Inverse Problems and Imaging
Volume X, No. 0X, 200X, X–XX
Web site: http://www.aimSciences.org

BESOV PRIORS FOR BAYESIAN INVERSE PROBLEMS

Masoumeh Dashti
Mathematics Institute, University of Warwick
Coventry CV4 7AL, UK

Stephen Harris
School of Mathematics, University of Edinburgh
Edinburgh EH9 3JZ, UK

Andrew Stuart
Mathematics Institute, University of Warwick
Coventry CV4 7AL, UK

(Communicated by the associate editor name)

Abstract. We consider the inverse problem of estimating a function u from noisy, possibly nonlinear, observations. We adopt a Bayesian approach to the problem. This approach has a long history for inversion, dating back to 1970, and has, over the last decade, gained importance as a practical tool. However most of the existing theory has been developed for Gaussian prior measures. Recently Lassas, Saksman and Siltanen (Inv. Prob. Imag. 2009) showed how to construct Besov prior measures, based on wavelet expansions with random coefficients, and used these prior measures to study linear inverse problems. In this paper we build on this development of Besov priors to include the case of nonlinear measurements. In doing so a key technical tool, established here, is a Fernique-like theorem for Besov measures. This theorem enables us to identify appropriate conditions on the forward solution operator which, when matched to properties of the prior Besov measure, imply the well-definedness and well-posedness of the posterior measure. We then consider the application of these results to the inverse problem of finding the diffusion coefficient of an elliptic partial differential equation, given noisy measurements of its solution.

1. Introduction

The Bayesian approach to inverse problems is an attractive one. It mathematizes the way many practitioners incorporate new data into their understanding of a given phenomenon; and it results in a precise quantification of uncertainty. Although this approach to inverse problems has a long history, starting with the paper [16], it is only in the last decade that its use has become widespread as a computationl tool [19]. The theoretical side of the subject, which is the focus of this paper, is far from fully developed, with many interesting open questions. The work [16] concerned linear Gaussian problems, and the mathematical underpinnings of such problems were laid in the papers [25, 24]. An important theme in subsequent theoretical work concerning linear Gaussian problems has been to study the effect of discretization, in both the state space and the data space, and to identify approaches which give rise to meaningful limits [21, 22]. In many imaging problems, the detection of edges

2000 *Mathematics Subject Classification:* Primary: 60H30, 60G50, 60G15 Secondary: 35J99.
Key words and phrases: Inverse problems, Besov measures, Elliptic partial differential equations.

and interfaces is important and such problems are not well modelled by Gaussian priors. This has led to two recent works which try to circumvent this issue: the paper [23] introduces the notion of Besov priors, based on wavelet expansions, and the paper [17] uses heirarchical Gaussian models to create a discretization of the Mumford-Shah penalization, in one dimension. The thesis [27] studies a number of related issues for quite general measurement models. A different series of papers has studied the development of Bayesian inversion with Gaussian priors and non-linear measurement functions [7, 30], in which general criteria for a Bayes theorem, interpreted to mean that the posterior distribution has density with respect to the prior, are given. This framework has been used to study the effect of approximation of both the space in which the prior lies and the forward model [8, 11], allowing the transfer of error estimates for the approximation of the forward problem, into estimates for the approximation of the posterior measure. The goal of the present paper is to extend this type of approximation theory from Gaussian priors to the Besov priors introduced in [23], and in the case of nonlinear measurement functions.

We consider the noisy nonlinear operator equation

$$y = \mathcal{G}(u) + \eta \tag{1}$$

with $\mathcal{G} : X \to Y$, X, Y Banach spaces and η a Y-valued random variable. We suppose in this paper that y, the operator $\mathcal{G}$ and the statistical properties of η are known, and an estimation of $u \in X$ is to be found. Such an inverse problem appears in many practical situations where the function of interest (here u) cannot be observed directly and has to be obtained from other observable quantities and through the mathematical model relating them.

This problem is in general ill-posed and therefore to obtain a reasonable approximation of u in a stable way, we need prior information about the solution [15, 19]. In particular if we expect the unknown function to be sparse in some specific orthonormal basis of X, implementing the prior information in a way that respects this sparsity will result in more efficient finite-dimensional approximation of the solution. For instance a smooth function with a few local irregularities has a more sparse expansion in a wavelet basis compared, for example, to a Fourier basis, and adopting regularization methods which respect the sparse wavelet expansion of the unknown function is of interest in many applications. Approximation techniques based on wavelet bases for recovering finite dimensional estimates of the unknown function are extensively studied in the approximation theory, signal processsing and statistics literature; see for example [1, 6, 13, 14]. The paper [2] introduced a nonparametric Bayesian approach to the problem of signal recovery, adopting a clever posterior construction tuned to the simple form of the observation operator. The paper [23] considers more general linear observation operators and constructs the wavelet-based Besov prior; the construction is through a generalization of the Karhunen-Loèvé expansion to non-Gaussian coefficients and wavelet bases.

We adopt a Bayesian approach to the above inverse problem, so that regularization is implicit in the prior measure on X. We study the Besov prior measure introduced in [23] as this measure is constructed so that it factors as the product of independent measures along members of a wavelet basis, see Section 3. We first make sense of Bayes rule in an infinite dimensional setting, and then study the finite dimensional approximations after having proved the well-posedness of the posterior over an infinite-dimensional Banach space. Bayes rule for functions is here interpreted as follows. We put a prior probability measure $\mu_0(\mathrm{d}u)$ on u and then specify the distribution of $y|u$ via (1), thereby defining a joint probability distribution on (u, y). If the posterior measure $\mu^y(\mathrm{d}u) = \mathbb{P}(\mathrm{d}u|y)$ is absolutely continuous with respect to the prior measure $\mu_0(\mathrm{d}u) = \mathbb{P}(\mathrm{d}u)$ then Bayes theorem is interpreted as the following formula for the Radon-Nikodym derivative:

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) \propto \exp\big(-\Phi(u; y)\big). \tag{2}$$

Here Φ depends on the the specific instance of the data y, the forward (or observation) operator $\mathcal{G}$ and the distribution of η (see Section 3). For simplicity we work in the case where X comprises periodic functions on the $d-$dimensional torus $\mathbb{T}^d$; generalizations are possible.

The problem of making sense of Bayes rule for nonlinear observation operators, and with a Gaussian prior, is addressed in [7, 30]; sufficient conditions on Φ and the Gaussian prior μ_0 which imply the well-definedness and well-posedness of the posterior are obtained. Key to obtaining the conditions on Φ is the Fernique theorem for Gaussian measures which gives an upper bound on the growth of μ_0-integrable functions. Our aim here is to generalize the results of [7, 30] to the case of Besov prior measures and so we need a similar Fernique-like result for the Besov measures. In Section 2, following [23], we construct the Besov measures using wavelet expansions with i.i.d. random coefficients (Karhunen-Loèvé expansions) of their draws, and prove a Fernique-like result for these measures (Theorem 2.1). We then use this result in Section 3 to find the conditions on the operator $\mathcal{G}$ which ensure the well-definedness (Theorem 3.2) and well-posedness (Theorem 3.3) of the posterior measure μ^y, provided that the Besov measure μ_0 is chosen appropriately. We then build on this theory to quantify finite dimensional approximation of the posterior measure, culminating in Theorem 3.4.

In Section 4 we apply these results to the inverse problem of finding the diffusion coefficient of a linear uniformly elliptic partial differential equation in divergence form, in a bounded domain in dimension $d \leq 3$, from measurements of the solution in the interior. Such an inverse problem emerges in geophysical applications where u is the log-permeability of the subsurface. It is studied using a Bayesian approach on function space in [11] for Gaussian priors and in [28] for non-Gaussian priors which give rise to an almost sure uniform lower bound on the permeability. In many subsurface applications it is natural to expect that u is a smooth function with a few local irregularities, and it is not natural to expect an almost sure uniform lower bound on the permeability, nor is a Gaussian prior appropriate. For these reasons a wavelet basis, and hence a Besov prior, provides a plausible candidate from a modelling perspective. We show that the conditions we require for Φ in this problem hold naturally when X is the space of Hölder continuous functions C^t. This also suggests the use of wavelet bases which are unconditional bases for C^t. Thus for the elliptic inverse problem with Besov priors we state a well-posedness result for the posterior measure (Theorem 4.1) and we also quantify the effect of the finite dimensional approximation of the log-permiablility in the wavelet basis used in construction of the prior, on the posterior measure, with result summarized in Theorem 4.2.

