in Applied Mathematical Sciences. Springer-Verlag New York, 2004.

[26] KALTENBACHER, B., NEUBAUER, A., AND SCHERZER, O. *Iterative regularization methods for nonlinear ill-posed problems*, vol. 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

[27] KNOWLES, I. Parameter identification for elliptic problems. *Journal of Computational and Applied Mathematics 131*, 1 (2001), 175 – 194.

[28] MONARD, F., NICKL, R., AND PATERNAIN, G. P. Consistent inversion of noisy non-Abelian X-ray transforms. *Comm. Pure Appl. Math. 74*, 5 (2021), 1045–1099.

[29] NICKL, R. *Bayesian non-linear statistical inverse problems*. Zurich Lectures in Advanced Mathematics. EMS Press, Berlin, [2023] ©2023.

[30] NICKL, R., VAN DE GEER, S., AND WANG, S. Convergence rates for penalized least squares estimators in PDE constrained regression problems. *SIAM/ASA J. Uncertain. Quantif. 8*, 1 (2020), 374–413.

[31] RICHTER, G. R. An inverse problem for the steady state diffusion equation. *SIAM J. Appl. Math. 41*, 2 (1981), 210–221.

[32] STUART, A. M. Inverse problems: a Bayesian perspective. *Acta Numer. 19* (2010), 451–559.

[33] TAYLOR, M. E. *Partial Differential Equations I*. Springer New York, NY, 2011.

[34] VOLLMER, S. J. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems 29*, 12 (2013), 125011, 32.

[35] WILBER, H., TOWNSEND, A., AND WRIGHT, G. B. Computing with functions in spherical and polar geometries II. The disk. *SIAM J. Sci. Comput. 39*, 3 (2017), C238–C262.

[36] YEH, W. W.-G. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research 22*, 2 (1986), 95–108.