2. Besov measures

We develop Besov measures following the construction in [23]. Let $\{\psi_l\}_{l=1}^{\infty}$ be a basis for $L^2(\mathbb{T}^d)$, $\mathbb{T}^d = (0,1]^d$, $d \le 3$, so that any $f \in L^2(\mathbb{T}^d)$ can be written as

$$f(x) = \sum_{l=1}^{\infty} f_l \, \psi_l(x).$$

Let $X^{s,q}$ be a Banach space with norm $\|\cdot\|_{X^{s,q}}$ defined as

$$\|f\|_{X^{s,q}} = \left(\sum_{l=1}^{\infty} l^{(sq/d+q/2-1)} |f_l|^q \right)^{1/q}$$

with $q \ge 1$ and $s > 0$. We now construct a probability measure on functions by randomizing the coefficients of an expansion in the basis $\{\psi_l\}_{l=1}^{\infty}$. The space $X^{s,q}$ will play a role analogous to the Cameron-Martin space for this measure. Indeed when $\{\psi_l\}_{l=1}^{\infty}$ is chosen to be the Karhunen-Loève basis for a Gaussian measure and $q = 2$, our choice of coefficients will ensure that $X^{s,2}$ is precisely the Cameron-Martin space.

Let $1 \le q < \infty$, $s > 0$, and $\kappa > 0$ be fixed and $\{\xi_l\}_{l=1}^{\infty}$ be real-valued i.i.d. random variables with probability density function

$$\pi_\xi(x) \propto \exp(-\frac{1}{2}|x|^q).$$

Let the random function u be defined as follows

$$u(x) = \sum_{l=1}^{\infty} l^{-(\frac{s}{d}+\frac{1}{2}-\frac{1}{q})} (\tfrac{1}{\kappa})^{\frac{1}{q}} \xi_l \psi_l(x). \tag{3}$$

We will refer to the induced measure on functions u as μ_0. We note that, since $\{\psi_l\}_{l=1}^{\infty}$ is an orthonormal basis and

$$u(x) = \sum_{l=1}^{\infty} u_l \psi_l(x)$$

with $u_l = l^{-(\frac{s}{d}+\frac{1}{2}-\frac{1}{q})} (\frac{1}{\kappa})^{\frac{1}{q}} \xi_l$, we have

$$\begin{aligned}
\prod_{l=1}^{\infty} \exp(-\frac{1}{2}|\xi_l|^q) &= \prod_{l=1}^{\infty} \exp\left(-\frac{\kappa}{2} l^{\frac{qs}{d}+\frac{q}{2}-1} |u_l|^q\right) \\
&= \exp\left(-\frac{\kappa}{2} \sum_{l=1}^{\infty} l^{\frac{qs}{d}+\frac{q}{2}-1} |u_l|^q\right) \\
&= \exp(-\frac{\kappa}{2} \|u\|_{X^{s,q}}^q) \qquad (4)
\end{aligned}$$

Thus, informally, u has a Lebesgue density proportional to $\exp(-\frac{\kappa}{2}\|u\|_{X^{s,q}}^q)$. We say that u is distributed according to an $X^{s,q}$ measure with parameter κ, or, briefly, a $(\kappa, X^{s,q})$ measure.

Remark 1. If $\{\psi_l\}_{l=1}^{\infty}$ in (3) is an r-regular wavelet basis [1] for $L^2(\mathbb{T}^d)$, with $r > s$, then $\|\cdot\|_{X^{s,q}}$ is the Besov B_{qq}^s [31] norm and u is distributed according to a (κ, B_{qq}^s)

[1] An r-regular wavelet basis for $L^2(\mathbb{R}^d)$ is a wavelet basis with r-regular scaling function and mother wavelets. A function f is r-regular if $f \in C^r$ and $|\partial^\alpha f(x)| \le C_m(1+|x|)^{-m}$, for any integer $m \in \mathbb{N}$ and any multi-index α with $|\alpha| = \alpha_1 + \cdots + \alpha_d \le r$. An r-regular basis for $L^2(\mathbb{T}^d)$ is obtained by periodification of the r-regular wavelets of $L^2(\mathbb{R}^d)$ [12, 26].

measure. Furthermore, if $q = 2$ with $\{\psi_l\}_{l=1}^{\infty}$ either a wavelet or Fourier basis, we obtain a Gaussian measure with Cameron-Martin space B^s_{22}, which is simply the Hilbert space $H^s = H^s(\mathbb{T}^d)$. Indeed (3) reduces to

$$u(x) = \sqrt{\frac{1}{\kappa}} \sum_{l=1}^{\infty} l^{-\frac{s}{d}} \, \xi_l \, \psi_l(x)$$

where $\{\xi_l\}_{l=1}^{\infty}$ are independent and identically distributed mean zero, unit variance Gaussian random variables. This is simply the Karhunen-Loève representation of draws from a mean zero Gaussian measure.

The following result shows that the random variable u is well-defined and characterizes apsects of its regularity.

Proposition 1. [23] *Let u be distributed according to a $(\kappa, X^{s,q})$ measure. The following are equivalent*

i) $\|u\|_{X^{t,q}} < \infty$ *almost surely.*
ii) $\mathbb{E}(\exp(\alpha\|u\|^q_{X^{t,q}})) < \infty$ *for any* $\alpha \in (0, \kappa/2)$;
iii) $t < s - d/q$.

Part (ii) of the above proposition provides a Fernique-like result [10] for Besov priors. Indeed for $q = 2$, this implies the Fernique theorem for all spaces H^t, $t < s - \frac{d}{q}$ on which the measure is supported. The Fernique result for Gaussian measures, however, is much stronger: it shows that if $u \in X$ almost surely with respect to the Gaussian measure μ then $\mathbb{E}^{\mu} \exp(\epsilon\|u\|^2_X) < \infty$ for ϵ small enough. Theorem 2.1 below goes some way towards showing a similar result for Besov measures by extending (ii) to C^t spaces (see Remark 2). Theorem 2.1 will also be useful in Section 3 when we deal with inverse problems, in the sense that it allows less restrictive conditions on the prior measure.

Before proving Theorem 2.1 we make a preliminary observation concerning the regularity of u given by (3) in C^t spaces. With the same conditions on s and t as are assumed in Proposition 1 one can show that $\mathbb{E}\|u\|_{C^t(\mathbb{T}^d)} < \infty$. Indeed, for any $\gamma \geq 1$, and u given by (3), using the definition of the Besov norm we can write

$$\|u\|^{\gamma q}_{B^t_{\gamma q, \gamma q}} = (\tfrac{1}{\kappa})^{\gamma} \sum_{l=1}^{\infty} l^{\frac{\gamma q t}{d} + \frac{\gamma q}{2} - 1} l^{-\gamma q(\frac{s}{d} + \frac{1}{2} - \frac{1}{q})} |\xi_l|^{\gamma q}.$$

Noting that $\mathbb{E}|\xi_l|^{\gamma q} = C(\gamma)$ and the exponent of l is smaller than -1 (since $t < s - d/p$), we have

$$\mathbb{E}\|u\|^{\gamma q}_{B^t_{\gamma q, \gamma q}} = C(\gamma)(\tfrac{1}{\kappa})^{\gamma} \sum_{l=1}^{\infty} l^{\frac{\gamma q}{d}(t-s) + \gamma - 1} \leq C_1(\gamma).$$

Now for a given $t < s - d/q$, choose γ large enough so that $\frac{d}{\gamma q} < s - d/q - t$. Then the embedding $B^{t_1}_{\gamma q, \gamma q} \subset C^t$ for any t_1 satisfying $t + \frac{d}{\gamma q} < t_1 < s - d/q$ [31] implies that $\mathbb{E}\|u\|_{C^t(\mathbb{T}^d)} < \infty$ and hence that $u \in C^t$ μ_0-almost surely. In the following theorem we show that, for small enough α, $\mathbb{E}\exp(\alpha\|u\|_{C^t(\mathbb{T}^d)}) < \infty$. For the proof we use the idea of the proof of a similar result for Radamacher series which appears in Kahane [20]. It is also key in this proof that the wavelet basis is an unconditional basis for Hölder spaces [32].

Theorem 2.1. *Let u be a random function defined as in (3) with $q \geq 1$ and $s > d/q$. Then for any $t < s - d/q$*

$$\mathbb{E}(\exp(\alpha||u||_{C^t})) < \infty$$

for all $\alpha \in (0, \kappa/(2\,r^*))$, with r^* a constant depending on q, d, s and t.

Proof. First, let $\kappa = 1$. We have [31]

$$\|u\|_{C^t} = \|u\|_{B^t_{\infty,\infty}} = \sup_{l\in\mathbb{N}} l^{(t-s)/d+1/q}|\xi_l| = \sup_{l\in\mathbb{N}} \lambda_l|\xi_l|$$

with $\lambda_l = l^{(t-s)/d+1/q}$. Note that, as shown above, $\|u\|_{C^t} < \infty$, μ_0-almost surely. Fix $r > 0$ and let

$$A = \{\omega \in \Omega : \sup_{l\in\mathbb{N}} \lambda_l|\xi_l(\omega)| > r\},$$
$$B = \{\omega \in \Omega : \sup_{l\in\mathbb{N}} \lambda_l|\xi_l(\omega)| > 2r\}.$$

Consider the following disjoint partition of A

$$A_m = \{\omega \in \Omega : \sup_{1\leq l\leq m-1} \lambda_l|\xi_l| \leq r,\ \lambda_m|\xi_m| > r\},$$

and define

$$B_m = \{\omega \in \Omega : \sup_{l\geq m} \lambda_l|\xi_l| > 2r\}.$$

We have

$$\mathbb{P}(A_m \cap B) = \mathbb{P}(A_m \cap B_m).$$

Noting that

$$B_m = C_m \cup B_{m+1} \quad \text{with} \quad C_m = \{\lambda_m|\xi_m| > 2r\},$$

and

$$A_m = D_m \cap E_m, \quad \text{with} \quad D_m = \{\omega \in \Omega : \sup_{1\leq l\leq m-1} \lambda_l|\xi_l| \leq r\},$$
$$E_m = \{\omega \in \Omega : \lambda_m|\xi_m| > r\},$$

we can write

$$\begin{aligned}\mathbb{P}(A_m \cap B) &\leq \mathbb{P}(A_m \cap C_m) + \mathbb{P}(A_m \cap B_{m+1})\\ &= \mathbb{P}(D_m \cap C_m) + \mathbb{P}(A_m)\mathbb{P}(B_{m+1})\\ &= \mathbb{P}(D_m)\mathbb{P}(C_m) + \mathbb{P}(A_m)\mathbb{P}(B_{m+1}).\end{aligned} \tag{6}$$

We now show that $\mathbb{P}(C_m) \leq (\mathbb{P}(E_m))^2$ for large enough r. First let $q > 1$. We have, with $\hat{r} = r/\lambda_m$, $c_0 = \int_{\mathbb{R}^d} \exp(-\frac{1}{2}|x|^q)\,\mathrm{d}x$ and $c_d = 2\pi^{d/2}/\Gamma(d/2)$,

$$\begin{aligned}\mathbb{P}(C_m) &= \frac{1}{c_0}\int_{\mathbb{R}^d} |x|\,\chi_{|x|>2\hat{r}}\,\mathrm{e}^{-\frac{1}{2}|x|^q}\mathrm{d}x\\ &= \frac{c_d}{c_0}\int_{2\hat{r}}^{\infty} \tilde{\rho}^d\,\mathrm{e}^{-\frac{1}{2}\tilde{\rho}^q}\mathrm{d}\tilde{\rho}\\ &= 2^{d+1}\frac{c_d}{c_0}\int_{\hat{r}}^{\infty} \rho^d\,\mathrm{e}^{-\frac{2^q}{2}\rho^q}\mathrm{d}\rho\\ &= 2^{d+1}\frac{c_d}{c_0}\int_{\hat{r}}^{\infty} \rho^d\,\mathrm{e}^{-\frac{1}{2}\rho^q}\,\mathrm{e}^{-\frac{2^q-1}{2}\rho^q}\mathrm{d}\rho\\ &\leq 2^{d+1}\frac{c_d}{c_0}\,\mathrm{e}^{-\frac{2^q-1}{2}\hat{r}^q}\int_{\hat{r}}^{\infty} \rho^d\,\mathrm{e}^{-\frac{1}{2}\rho^q}\,\mathrm{d}\rho\\ &= 2^{d+1}\,\mathrm{e}^{-\frac{2^q-1}{2}\hat{r}^q}\,\mathbb{P}(E_m).\end{aligned} \tag{7}$$

Then one can show that $2^{d+1}\mathrm{e}^{-\frac{2^q-1}{2}\hat{r}^q} \le \mathbb{P}(E_m)$ for sufficiently large $\hat{r}$. Indeed there exists $c_{q,d}$ depending only on q and d such that for $\rho \ge \hat{r} = c_{q,d}$ and $\nu \le 2^q - 2$ we have

$$\rho^d > 2^{d+1} c_0 \frac{\nu q + q}{2\, c_d}\, \rho^{q-1}\, \mathrm{e}^{-\nu\, \rho^q/2}. \tag{8}$$

Therefore we can write

$$\begin{aligned} \mathbb{P}(E_m) &= \frac{c_d}{c_0} \int_{\hat{r}}^{\infty} \rho^d\, \mathrm{e}^{-\frac{1}{2}\rho^q} \mathrm{d}\rho \\ &> 2^{d+1} \int_{\hat{r}}^{\infty} \frac{\nu q + q}{2}\, \rho^{q-1}\, \mathrm{e}^{-(\nu+1)\, \rho^q/2} \mathrm{d}\rho \\ &= 2^{d+1}\mathrm{e}^{-\frac{1+\nu}{2}\, \hat{r}^q} \;\ge\; 2^{d+1}\mathrm{e}^{-\frac{2^q-1}{2}\, \hat{r}^q} \end{aligned}$$

Hence going back to (7) we have

$$\mathbb{P}(C_m) \le (\mathbb{P}(E_m))^2. \tag{9}$$

For $q = 1$ we can calculate $\mathbb{P}(E_m)$ and $\mathbb{P}(C_m)$ as follows

$$\mathbb{P}(E_m) = \frac{c_d}{c_0} \sum_{k=0}^{d} 2^{k+1}\, r^{d-k} \binom{d}{k} \mathrm{e}^{-\frac{r}{2}}, \quad \text{and} \quad \mathbb{P}(C_m) = \frac{c_d}{c_0} \sum_{k=0}^{d} 2^{k+1}\, (2r)^{d-k} \binom{d}{k} \mathrm{e}^{-r}$$

where $c_d/c_0 = 2^d$ when $q = 1$. This readily shows (9) for the case of $q = 1$ as well. Substituting (9) in (6) we get

$$\begin{aligned} \mathbb{P}(A_m \cap B) &\le \mathbb{P}(D_m)(\mathbb{P}(E_m))^2 + \mathbb{P}(A_m)\mathbb{P}(B_{m+1}) \\ &= \mathbb{P}(D_m \cap E_m)\mathbb{P}(E_m) + \mathbb{P}(A_m)\mathbb{P}(B_{m+1}) \\ &= \mathbb{P}(A_m)\mathbb{P}(E_m) + \mathbb{P}(A_m)\mathbb{P}(B_{m+1}). \end{aligned}$$

This, since $E_m \subseteq A$ and $B_{m+1} \subseteq A$, implies that

$$\mathbb{P}(A_m \cap B) \le 2\mathbb{P}(A_m)\, \mathbb{P}(A).$$

Writting the above inequality for $m = 1, 2, \dots$ and then adding them we obtain

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) \le 2(\mathbb{P}(A))^2, \tag{10}$$

for $r > c_{q,d}$ (note that $\hat{r} = r/\lambda_m > c_q$ and $\lambda_m < 1$).

Let $\mathscr{P}(\rho) = \mathbb{P}(\|u\|_{C^t} > \rho)$. We have

$$\int_{C^t(\mathbb{T}^d)} \mathrm{e}^{\epsilon\|u\|_{C^t}}\, \mu_0(\mathrm{d}u) = \int_{\Omega} \mathrm{e}^{\epsilon\|u(\omega)\|_{C^t}}\, \mathbb{P}(\mathrm{d}\omega) = -\int_0^{\infty} \mathrm{e}^{\epsilon\rho}\, \mathrm{d}\mathscr{P}(\rho).$$

Let $\mathscr{P}(r) = \mathbb{P}(A) = \beta$ and note that by (10) we have

$$\mathscr{P}(r) = \frac{1}{2}(2\beta), \quad \mathscr{P}(2r) \le \frac{1}{2}(2\beta)^2, \;\dots, \mathscr{P}(2^n r) \le \frac{1}{2}(2\beta)^{2^n},$$

and therefore, choosing r large enough so that $\beta < \frac{1}{2}$,

$$-\int_{2^n r}^{2^{n+1} r} \mathrm{e}^{\epsilon\rho}\, \mathrm{d}\mathscr{P}(\rho) \le \mathrm{e}^{2^{n+1}\epsilon r}\, \mathscr{P}(2^n r) = \frac{1}{2}(2\beta)^{2^n}\, \mathrm{e}^{2^{n+1}\epsilon r}.$$

Hence, letting $\beta = 1/4$, we can write

$$\begin{aligned} - \int_0^\infty \mathrm{e}^{\epsilon\rho}\, \mathrm{d}\mathscr{P}(\rho) &\le - \int_0^r \mathrm{e}^{\epsilon\rho}\, \mathrm{d}\mathscr{P}(\rho) + \sum_{n=0}^\infty \frac{1}{2}(2)^{-2^n}\, \mathrm{e}^{2^{n+1}\epsilon r} \\ &= - \int_0^r \mathrm{e}^{\epsilon\rho}\, \mathrm{d}\mathscr{P}(\rho) + \sum_{n=0}^\infty \frac{1}{2}\mathrm{e}^{-2^n(\ln 2 - 2\epsilon r)} \end{aligned}$$

which is finite for $\epsilon < \frac{\ln 2}{2r}$. This proves the desired result for $\kappa = 1$, once we can identify the lower bound on r; a simple rescaling gives the result for general κ.

We need to choose r large enough so that $\mathbb{P}(A) = 1/4$ and (8) holds true as well. We have

$$\begin{aligned} \mathbb{P}(A) &\le \sum_{j=1}^\infty \mathbb{P}\{\omega : \lambda_j |\xi_j(\omega)| > r\} \le \frac{1}{c_{01}} \sum_{j=1}^\infty \int_{\frac{r}{\lambda_j}}^\infty x^d \mathrm{e}^{-x/2}\, \mathrm{d}x \\ &= \frac{1}{c_{01}} \sum_{j=1}^\infty \sum_{k=0}^d 2^{k+1} \left(\frac{r}{\lambda_j}\right)^{d-k} \binom{d}{k} \mathrm{e}^{-\frac{r}{2\lambda_j}} \\ &\le \frac{1}{c_{01}} \sum_{k=0}^d 2^{k+1} \binom{d}{k} \int_r^\infty x^{(d-k)(\frac{s-t}{d} - \frac{1}{q})} \exp(-\frac{1}{2} x^{\frac{s-t}{d} - \frac{1}{q}})\, \mathrm{d}x \end{aligned}$$

with $c_{01} = c_d^{-1} \int_{\mathbb{R}^d} \mathrm{e}^{-|x|/2}\, \mathrm{d}x$ and noting that $\lambda_1 = 1$ and $\lambda_j = j^{(t-s)/d + 1/q}$. Now choose $r_1 = r_1(s,t,d,q)$ such that

$$\int_{r_1}^\infty x^{(d-k)(\frac{s-t}{d} - \frac{1}{q})} \exp(-\frac{1}{2} x^{\frac{s-t}{d} - \frac{1}{q}})\, \mathrm{d}x < \frac{c_{01}}{4(d+1)} \frac{1}{2^{k+1}\binom{d}{k}}, \quad \text{for} \quad k = 0, \dots, d,$$

and therefore $\mathbb{P}(A) < 1/4$. To have (8) true as well, we set

$$\begin{aligned} \nu = 0, &\quad \text{and} \quad c_{q,d} = (2^d q\, c_0/c_d)^{1/(d-q+1)}, \quad \text{for} \quad 1 \le q < 1 + 3/4 \\ \nu = 1, &\quad \text{and} \quad c_{q,d} = \max\{1, 2^{d+2} q\, c_0/c_d\}, \quad \text{for} \quad q \ge 1 + 3/4. \end{aligned}$$

Since for $\kappa \ne 1$, $\epsilon = \alpha/\kappa$, this implies that $\alpha \le \kappa/(2\, r^*)$ with

$$r^* = (\ln 2)\, \max\{r_1, c_{q,d}\}. \tag{11}$$

□

Remark 2. Note that the bound on α in Theorem 2.1 is not sharp. Also, under the same condition as in Theorem 2.1, it is natural to expect a similar result to hold with power q of $\|u\|_{C^t}$ in the exponent and for the result to extend to a norm in any space which has full measure; it would then be consistent with the Gaussian Fernique theorem that arises when $q = 2$ (see Theorem 2.6 in [10]). However we have not been able to prove the result with this level of generality.

3. Bayesian approach to inverse problems for functions

Recall the probabilistic inverse problem we introduced in Section 1: find the posterior distribution μ^y of $u \in X$, given a prior distribution μ_0 of $u \in X$, and $y \in Y$ given by (1) for a single realization of the random variable η. We denote the distribution of η on Y by $\mathbb{Q}_0(\mathrm{d}y)$. By (1) the distribution of y given u is known as well; we denote it by $\mathbb{Q}^u(\mathrm{d}y)$ and, provided $\mathbb{Q}^u$ is absolutely continuous with respect to $\mathbb{Q}_0$, we may define $\Phi : X \times Y \to \mathbb{R}$ so that

$$\frac{\mathrm{d}\mathbb{Q}^u}{\mathrm{d}\mathbb{Q}_0}(y) = \exp\big(-\Phi(u;y)\big), \tag{12}$$

and

$$\int_Y \exp\big(-\Phi(u;y)\big)\,\mathbb{Q}_0(\mathrm{d}y) = 1. \tag{13}$$

For instance if η is a mean zero random Gaussian field on Y with Cameron-Martin space $\big(E, \langle\cdot,\cdot\rangle_E, \|\cdot\|_E\big)$ then the Cameron-Martin formula (Proposition 2.24 in [10]) gives

$$\Phi(u;y) = \frac{1}{2}\|\mathcal{G}(u)\|_E^2 - \langle y, \mathcal{G}(u)\rangle_E. \tag{14}$$

For finite-dimensional $Y = \mathbb{R}^K$, when η has Lebesgue density ρ, then we have the identity $\exp(-\Phi(u;y)) = \rho(y - \mathcal{G}(u))/\rho(y)$.

The previous section shows how to use wavelet or Fourier bases to construct probability measures μ_0 which are supported on a given Besov space B^t_{qq} (and consequently on the Hölder space C^t). Here we show how use of such priors μ_0 may be combined with properties of Φ, defined above, to deduce the existence of a well-posed Bayesian inverse problem. To this end we assume the following conditions on Φ:

Assumption 3.1. Let X and Y be Banach spaces. The function $\Phi : X \times Y \to \mathbb{R}$ satisfies:

(i) there is an $\alpha_1 > 0$ and for every $r > 0$, an $M \in \mathbb{R}$, such that for all $u \in X$, and for all $y \in Y$ such that $||y||_Y < r$,
$$\Phi(u, y) \geq M - \alpha_1 ||u||_X;$$

(ii) for every $r > 0$ there exists $K = K(r) > 0$ such that for all $u \in X$, $y \in Y$ with $\max\{||u||_X, ||y||_Y\} < r$
$$\Phi(u, y) \leq K;$$

(iii) for every $r > 0$ there exists $L = L(r) > 0$ such that for all $u_1, u_2 \in X$ and $y \in Y$ with $\max\{||u_1||_X, ||u_2||_X, ||y||_Y\} < r$
$$|\Phi(u_1, y) - \Phi(u_2, y)| \leq L||u_1 - u_2||_X;$$

(iv) there is an $\alpha_2 > 0$ and for every $r > 0$ a $C \in \mathbb{R}$ such that for all $y_1, y_2 \in Y$ with $\max\{||y_1||_Y, ||y_2||_Y\} < r$ and for every $u \in X$
$$|\Phi(u, y_1) - \Phi(u, y_2)| \leq \exp(\alpha_2 ||u||_X + C)||y_1 - y_2||.$$

3.1. Well-defined and well-posed Bayesian inverse problems.

Recall the notation μ_0 for the Besov prior measure defined by(3) and μ^y for the resulting posterior measure. We now prove well-definedness and well-posedness of the posterior measure. The following theorems generalize the results of [30] from the case of Gaussian priors to Besov priors.

Theorem 3.2. *Let Φ satisfy (13) and Assumption 3.1(i)–(iii). Suppose that for some $t < \infty$, C^t is continuously embedded in X. There exists $\kappa^* > 0$ such that if μ_0*

is a $(\kappa, X^{s,q})$ *measure with* $s > t + \frac{d}{q}$ *and* $\kappa > \kappa^*$*, then* μ^y *is absolutely continuous with respect to* μ_0 *and satisfies*

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)} \exp\big(-\Phi(u;y)\big), \tag{15}$$

with the normalizing factor $Z(y) = \int_X \exp(-\Phi(u;y))\, \mu_0(\mathrm{d}u) < \infty$*. The constant* $\kappa^* = 2\, c_e\, r^*\, \alpha_1$*, where* c_e *is the embedding constant satisfying* $\|u\|_X \le c_e \|u\|_{C^t}$*, and* r^* *is as in (11).*

Proof. Define $\pi_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u) \otimes \mathbb{Q}_0(\mathrm{d}y)$ and $\pi(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mathbb{Q}^u(\mathrm{d}y)$. Assumption 3.1(iii) gives continuity of Φ on X and since $\mu_0(X) = 1$ we have that $\Phi : X \to \mathbb{R}$ is μ_0-measurable. Therefore $\pi \ll \pi_0$ and π has Radon-Nikodym derivative given by (12) as (noting that by (13) and since $\mu_0(X) = 1$, we have $\int_{X\times Y} \exp\big(-\Phi(u;y)\big)\, \pi_0(\mathrm{d}u, \mathrm{d}y) = 1$)

$$\frac{\mathrm{d}\pi}{\mathrm{d}\pi_0}(u;y) = \exp\big(-\Phi(u;y)\big).$$

This then by Lemma 5.3 of [18], implies that $\mu^y(du) = \pi(\mathrm{d}u, \mathrm{d}\xi | \xi = y)$ is absolutely continuous with respect to $\mu_0(y) = \pi_0(\mathrm{d}u, \mathrm{d}\xi | \xi = y)$, since π_0 is an independent product. This same lemma also gives (15) provided that the normalization constant is positive, which we now establish. We note that all integrals over X may be replaced by integrals over $X^{t,q}$ for any $t < s - \frac{d}{q}$ since $\mu_0(X^{t,q}) = 1$. First by Assumption 3.1(i) note that there is $M = M(y)$ such that

$$\begin{aligned} Z(y) &= \int_{X^{t,q}} \exp(-\Phi(u;y)) d\mu_0(u) \\ &\le \int_{X^{t,q}} \exp(\alpha_1 \|u\|_X - M) d\mu_0(u) \\ &\le \int_{X^{t,q}} \exp(\alpha_1\, c_e \|u\|_{C^t} - M) d\mu_0(u) \end{aligned}$$

This upper bound is finite by Theorem 2.1 since $\kappa > 2\, c_e\, r^*\, \alpha_1$. We now prove that the normalisation constant does not vanish. Let $R = \mathbb{E}\|u\|_{X^{t,q}}$ noting that $R \in (0, \infty)$ since $t < s - \frac{d}{q}$. As $\|u\|_{X^{t,q}}$ is a nonnegative random variable we have that $\mu_0(\|u\|_{X^{t,q}} < R) > 0$. Taking $r = \max\{\|y\|_Y, R\}$, Assumption 3.1(ii) gives

$$\begin{aligned} Z(y) &= \int_{X^{t,q}} \exp(-\Phi(u;y)) d\mu_0(u) \\ &\ge \int_{\|u\|_{X^{t,q}} < R} \exp(-K) d\mu_0(u) \\ &= \exp(-K)\mu_0(\|u\|_{X^{t,q}} < R) \end{aligned}$$

which is positive. □

We now show the well-posedness of the posterior measure μ^y with respect to the data y. Recall that the Hellinger metric d_{Hell} is defined by

$$d_{\mathrm{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2} \int \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\nu}} \right)^2 d\nu}.$$

The Hellinger metric is independent of the choice of reference measure ν, the measure with respect to which both μ and μ' are absolutely continuous. The posterior measure is Lipschitz with respect to data y, in this metric.

Theorem 3.3. *Let Φ satisfy (13) and Assumption 3.1(i)–(iv). Suppose that for some $t < \infty$, C^t is continuously embedded in X. There exists $\kappa^* > 0$ such that if μ_0 is a $(\kappa, X^{s,q})$ measure with $s > t + \frac{d}{q}$ and $\kappa > \kappa^*$ then*

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \le C\, \|y - y'\|_Y$$

where $C = C(r)$ with $\max\{\|y\|_Y, \|y'\|_Y\} \le r$. The constant $\kappa^* = 2\, c_e\, r^*(\alpha_1 + 2\alpha_2)$, where c_e is the embedding constant satisfying $\|u\|_X \le c_e \|u\|_{C^t}$, and r^* is as in (11).

Proof. As in Theorem 3.2, $Z(y), Z(y') \in (0, \infty)$. An application of the mean value theorem along with Assumption 3.1(i), and (iv) gives

$$\begin{aligned}
|Z(y) - Z(y')| &\le \int_{X^{t,q}} |\exp(-\Phi(u;y)) - \exp(-\Phi(u;y'))|\, d\mu_0(u) \\
&\le \int_{X^{t,q}} \exp(\alpha_1 \|u\|_X - M) |\Phi(u;y) - \Phi(u;y')|\; d\mu_0(u) \\
&\le \int_{X^{t,q}} \exp\big((\alpha_1 + \alpha_2)\|u\|_X - M + C\big) \|y - y'\|_Y\, d\mu_0(u) \\
&\le C\|y - y'\|_Y, \qquad (16)
\end{aligned}$$

since $\|u\|_X \le c_e \|u\|_{C^t}$ and $c_e(\alpha_1 + \alpha_2) < \kappa/(2r^*)$. Using the inequality $(a+b)^2 \le 2(a^2 + b^2)$

$$\begin{aligned}
2d_{\text{Hell}} &= \int_{X^{t,q}} \left(Z(y)^{-\frac{1}{2}} \exp(-\frac{1}{2}\Phi(u;y) - (Z(y')^{-\frac{1}{2}} \exp(-\frac{1}{2}\Phi(u;y') \right)^2 d\mu_0(u) \\
&\le I_1 + I_2
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \frac{2}{Z(y)} \int_{X^{t,q}} \left(\exp(-\frac{1}{2}\Phi(u;y)) - \exp(-\frac{1}{2}\Phi(u;y')) \right)^2 d\mu_0(u) \\
I_2 &= 2|Z(y)^{-\frac{1}{2}} - Z(y')^{-\frac{1}{2}}|^2 \int_{X^{t,q}} \exp(-\Phi(u;y')) d\mu_0(u) \\
&= 2|Z(y)^{-\frac{1}{2}} - Z(y')^{-\frac{1}{2}}|^2 Z(y').
\end{aligned}$$

Again, an application of the mean value theorem, and use of Assumptions 3.1(i) and (iv), gives

$$\begin{aligned}
\frac{Z(y)}{2} I_1 &\le \int_{X^{t,q}} \frac{1}{4} \exp(\alpha_1 \|u\|_X - M) \exp(2\alpha_2 \|u\|_X + 2C) \|y - y'\|^2\, \mathrm{d}\mu_0(u) \\
&\le C\|y - y'\|_Y^2,
\end{aligned}$$

since $c_e(\alpha_1 + 2\alpha_2) < \kappa/(2r^*)$. Recall that $Z(y)$ and $Z(y')$ are positive and bounded from above. Thus by the mean value theorem and (16)

$$I_2 = 2Z(y')|Z(y)^{-\frac{1}{2}} - Z(y')^{-\frac{1}{2}}|^2 \le C|Z(y) - Z(y')|^2 \le C\|y - y'\|_Y^2.$$

The result follows. □

3.2. Approximation of the posterior. Consider Φ^N to be an approximation of Φ. Here we state a result which quantifies the effect of this approximation in the posterior measure in terms of the aproximation error in Φ.

Define $\mu^{y,N}$ by

$$\frac{\mathrm{d}\mu^{y,N}}{\mathrm{d}\mu_0}(u) = \frac{1}{Z^N(y)} \exp\big(-\Phi^N(u)\big), \qquad (17a)$$

$$Z^N(y) = \int_X \exp\big(-\Phi^N(u)\big)\mathrm{d}\mu_0(u). \tag{17b}$$

We suppress the dependence of Φ and Φ^N on y in this section as it is considered fixed.

Theorem 3.4. *Assume that the measures μ and μ^N are both absolutely continuous with respect to μ_0, and given by (15) and (17) respectively. Suppose that Φ and Φ^N satisfy Assumption 3.1(i) and (ii), uniformly in N, and that there exist $\alpha_3 \geq 0$ and $C \in \mathbb{R}$ such that*

$$|\Phi(u) - \Phi^N(u)| \leq \exp(\alpha_3\|u\|_X + C)\psi(N)$$

where $\psi(N) \to 0$ as $N \to \infty$. Suppose that for some $t < \infty$, C^t is continuously embedded in X. Let μ_0 be a $(\kappa, X^{s,q})$ measure with $s > t + \frac{d}{q}$ and $\kappa > 2\,c_e\,r^*(\alpha_1 + 2\alpha_3)$ where r^* is as in (11) and c_e is the embedding constant satisfying $\|u\|_X \leq c_e\|u\|_{C^t}$. Then there exists a constant independent of N such that

$$d_{\mathrm{Hell}}(\mu, \mu^N) \leq C\psi(N).$$

The proof is very similar to the proof of Theorem 3.3 and, in the Gaussian case, is given in [30]; hence we omit it.

4. Application to an elliptic inverse problem

We consider the elliptic equation

$$-\nabla\cdot\big(\mathrm{e}^{u(x)}\nabla p(x)\big) = f + \nabla\cdot g, \quad x \in \mathbb{T}^d, \tag{18}$$

with periodic boundary conditions and with $\mathbb{T}^d = (0,1]^d$, $d \leq 3$, p, u and f scalar functions and g a vector function on $\mathbb{T}^d$. Given any $u \in L^\infty(\mathbb{T}^d)$ we define $\lambda(u)$ and $\Lambda(u)$ by

$$\lambda(u) = \operatorname{ess}\inf_{x\in\mathbb{T}^d} \mathrm{e}^{u(x)}, \qquad \Lambda(u) = \operatorname{ess}\sup_{x\in\mathbb{T}^d} \mathrm{e}^{u(x)}.$$

Where it causes no confusion we will simply write λ or Λ. Equation (18) arises as a model for flow in a porous medium with p the pressure (or the head) and e^u the permeability (or the transmissivity); the velocity v is given by the formula $v \propto -\mathrm{e}^u\nabla p$.

Consider making noisy pointwise observations of the pressure field p. We write the observations as

$$y_j = p(x_j) + \eta_j, \quad x_j \in \mathbb{T}^d \quad j = 1, \cdots, K. \tag{19}$$

We assume, for simplicity, that $\eta = \{\eta_j\}_{j=1}^K$ is a mean zero Gaussian with covariance Γ. Our objective is to determine u from $y = \{y_j\}_{j=1}^K \in \mathbb{R}^K$. Concatenating the data, we have

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

with

$$\mathcal{G}(u) = \big(p(x_1), \cdots, p(x_{\mathrm{K}})\big)^T. \tag{20}$$

In order to apply Theorem 3.2, 3.3 and 3.4 to the elliptic inverse problem we need to prove certain properties of the forward operator $\mathcal{G}$ given by (20), viewed as a mapping from a Banach space X into $\mathbb{R}^m$. The space X must be chosen so that C^t is continuously embedded into X and then the Besov prior μ_0 chosen with $s > t + \frac{d}{q}$. In this section $|\cdot|$ stands for the Euclidean norm. The following result is proved in [11].

Proposition 2. *Let $f \in L^r(\mathbb{T}^d)$, $g \in L^{2r}(\mathbb{T}^d)$. Then for any $u \in L^\infty(\mathbb{T}^d)$ there exists $C = C(K, d, r, \|f\|_{L^r}, \|g\|_{L^{2r}})$ such that*

$$|\mathcal{G}(u)| \leq C \, \exp(\|u\|_{L^\infty(D)}).$$

If $u_1, u_2 \in C^t(D)$ for any $t > 0$. Then, for any $\epsilon > 0$,

$$|\mathcal{G}(u_1) - \mathcal{G}(u_2)| \leq C \, \exp\left(c \max\{\|u_1\|_{C^t(D)}, \|u_2\|_{C^t(D)}\}\right) \|u_1 - u_2\|_{L^\infty(D)}.$$

with $C = C(K, d, t, \epsilon, \|f\|_{L^r}, \|g\|_{L^{2r}})$ and $c = 4 + (4 + 2d)/t + \epsilon$.

Note that if instead of pointwise measurements of p, we consider the observations to be of the form $(l_1(p), \ldots, l_{\mathrm{K}}(p))$ where $l_j : H^1 \to \mathbb{R}$, $j = 1, \ldots, \mathrm{K}$, are bounded linear functionals, then one can get similar boundedness and continuity properties of $\mathcal{G}$ assuming u to be only essentially bounded on $\mathbb{T}^d$: Hölder continuity is not needed. However, since we construct the prior μ_0 using a countable orthonormal basis $\{\psi_l\}_{l\in\mathbb{N}}$, requiring the draws of μ_0 to be bounded in $L^\infty(\mathbb{T}^d)$ results, in any case, in more regular draws which lie in a Hölder space. This is because $L^\infty(\mathbb{T}^d)$ is not separable but any draw from μ_0 can be expanded in $\{\psi_l\}_{l\in\mathbb{N}}$. It is thus natural to consider the case of pointwise measurements since very similar arguments will also deal with the case of measurements which are linear functionals on H^1.

4.1. Well-definedness and continuity of the posterior measure. Now we can show the well-definedness of the posterior measure and its continuity with respect to the data for the elliptic problem. As we noted in Remark 1 by choosing $\{\psi_l\}_{l\in\mathbb{N}}$ of (3) as a wavelet or Fourier basis we can construct a Besov (κ, B^s_{qq}) or a Gaussian (κ, H^s) prior measure $(B^s_{22} \equiv H^s)$. We have the following theorem:

Theorem 4.1. *Consider the inverse problem for finding u from noisy observations of p in the form of (19) and with p solving (18). Let $f \in L^r(\mathbb{T}^d)$, $g \in L^{2r}(\mathbb{T}^d)$ and consider μ_0 to be distributed as a Besov (κ, B^s_{qq}) prior with $1 \leq q < \infty$, $s > d/q$, $\kappa > 0$ for $q = 2$ and $\kappa > 4r^*$ for $q \neq 2$ and r^* as in (11). Then the measure $\mu^y(\mathrm{d}u)$ is absolutely continuous with respect to μ_0 with Radon-Nikodym derivative satisfying*

$$\frac{d\mu^y}{d\mu_0}(u) \propto \exp\Big(-\frac{1}{2}\left|\Gamma^{-1/2}\big(y - \mathcal{G}(u)\big)\right|^2 + \frac{1}{2}|\Gamma^{-1/2}y|^2\Big).$$

Furthermore, the posterior measure is continuous in the Hellinger metric with respect to the data

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y'}) \leq C|y - y'|.$$

Proof. Let $t < s - d/q$ and $X = C^t(\mathbb{T}^d)$. The function

$$\Phi(u; y) := \frac{1}{2}\left|\Gamma^{-1/2}\big(y - \mathcal{G}(u)\big)\right|^2 - \frac{1}{2}|\Gamma^{-1/2}y|^2$$

satisfies (13) and Assumption 3.1(i) with $M = c\,r^2$, c depending on Γ, and $\alpha_1 = 0$. Using Proposition 2, Assumption 3.1(ii) and (iii) follow easily. By Theorem 2.1, $\mu_0(C^t(D)) = 1$ for any t such that $t < s - d/q$ and the absolute continuity of μ^y with respect to μ_0 follows by Theorem 3.2.

We note that

$$\begin{aligned}|\Phi(u; y_1) - \Phi(u; y_2)| &\leq \frac{1}{2}\big|\Gamma^{-\frac{1}{2}}\mathcal{G}(u)\big|\big|\Gamma^{-\frac{1}{2}}\big(y_1 - y_2\big)\big| \\ &\leq c_1 \exp\big(\|u\|_X\big)|y_1 - y_2|\end{aligned}$$

Hence Assumption 3.1 (iv) holds, and noting that $\alpha_2 = 1$, the continuity of μ^y with respect to the data follows from Theorem 3.3. □

4.2. APPROXIMATING THE POSTERIOR MEASURE. In this section, we consider the approximation of a sufficiently regular u in a finite-dimensional subspace of $L^2(\mathbb{T}^d)$ and use the Lipschitz continuity of $\mathcal{G}$ together with Theorem 3.4 to find an error estimate for the corresponding approximate posterior measure.

Let $\{\psi_l\}_{l\in\mathbb{N}}$ be an orthonormal basis of $L^2(\mathbb{T}^d)$ and $W^N = \operatorname{span}\{\psi_1,\dots,\psi_N\}$. Denote the orthogonal projection of $L^2(\mathbb{T}^d)$ onto W^N by P^N and let $\mathcal{G}^N = \mathcal{G}(P^N u)$. Define the approximated posterior measure $\mu^{y,N}$ by

$$\frac{\mathrm{d}\mu^{y,N}}{\mathrm{d}\mu_0}(u) = \frac{1}{Z^N(y)} \exp\big(-\frac{1}{2}|\Gamma^{-1/2}(y-\mathcal{G}^N(u))|^2\big) \tag{21}$$

with Z^N the normalizing factor.

Now we write $u \in L^2(\mathbb{T}^d)$ in a wavelet basis:

$$u(x) = u_1\phi(x) + \sum_{j=0}^{\infty} \sum_{(m,k)\in\Lambda_j} u_{m,k}\, \hat{\psi}_{m,k}(x). \tag{22}$$

In the above equation ϕ is the scaling function for $L^2(\mathbb{T}^d)$, $k = (k_1,\dots,k_d)$, $\Lambda_j = \{1,\dots,2^d-1\}\times\{0,\dots,2^j-1\}^d$, and for each fixed j,

$$\hat{\psi}_{m,k}(x) = 2^{j/2}\sum_{n\in\mathbb{Z}^d}\bar{\psi}_m(2^j(x-\frac{k}{2^j}-n))$$

where $\bar{\psi}_m$ are the mother wavelet functions for $L^2(\mathbb{R}^d)$(see Chapter 3 of [26]). We also assume that the above wavelet basis is r-regular, with r sufficiently large (see Remark 1).

We now impose one-dimensional indexing on the basis by setting $\psi_1 = \phi$ and using the following numbering [23, 26] for $\psi_l = \hat{\psi}_{m,k}$, $l > 1$,

$$\begin{aligned} &\text{for } j=0: \quad l = 2,\dots,2^d,\\ &\text{for } j=1: \quad l = 2^d+1,\dots,2^{2d},\\ &\qquad\vdots \end{aligned}$$

With this notation, the Karhunen-Loève expansion of a function u drawn from a (κ, B^s_{qq})-Besov prior μ_0 is the same as (3) and therefore the measure $\mu^{y,N}$ is an approximation to μ^y found by truncating the Karhunen-Loève expansion of the prior measure to N terms using the orthogonal projection P^N defined above.

We have the following result on the convergence of $\mu^{y,N}$ to μ^y as $N\to\infty$:

Theorem 4.2. *Consider the inverse problem of finding $u \in C^t(\mathbb{T}^d)$, with $t > 0$, from noisy observations of p in the form of (19) and with p solving (18) with periodic boundary conditions. Assume that the prior μ_0 is a (κ, B^s_{qq}) measure with $s > d/q + t$, $\kappa > 0$ for $q = 2$ and $\kappa > 8r^*(2 + (2+d)/t)$ otherwise. Then*

$$d_{\mathrm{Hell}}(\mu^y, \mu^{y,N}) \le C\, N^{-t/d}.$$

We note that, although for a fixed N the rate of convergence of approximated posterior measure to μ^y in the wavelet case is smaller than that of the Fourier case, where $d_{\mathrm{Hell}}(\mu^y,\mu^{y,N}) \le C\,N^{-t}$ (see [11]), one should take into account that we expect that the functions that solve the elliptic inverse problem of this section, have a more sparse expansion in a wavelet basis compared to the Fourier basis (see also section 9.4 of [12] or section 3.11 of [26]).

Proof of Theorem 4.2. Let V_0 and W_j be the spaces spanned by $\{\phi\}$ and $\{\hat{\psi}_{m,k}\}_{(m,k)\in\Lambda_j}$ respectively. Consider Q_j to be the orthogonal projection in $L^2(\mathbb{T}^d)$ onto W_j, and P_j the orthogonal projection of L^2 onto $\oplus_{k=1}^{j-1} W_k \oplus V_0$. For any $f \in C^t(\mathbb{T}^d)$ we can write [32, Proposition 9.5 and 9.6]

$$\|f - P_j f\|_{L^\infty} \leq C \sup_{0<|x-y|<2^{-j}} \|f(x) - f(y)\|_{L^\infty} \leq C\ 2^{-jt}\ \|f\|_{C^t}.$$

Here and in the rest of this proof we represent any constant independent of f and j by C. Using the above inequality, we have

$$\begin{aligned}\|Q_j f\|_{L^\infty} &= \|P_{j+1} f - P_j f\|_{L^\infty} \\ &\leq \|f - P_j f\|_{L^\infty} + \|f - P_{j+1} f\|_{L^\infty} \ \leq\ C\, 2^{-jt} \|f\|_{C^t}.\end{aligned}$$

Hence

$$\begin{aligned}\|u - P^N u\|_{L^\infty(D)} &\leq \sum_{j=J+1}^{\infty} \|Q_j u\|_{L^\infty} \\ &\leq\ C\, \|u\|_{C^t} \sum_{j=J+1}^{\infty} 2^{-jt} = C\, \|u\|_{C^t}\, 2^{-(J+1)t} \sum_{j=0}^{\infty} 2^{-jt} \\ &\leq\ C\, \|u\|_{C^t}\, 2^{-(J+1)t}\ \leq\ C\, \|u\|_{C^t}\, N^{-t/d}.\end{aligned}$$

By Proposition 2 we have

$$|\Phi(u) - \Phi(P^N u)| \leq C \exp\big(c_1\, \|u\|_{C^t(D)}\big)\, N^{-t/d},$$

with $c_1 > 4 + (4 + 2d)/t$. The result therefore follows by Theorem 3.4. ☐

Remark 3. Let $W^\perp$ be the orthogonal complement of W^N in $L^2(\mathbb{T}^d)$. Since μ_0 is defined by the Karuhnen-Loève expansion of its draws as in (3) using $\{\psi_l\}_{l\in\mathbb{N}}$, it factors as the product of two measures $\mu_0^N \otimes \mu_0^\perp$ on $W^N \oplus W^\perp$. Since $\mathcal{G}^N(u) = \mathcal{G}(P^N u)$ depends only on $P^N u$, we may factor $\mu^{y,N}$ as $\mu^{y,N} = \mu^N \otimes \mu^\perp$ where μ^N satisfies

$$\frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0^N}(u) = \frac{1}{Z^N} \exp\big(-\frac{1}{2}|\Gamma^{-1/2}(y - \mathcal{G}^N(u))|^2\big) \tag{23}$$

and $\mu^\perp = \mu_0^\perp$. With this definition of μ^N as a measure on the finite dimensional space W^N and having the result of Theorem 4.2 one can estimate the following weak errors (see Theorem 2.6 of [11]):

$$\|\mathbb{E}^{\mu^y} p - \mathbb{E}^{\mu^N} p^N\|_{L^\infty(\mathbb{T}^d)} \leq C\, N^{-t/d},$$

$$\|\mathbb{E}^{\mu^y}(p - \bar{p}) \otimes (p - \bar{p}) - \mathbb{E}^{\mu^N}(p^N - \bar{p}^N) \otimes (p^N - \bar{p}^N)\|_S \leq C\, N^{-t/d},$$

with p^N the solution of (18) for $u = P^N u$, $\bar{p} = \mathbb{E}^{\mu^y} p$, $\bar{p}^N = \mathbb{E}^{\mu^N} p^N$ and $S = \mathcal{L}(H^1(\mathbb{T}^d), H^1(\mathbb{T}^d))$.

5. Conclusion

We used a Bayesian approach [19] to find a well-posed probabilistic formulation of the solution to the inverse problem of finding a function u from noisy measurements of a known function $\mathcal{G}$ of u. The philosophy underlying this approach is that formulation of the problem on function space leads to greater insight concerning both the structure of the problem, and the development of effective algorithms to probe it. In particular it leads to the formulation of problems and algorithms which are robust under mesh-refinement [30]. Motivated by the sparsity promoting features of the wavelet bases for many classes of functions appearing in applications, we studied the use of the Besov priors introduced in [23] within the Bayesian formalism.

Our main goal has been to generalize the results of [30] on well-definedness and well-posedness of the posterior measure for the Gaussian priors, to the case of Besov priors. We showed that if the operator $\mathcal{G}$ satisfies certain regularity conditions on the Banach space X, then provided that the Besov prior is chosen appropriately, the posterior measure over X is well-defined and well-posed (Theorems 3.2 and 3.3). Using the well-posedness of the posterior on the infinite-dimensional space X, we then studied the convergence of the appropriate finite-dimensional approximations of the posterior. In finding the required conditions on $\mathcal{G}$, it is essential to know which functions of u have finite integral with respect to the Besov prior μ_0. In other words we need a result similar to the Fernique theorem from Gaussian measures, for the Besov case. A Fernique-like result for Hölder norms is proved in Theorem 2.1, and may be of independent interest.

As an application of these results, we have considered the problem of finding the diffusion coefficient of an elliptic partial differential equation from noisy measurements of its solution. We have found the conditions on the Besov prior which make the Bayesian formalism well-posed for this problem. We have also considered the approximation of the posterior measure on a finite-dimensional space spanned by finite number of elements of the same wavelet basis used in constructing the prior measure, and quantified the error incurred by such an approximation.

A question left open by the analysis in this paper is how to extract information from the posterior measure. Typical information desired in applications involves the computation of expectations with respect to the posterior. A natural approach to this is through the use of Markov chain-Monte Carlo (MCMC). For Gaussian priors there has been considerable recent effort to develop new MCMC methods which are discretization invariant [3, 9] in that they are well-defined in the infinite dimensional limit; it would be interesting to extend this methodology to Besov priors. In the meantime the analysis of standard Random Walk and Langevin algorithms in [4] applies to the posterior meausures constructed in this paper and quantifies the increase in computational cost incurred as dimension increases, resulting from the fact that the infinite dimensional limit is not defined for these standard algorithms. A second approach to integration in high dimensions is via polynomial chaos approximation [29] and a recent application of this approach to an inverse problem may be found in [28]. A third approach is the use of quasi-Monte Carlo methods; see [5]. It would be of interest to study the application of all of these methodologies to instances of the Besov-prior inverse problems constructed in this paper.

Acknowledgements

MD and AMS are grateful to the EPSRC (UK) and ERC for financial support.

References

[1] Abramovich, F.; Silverman, B. W. 1998, Wavelet decomposition approaches to statistical inverse problems. *Biometrika* 85, no. 1, 115–129.

[2] Abramovich F., Sapatinas T., Silverman B. W. 1998, Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60, no. 4, 725–749.

[3] A. Beskos, G. O. Roberts, A. M. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stochastic Dynamics*, 8(3):319–350, Sep 2008.

[4] A. Beskos, G. O. Roberts, and A. M. Stuart. Optimal scalings for local Metropolis-Hastings chains on non-product targets in high dimensions. *Ann. Appl. Prob.*, 19:863–898, 2009.

[5] , R.E. Caflisch. Monte carlo and quasi-monte carlo methods, Acta Numerica 1998, pages 1–49.

[6] Chambolle A., DeVore R. A., Lee, N., Lucier B. J. 1998, Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.* 7. no. 3, 319–335.

[7] S.L. Cotter, M. Dashti, J.C. Robinson, and A.M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse Problems*, 25:doi:10.1088/0266–5611/25/11/115008, 2009.

[8] S.L. Cotter, M. Dashti, and A.M. Stuart. Approximation of Bayesian inverse problems. *SIAM J. Num. Anal.*, **48**(2010), 322-345.

[9] S.L.Cotter, M. Dashti and A.M.Stuart. Variational data assimilation using targetted random walks. Int. J. Num. Meth. Fluids, To appear 2011.

[10] Da Prato G. & Zabczyk J. 1992, *Stochastic equations in infinite dimensions.* Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge.

[11] Dashti M. and Stuart A. 2011, Uncertainty quantification and weak approximation of an elliptic inverse problem. SIAM J. Num. Anal., to appear. See `http://arxiv.org/abs/1102.0143`

[12] Daubechies I. 1992 *Ten lectures on wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics, 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[13] Daubechies I., Defrise M. and De Mol C. 2004m An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* 57, no. 11, 1413–1457.

[14] Donoho D. L., Johnstone I. M. 1994, Ideal spatial adaptation by wavelet shrinkage. *Biometrika* 81, no. 3, 425–455.

[15] Engl H. K., Hanke M., and Neubauer A. 1996 *Regularization of Inverse Problems*, Kluwer.

[16] J.N. Franklin. Well-posed stochastic extensions of ill-posed linear problems. *J. Math. Anal. Appl.*, 31:682–716, 1970.

[17] T. Helin and M. Lassas. Hierarchical models in statistical inverse problems and the Mumuford-Shah functional. *Inverse Problems*, 32:015008, 2011.

[18] Hairer M., Stuart A.M. and Voss J. 2007, *Analysis of SPDEs arising in path sampling part II: the nonlinear case.* Annals of Applied Probability 17(5-6) 1657-1706.

[19] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer, 2005.

[20] Kahane J.-P. 1985, *Some random series of functions.* Cambridge University Press, Cambridge.

[21] S. Lasanen. Discretizations of generalized random variables with applications to inverse problems. *Ann. Acad. Sci. Fenn. Math. Diss., University of Oulu*, 130, 2007.

[22] S. Lasanen. Measurements and infinite-dimensional statistical inverse theory. *PAMM*, 7:1080101–1080102, 2007.

[23] Lassas M., Saksman E. and Siltanen S. 2009, Discretization invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging* **3**, 87-122.

[24] M.S. Lehtinen, L. Paivarinta, and E. Somersalo. Linear inverse problems for generalized random variables. *Inverse Problems*, 5:599–612, 1989.

[25] A. Mandelbaum. Linear estimators and measurable linear transformations on a hilbert space. *Prob. Theo. Rel. Fields*, 65:385–397, 1984.

[26] Meyer, Y. 1992, *Wavelets and operators.* Translated from the 1990 French original by D. H. Salinger. Cambridge Studies in Advanced Mathematics, 37. Cambridge University Press, Cambridge.

[27] P. Piiroinen. Statistical measruements, experiments and applications. *Ann. Acad. Sci. Fenn. Math. Diss., University of Helsinki*, 143, 2005.

[28] Schwab, Ch. and Stuart, A.M. 2011, *Sparse deterministic approximation of Bayesian inverse problems.* Submitted. See `http://arxiv.org/abs/1103.4522`

[29] P.D. Spanos and R. Ghanem. Stochastic finite element expansion for random media. *J. Eng. Mech.*, 115:1035–1053, 1989.

[30] Stuart A. M., 2010, Inverse Problems: A Bayesian Approach. *Acta Numerica*.

[31] Triebel H., 1983 *Theory of Function Spaces.* Birkhäuser, Boston.

[32] Wojtaszczyk P. 1997, A mathematical introduction to wavelets. London Mathematical Society Student Texts, 37. Cambridge University Press, Cambridge.